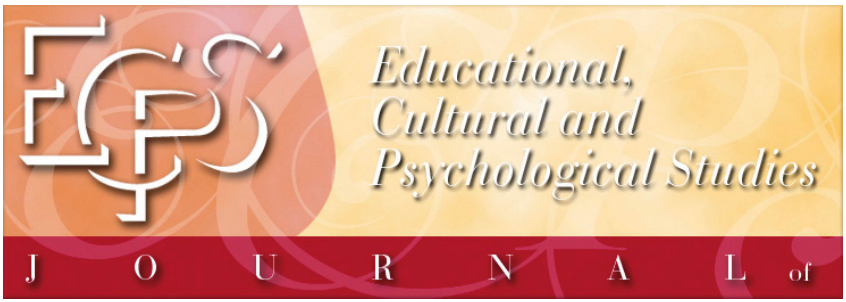

25
June 2022

Note di Ricerca
Research Notes

Commenti, Riflessioni, Presentazioni, Resoconti, Dibattiti, Interviste
Comments, Reflections, Presentations, Reports, Debates, Interviews

Investigating the Effect of Missing Data on the Experimental Test of Mathematics for the Second-Secondary Students

Aiman Freihat

Al-Balqa Applied University - Educational Sciences Department (Salt, Jordan); Ajloun University College (Jordan)

DOI: https://dx.doi.org/10.7358/ecps-2022-025-frei

aiman.freihat@bau.edu.jo

INDAGARE L'EFFETTO DEI DATI MANCANTI IN UN TEST SPERIMENTALE DI MATEMATICA PER GLI STUDENTI DELLA SCUOLA SECONDARIA DI SECONDO GRADO

Abstract

This study aimed to uncover the effect of the missing values using the Rasch model on an experimental test of mathematics for the second-secondary students. To achieve the goal of the study, a multiple-choice test was designed testing all basic skills found on the student's book. The test consisted of 103 items. This study was conducted in September 2019. The results were analyzed using SPSS, Winstep, and MG3-BILOG. To determine the correlation coefficients, the results showed that the test had good psychometric properties and the test had high reliability, validity indications, acceptable difficulty, and discrimination coefficients. The findings also showed that the differentiation coefficients of items were equal. The results also showed that the premises of the Rasch model were fulfilled. Further, the results showed that there were statistically significant differences between the three models according to the percentage of data loss, standard error, difficulty, internal and external matching, and bias, and were in favor of the original model before missing the values.

Keywords: Difficulty factor; Missing values; Modern evaluation; Psychological measures; Rasch model; Standard error.

1. Introduction

The Rasch model, with its characteristics and assumptions, cannot be realized in the presence of missing data. The methods of estimating responses and the ability parameter for a determinant are affected by the missing data. This model ignores these values and deals with them as zero. This may affect the standard error, difficulty curve of the information function, and the external and internal match. The process of missing data directly affects the accuracy of the parameters of the items, especially in psychological and educational tests. These tests were used in the production of question banks that may be adopted in diagnostic and achievement tests, especially those related to the evaluation process in schools. To study the impact of the Rasch model on the missing data, three data models were used as defined by (Embretson & Reise, 2000). The first one is Missing Completely at Random (MCAR). This type of model is not related in any way to the study variables, and it does not depend on the characteristics of the participants. Instead, it may be due to unintended reasons from the participants, such as forgetfulness. It may occur because participants prefer to withdraw from participating in the test.

The second model is Missing at Random (MAR). This type is affected by the participant's desire to disclose the response, which is directly affected by the nature of the independent variables. The participants may consider their responses as private things related to their tendencies and qualifications. The third model is Non-Ignorable Non-Responses (NINR). This type is related to the difficulty factor, the quality of matching the internal and external data, and the ability of the individuals. The missing is related to the characteristics of the item or the participants. The reasons for the missing values in the data are due to different reasons, such as the participants' lack of conviction about the relevance of alternatives to the content of the item, the participant's ignorance of the answer, personal matters, forgetfulness, or the time factor (Hidi & Renninger, 2019). Despite the ability of objective tests to accurately reveal students' abilities, they may face many defects because of missing values, which cause a variation in the responding to the test and leads to distortions in the logarithmic structure (Schneider *et al.*, 2013). Additionally, ignoring the missing values affected the items of the tests in the sense that they may not measure students' ability to synthesize and evaluate and their ability to organize information and link them with each other (Pidgeon & Yates, 2018). Researchers also focus on investigating the effect of missing values on the accuracy of the horizontal equation of the test using the paragraph response theory. In order for the researchers to obtain generalizations, they often generate data using specific

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

programs such as (WINGEN). The majority of studies have confirmed that there is a difference in estimating the capabilities of individuals depending on different methods of compensating for missing values. They also found statistically significant differences between the arithmetic means of the equation errors due to a variable of compensating for the missing values.

1.1. *The importance of the study*

The impact of missing values on the tests cannot be ignored at all, and the researchers cannot neglect them without knowing their effect on the tests. Hence, the researcher decided to reveal the impact of missing values on the mathematics test. At theoretical level, the study is very important to determine the most effective way to compensate for the missing data in terms of its impact on detecting the differential performance of the test items. At practical level, it gives a clear picture of the efficiency of the methods of compensation for missing data in terms of their impact on the detection of differential performance, which makes it easier for researchers to know the efficient way to deal with missing data. It provides researchers with the most efficient methods to ensure the validity, stability and impartiality of test results.

This study aims to uncover the effect of missing data on the mathematics test, using Rasch model. It investigates whether there are statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of difficulty and standard error of the test items according to the method of missing values. The study also explores whether there are statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of capacity estimates and standard error of the test items according to the method of missing values? It also investigates whether the level of bias in the test items differs according to the method of missing.

1.2. *The problem of the study*

The problem of the study lies in revealing the effect of the missing values by using the Rasch model on the national exam of mathematics for the tenth grade. The study answers the following questions: First: Are there statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of difficulty and standard error of the test items according to the method of missing values? Second: Are there statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of capacity estimates and standard error of

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

the test items according to the method of missing values? Third: Does the level of bias in the test items differ according to the method of missing?

2. Literature review

The impact of missing values on the data has gained the attention of many researchers, as many of them indicated the effect of missing values on the results of their research. Many studies showed that if missing data were neglected, the wrong outputs would be obtained, especially in the field of psychological measurement and educational studies (Weiss, 2014). Some studies dealt with the impact of missing values on the accuracy of estimating the parameters of the items according to the classical theories. In contrast, others dealt with methods of treatment.

In another study (Awwad, 2010), the authors conducted a study aimed at revealing the effect of methods of treating missing values in estimating the parameters of items and abilities. The sample of the study consisted of (1600) participants from eighth-grade students. The Otis-Lennon test was used for mental ability, which was composed of (80) items. The results indicated that there were statistically significant differences in the mean and standard error in estimating the ability of individuals and the characteristics of the item in favor of response function.

Similarly (Sarayra, 2018) conducted a study aimed at revealing the effect of missing values on the accuracy of estimating the information function and reliability. The sample consisted of (1000) of the tenth-grade students in Karak governorate. The researcher used a test consisting of (50) items and used three methods for the missing data. The results indicated that there were differences in estimation maximum information due to the method of missing and in favor of the least. Another study was conducted by (Hambleton & Cook, 1977), which aimed at identifying the best methods used in treating data missing. The sample consisted of (1200) participants. The results concluded that the best methods of dealing with missing data were two methods: calculating the arithmetic mean of participants and calculating the compensating value from an unconditional distribution. Another study conducted by (Aldarabsah, 2012) aimed at revealing the effect of the method of treating the missing values on the accuracy of the capacity estimates and the parameters of the items. The study generates 1500 responses using WINGEN program based on a test consisting of (80) double-response items. The results indicated that there are statistically significant differences in the accuracy of estimating the ability and the parameters of the items due to the method of missing and in favor of the least missing.

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

A study was conducted by (Waterbury, 2019) used the Rasch model to reveal the effect of missing data mechanisms, missing data ratios, sample size, and test length on standard biases and errors of item parameters. The results indicated that there are statistically significant differences due to the length of the test, the sample size, and the method of missing. The review of previous studies revealed that the majority of the previous studies have relied on the principle of simulation by generating virtual data and pre-prepared measures, using the classical measurement and the Rasch model.

3. Research model

George Rasch has built a mathematical model that provides the relationship between the ability of the individual and the difficulty of the item, which mainly affects the response of the individual to the item (Engelhard & Wang, 2019). This relationship can be mathematically expressed as follows:

$$\mathrm{P}(x_{vi}) = 1 - \frac{e^{\beta v - \delta_i}}{1+e^{\beta v \delta_i}}, (x_{vi} = 0) \tag{1}$$

$$\mathrm{P}(x_{vi}) = \frac{1}{1+e^{\beta v - \delta_i}} \tag{2}$$

This theory enables us to independently estimate the individual's ability from the sample. It interprets the individual's performance on a particular test through specific features. It also gives estimates for these features (Hambleton & Swaminathan, 2013). The relationship between ability and difficulty enables the researcher to observe the behavior of the individual and the frequency expressed through responses, and thus he/she predicts the performance of the individual on the test items through a set of factors

Rasch model can be used in most types of educational measures, and it is based on a set of basic assumptions that must be met in any model (Hambleton & Swaminathan, 2013), which includes: First: Unidimensionality: This theory assumes that the data should measure only one ability so that the response of the individual is attributed to that ability. This assumption is confirmed by using factor analysis. Second: Local Independence: This assumption is achieved through different factors: First, the independence of the individual's ability from the ability of other individuals who perform the same test. Second, item difficulty does not depend on other items. These factors make the responses of individuals statistically independent and not correlated at a specific ability level. That

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

is, the correlation coefficient between the answers of the items must equal zero In order to achieve local independence, the items must be readable, and the participants should be suitable in terms of the validity of their responses on the scale. It is possible to infer how closely the item fits the model by referring to specific criteria, as follows. The FIRST criterion: the extent to which the item measures the value in question and the extent of its compatibility with the rest of the items. This can be done by using the (T) test for Total Fit Statistics, which assumes that the test measures the same value in question. Thus, this test provides Unidimensionality (Bond *et al.*, 2020). The SECOND criterion: the extent to which the difficulty of the items is independent of the ability of the participants. This criterion can be tested through the use of Between Groups Statistic, which reveals the relative stability of the difficulty coefficients within the levels of ability and the deviation in the characteristics of the curves of the items.

3.1. *Participants*

The sample of the study consisted of 1000 students, chosen randomly, enrolled in public and private schools. The structure of participants is shown in *Table 1*.

Table 1. – The structure of participants.

GENDER	NUMBER
Males	500
Females	500
TOTAL	1000

3.2. *Data collection tools*

FIRST: The researcher designed a test consisting of 103 items. The sample consisted of 1000 students who were distributed into four groups according to the method of missing. Each group was asked to delete the questions according to the model and the percentage of missing that the researcher wants. The researcher calculated the mean and standard deviation of the difficulty estimates for the items, the standard deviation, and the internal and external fit of the standard missing squares.

SECOND: Missing of data: the researcher studied missing data at three percentages 50%, 10%, and 15% by assuring students not to answer questions according to the model and the desired percentage. Thus, the researcher would have three files according to the missing rates, as shown in *Table 2*.

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

Table 2. – Methods of missing data.

MAR	MCAR	MNAR
50%	10%	15%

Third: Verification of the item response theory according to the Rasch model This step aims to verify the assumptions of the item response theory according to the Rasch model for the three models. The unidimensional examination was performed using factor analysis, which is based on the basic components method. This assumption was verified using the SPSS program, Exploratory Factor Analysis, Principal Component Analysis, and Varimax Rotation method. Furthermore, Eigen Values and Explained Variance were calculated for each of the factors, as shown in *Table 3*. It is noticed from *Table 3* that the ratio of the explained variance for the first factor to the explained variance for the second factor is greater than 2. This confirms the fulfillment of the unidimensional assumption. *Figure 1* shows the scree of Eigenvalues of the test's components before missing the data.

Table 3. – Results of factor analysis.

Components	Data before missing	Data after missing		
		50%	10%	15%
1	10.9	10.381	5.232	4.173
2	1.801	1.715	1.219	1.161
Total of division	6.05	6.10	4.35	3.59

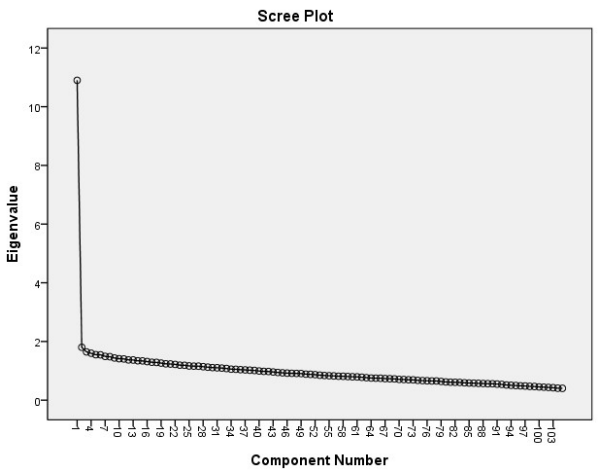

Figure 1. – Eigenvalues of the test's components before missing the data.

Fourth: Calculating information function BILOG-MG3 software was used to determine the amount of information provided by the test at different ability levels. It also provides the relationship between raw scores and ability, the estimation of the corresponding information function for each raw score and ability, and the estimation of the standard error and ability, as shown in *Table 4* and *Table 5*.

Table 4. – The amount of information provided by the test at different ability levels, and the relationship between raw score and capacity.

Raw score	Frequency	Free values	Standard error	Information function
0	0	0.89	1.83	0.3
1	0	0.94	1.01	0.99
2	0	0.98	0.71	1.96
3	0	1.03	0.51	3.84
4	0	1.08	0.46	4.75
5	0	1.14	0.42	5.64
6	0	1.19	0.39	6.5
7	0	1.24	0.37	7.35
8	0	1.3	0.35	8.18
9	0	1.36	0.33	8.99
10	0	1.42	0.32	9.78
11	0	1.61	0.31	10.55
12	0	1.68	0.3	11.3
13	0	1.75	0.29	12.03
14	0	1.83	0.28	12.74
15	0	1.91	0.27	13.43
16	0	2	0.27	14.1
17	0	2.09	0.26	14.76
18	0	2.19	0.25	15.39
19	0	2.29	0.25	16
20	0	2.41	0.25	16.59
21	0	2.54	0.24	17.17
22	1	2.69	0.24	17.72
23	2	2.85	0.23	18.26
24	1	3.05	0.23	18.78
25	1	3.28	0.23	19.27
26	5	3.58	0.23	19.75
27	7	4	0.22	20.21

Raw score	Frequency	Free values	Standard error	Information function
28	6	4.7	0.22	20.65
29	7	5.91	0.22	21.07
30	8	0.89	0.22	21.48
31	6	0.94	0.21	21.86
32	6	0.98	0.21	22.23
33	7	1.03	0.21	22.57
34	5	1.08	0.21	22.9
35	7	1.14	0.21	23.21
36	6	1.19	0.21	23.5
37	5	1.24	0.21	23.77
38	5	1.3	0.2	24.02
39	6	1.36	0.2	24.26
40	5	1.42	0.2	24.47
41	7	1.61	0.2	24.67
42	10	1.68	0.2	24.85
43	14	1.75	0.2	25.01
44	12	1.83	0.2	25.15
45	10	1.91	0.2	25.27
46	21	2	0.2	25.38
47	19	2.09	0.2	25.47
48	30	2.19	0.2	25.53
49	31	2.29	0.2	25.58
50	22	2.41	0.2	25.62
51	28	2.54	0.2	25.62

Table 5. – The amount of information provided by the test at different ability levels, and the relationship between raw score and capacity.

Raw score	Frequency	Free values	Standard error	Information function
52	37	2.69	0.2	25.63
53	34	2.85	0.2	25.60
54	32	3.05	0.2	25.56
55	44	3.28	0.2	25.5
56	34	3.58	0.2	25.42
57	45	4	0.2	25.33
58	33	4.7	0.2	25.22

Raw score	Frequency	Free values	Standard error	Information function
59	25	5.91	0.2	25.08
60	30	0.89	0.2	24.93
61	23	0.94	0.2	24.77
62	21	0.98	0.2	24.58
63	15	1.03	0.2	24.38
64	14	1.08	0.2	24.15
65	10	1.14	0.2	23.91
66	22	1.19	0.21	23.66
67	16	1.24	0.21	23.38
68	16	1.3	0.21	23.09
69	12	1.36	0.21	22.77
70	14	1.42	0.21	22.44
71	23	1.61	0.21	22.09
72	13	1.68	0.21	21.72
73	17	1.75	0.22	21.34
74	15	1.83	0.22	20.94
75	9	1.91	0.22	20.51
76	10	2	0.22	20.07
77	20	2.09	0.23	19.61
78	12	2.19	0.23	19.14
79	8	2.29	0.23	18.64
80	10	2.41	0.23	18.13
81	8	2.54	0.24	17.59
82	9	2.69	0.24	17.04
83	12	2.85	0.25	16.47
84	11	3.05	0.26	14.65
85	11	3.28	0.27	14
86	7	3.58	0.27	13.34
87	10	4	0.28	12.66
88	2	4.7	0.29	11.95
89	1	5.91	0.3	11.23
90	3	0.89	0.31	10.49
91	9	0.94	0.32	9.72
92	0	0.98	0.33	8.94
93	0	1.03	0.35	8.14
94	0	1.08	0.37	7.32

Raw score	Frequency	Free values	Standard error	Information function
95	0	1.14	0.39	6.48
96	0	1.19	0.42	5.61
97	0	1.24	0.46	4.73
98	0	1.3	0.51	3.83
99	0	1.36	0.59	2.9
100	0	1.42	0.71	1.96
101	0	1.61	1.01	0.99
102	0	1.68	1.83	0.3

It is noticed from *Table 4* and *Table 5* that the average ability is zero, meaning that it is at the maximum function of the information. The test gives the maximum function of the information at the average level of ability, as its value reached (25.62) at the raw degree (52), which corresponds to zero scales, as shown in the *Figure 2*.

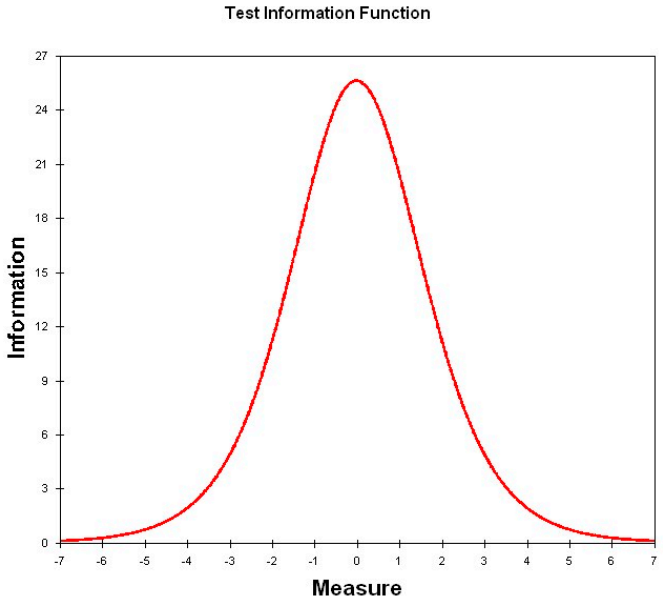

Figure 2. – Information function association resulting from the output of the Bilog-MG3 software.

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

It is also clear from *Table 4* and *Table 5*, that the following equations can be used to predict raw degree or ability.

$$\text{Rawscore} = \text{ability} \times 15.3850 + 52.4991 \quad (3)$$

$$\text{Ability} = \text{rawscore} \times 0.0593 + -3.1112 \quad (4)$$

4. Results and discussion

First: The results related to the first question, which states: Are there statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of difficulty and standard error of the test items according to the method of missing values? The mean and standard deviation were found for each estimate of item difficulty log units, the standard error, and the internal and external fit for the three models according to the missing method.

Table 6. – The mean and standard deviation of item difficulty for the items in log units, the standard error in measuring this difficulty, and the internal and external fit of statistical values for the three models according to the missing method.

Case 0.10	Difficulty	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	- 0.6260	0.1805	1.0191	- 0.4760	1.0528	- 0.443
Standard deviation	0.5198	0.00584	0.1448	0.607	0.2307	0.880
Case 0.15	Difficulty	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	- 1.22	0.221	1.331	- 0.22	1.161	- 0.30
Standard deviation	1.97	0.021	0.225	0.98	0.552	1.35
Case 0.50	Difficulty	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	- 3.54	0.2955	1.782	- 0.16	1.226	- 0.22
Standard deviation	1.09	0.092	0.659	1.01	0.606	1.55

It is evident from *Table 6* that there is an apparent difference for the mean values of the internal and external square mean, in the sense that they are close to one, which is the ideal situation that the model assumes. The results showed an apparent difference for the values of the standard error

of the internal and external averages of squares since they deviate from the values of the ideal of the model, which is (zero). We also notice that the values of the mean and the standard error of the (ZSTD) depart somewhat from the ideal values. The table also shows apparent differences in the difficulty factor according to the three missing methods. To demonstrate the significance of the differences, a factor analysis of repeated measurements was used, which requires that the data be spherical, followed by determining the appropriate statistic. The researcher used Mauchly's statistic, Geisser-Greenhouse and Lower-bound, as shown in *Table 7*.

It is evident from *Table 7* that there are statistically significant differences at the level ($\alpha \leq 0.05$) between the levels of the independent variable, which is the elimination method, since the level of significance was less than (0.05). Consequently, the sphericity condition was not met, and thus Greenhouse-Geisser was used to indicate the differences in the missing ratio.

The results in *Table 8* indicate the existence of statistically significant differences in the missing ratio. To find out the significance of the differences, the Scheffe test was used for dimensional comparisons, as shown in *Table 9*.

Table 7. – Sphericity data: results of Mauchly's statistic, Geisser-Greenhouse, and Lower-bound.

Effect	Mauchly's	χ^2	Freedom values	Level of significance	Greenhouse-Geisser	Lower-bound
Missing ratio	0.37	39.18	2	000	0.492	0.500

Table 8. – Results of the Greenhouse- Geisser test.

Source	Sum of squares	Freedom values	Mean squares	Value of corrected statistic	Level of significance
Missing data	0.033	1.28	0.55	32.18	0.00
Error	0.049	92.22	0.00038		

Table 9. – Results of the Scheffe test for dimensional comparisons.

Missing data	0.10	0.15	0.50
Scheffe			
0.10	–	0.000*	0.000*
0.15		–	0.000*
0.50	0.10	0.15	–

It is evident from a table shown in *Table 9* that there are statistically significant differences at the level of significance ($\alpha \leq 0.05$) in the mean of the

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

percentage missing of the degrees of the participants in favor of the ratio (0.5) followed by (0.15).

Second: The results related to the second question, which states: Are there statistically significant differences at the level of ($\alpha \leq 0.05$) in the level of the ability estimates and the standard error of the test items according to the method of missing values? The mean and standard deviation were found for the ability estimates for the items in log units, the standard error in measuring ability, and the internal and external conformity statistical values for the three models according to the missing method. It is evident from *Table 10* that there is an apparent difference for the mean values of the internal and external square means, in the sense that they are close to one, which is the ideal situation that the model assumes. The results revealed an apparent difference for the values of the standard error of the internal and external averages of squares since they deviate from the values of the ideal of the model, which is (zero). We also notice that the values of the mean and the standard error of the values of the (ZSTD) depart somewhat from the ideal values assumed by the model. The results also showed the existence of apparent differences in the ability estimates according to the difference in the missing methods. Therefore, the Greenhouse-Geisser statistic was used to show the differences in the capacity estimates of the missing ratio.

Table 10. – The mean and standard deviation of the ability estimates for the items in log units, the standard error in measuring ability, and the internal and external conformity statistical values for the three models according to the missing method.

Case 0.10	Ability	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	0.0907	0.2955	0.9988	0.00	1.0068	0.3
Standard deviation	1.80092	0.20877	0.0542	0.9	.069090	1.3
Case 0.15	Ability	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	0.0887	0.2867	0.9976	0.00	1.0005	0.31
Standard deviation	1.8651	0.2187	0.0577	0.92	.0665	1.31
Case 0.50	Ability	Standard error	Internal fit		External fit	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	0.0932	0.2876	0.9998	0.00	1.0075	0.29
Standard deviation	1.8051	0.2008	0.0576	0.9	.06541	1.29

Online ISSN 2037-7924 - Print ISSN 2037-7932 - ISBN 978-88-5513-063-9

Table 11. – The differences in the capacity estimates of the missing ratio.

Source	Sum of squares	Freedom values	Mean squares	Value of corrected statistic	Level of significance
Missing data	0.068	1.45	0.55	43.17	0.09
Error	0.088	98.11	0.00022		

The results in *Table 11* indicate that there are no statistically significant differences in the capacity estimates according to the missing ratio variable. Third: The results related to the second question, which states: Does the level of bias in the test items differ according to missing method? To answer the question, the maximum likelihood values of the bias and item difficulty were obtained, as shown in the following figures (*Fig. 3*).

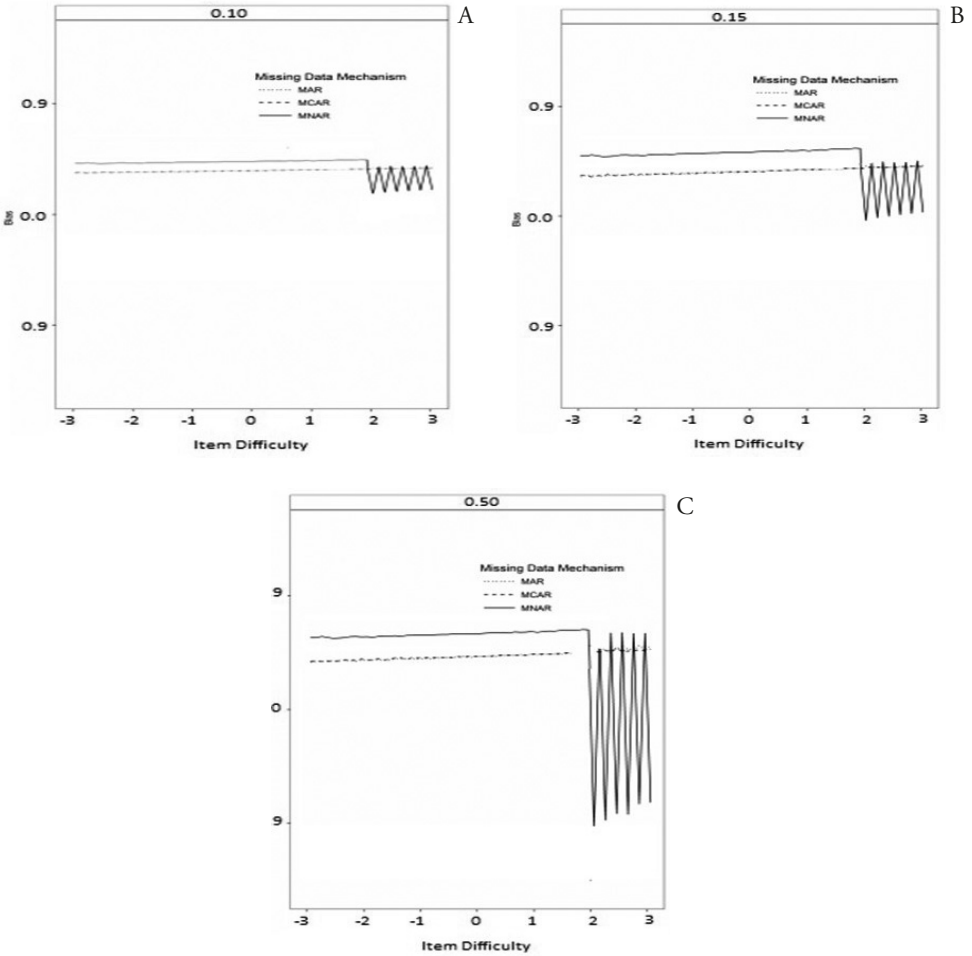

Figure 3. – The maximum likelihood values of the bias and item difficulty.

It is evident from the previous figures that the level of difficulty rises as the percentage of missing and bias decreases, as the differential performance of the vertebra increases. It appears from the MANAR figure that the answers were biased to a low degree. It is likely that the positive bias for residual material difficulties is greater than in the case of 50.

5. Conclusion

The study aimed to uncover the effect of missing values on an experimental mathematics test for the high school science secondary class using the one-parameter Rasch model. The results indicated the existence of statistical differences attributed to the method of missing, as the difficulty coefficients come closer to the ideal and acceptable values. The lower the missing ratio, the lower the standard error. The researcher attributes this result to the fact that the statistical program deals with missing values as zero, which affects the characteristics of the items and increases the amount of deviation of their values from the mean, and thus the standard errors increase. Also, the standard error in the ability estimates is affected by the difference in the missing ratios. The researcher believes that the percentage of decrease in the percentage of missing values in the data helps in raising the level of empirical reliability, which raises the level of estimation. Also, the level of difficulty rises as the percentage of missing and bias is less as the differential performance of the item increases. The current study recommends taking into account the missing values, levels of questions and students' ability when analyzing the test results. It also recommends to use essay questions in the tests. The study has some limitations. It analyzed the effect of the missing values using the Rasch model. The current study was limited to a sample taken public and private schools in Jordan.

References

Aldarabsah, Ryad (2012). *The effect of ability estimation method and handling method with missing values on the accuracy of items and persons' parameters*. Unpublished PhD Thesis, Yarmuk University (Jordan), Faculty of Education.

Awwad, Ali B. (2010). *Comparing methods of dealing with missing data in estimating items and persons parameters*. Unpublished PhD Thesis, Yarmuk University (Jordan), Faculty of Education.

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Eribaum Associates.

Engelhard, G., & Wang, J. (2019). Developing a concept map for Rasch Measurement Theory. In *Quantitative Psychology: 84th Annual Meeting of the Psychometric Society* (pp. 19-29), Santiago, Chile, 15-19 July. Cham (Switzerland): Springer International Publishing AG.

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, *14*, 75-96.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. New York: Springer Science & Business Media.

Hidi, S. E., & Renninger, K. A. (2019). Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review*, *31*(4), 833-852.

Pidgeon, D., & Yates, A. (2018). *An introduction to educational measurement*. London: Routledge.

Sarayra, Rajee (2018). The impact of missing data proportion and the handling methods in the accuracy of estimating the maximum value of item information function, and value of item reliability index. *Journal of Studies. Faculty of Educational Scienses*, *45*(8), 189-205.

Schneider, M. C., Egan, K. L., & Julian, M. W. (2013). Classroom assessment in the context of high-stakes testing. *SAGE Handbook of Research on Classroom Assessment*, 55-71.

Waterbury, G. T. (2019). Missing data and the Rasch model: The effects of missing data mechanisms on item parameter estimation. *Journal of Applied Measurement*, *20*(2), 154-166.

Weiss, D. J. (2014). *New horizon testing: Latent trait test theory and computerized adaptive testing*. New York: Elsevier.

Riassunto

Questo studio mirava a scoprire l'effetto dei valori mancanti utilizzando il modello Rasch su un test sperimentale di matematica per gli studenti della seconda secondaria. Per raggiungere questo obiettivo nel settembre 2019 è stato progettato un test comprendente 103 item a risposta multipla, che ha messo alla prova tutte le abilità di base su cui fa leva il manuale che deve preparare lo studente per sostenere l'esame. I risultati sono stati analizzati utilizzando SPSS, Winstep e MG3-BILOG. Nel calcolare i coefficienti

di correlazione, i dati hanno mostrato che il test ha buone proprietà psicometriche: alta affidabilità e validità, livelli di difficoltà e discriminatività accettabili. I risultati hanno anche mostrato che i coefficienti di differenziazione degli item sono uguali. Sono state soddisfatte le premesse del modello Rasch. Inoltre, sono state evidenziate differenze statisticamente significative fra i tre modelli elaborati in base alla percentuale di perdita di dati, errore standard, livelli di difficoltà, corrispondenza interna ed esterna e bias, esse risultano a favore del modello originale prima di perdere i valori.

Parole chiave: Errore standard; Indice di difficoltà; Misure psicologiche; Modello Rasch; Valori mancanti; Valutazione moderna.

How to cite this Paper: Freihat, A. (2022). Investigating the effect of missing data on the experimental test of mathematics for the second-secondary students [Indagare l'effetto dei dati mancanti in un test sperimentale di matematica per gli studenti della scuola secondaria di secondo grado]. *Journal of Educational, Cultural and Psychological Studies*, *25*, 59-76. DOI: https://dx.doi.org/10.7358/ecps-2022-025-frei