Journal of Environment and Sustainability Education
1 (1), 2023, 7-19
Available at: joease.id/index.php/joease

Characteristics analysis of integrated science question items with the rasch model

Sri Yamtinah[1]*, Ashadi[1], Bakti Mulyani[1], Budi Utami[1], Marantika Lia Kristyasari[2]

[1]Universitas Sebelas Maret, Ir. Sutami St., No. 36A, Surakarta, Central Java, 57126, Indonesia
[2]Universitas Masamus, Jl. Kamizaun Mopah Lama, Merauke, Papua, 99611, Indonesia
e-mail: jengtina@staff.uns.ac.id
* Corresponding Author.

Received: 6 January 2023; Revised: 15 February 2023; Accepted: 27 March 2023

Abstract: This study aims to analyze the characteristics of the Integrated Science question item. This is quantitative descriptive research. The research subjects were 90 students of class VIII Junior High School in the high, medium, and low categories in Surakarta City. The characteristics of the items were analyzed using the Winstep® program with the Rasch Model. Based on data analysis, the results obtained are: 1) the validity of conformity level of the items is valid; 2) the reliability value of the students was 0,79 which was sufficient, the reliability of the items was 0,93 which was very good, and the reliability between the students and the items (Alpha Cronbach) was 0,81 which was very good; 3) the value of separation item > value separation of students; 4) the distribution of the difficulty level of the items is generally in the medium category; and 5) the discriminating power of the questions in the very good category is 35%, 25% in the good, 30% in the moderate, and 10% in the unable to discriminate category. Thus, it can be concluded that the analysis of the characteristics of the Integrated Science question items using the Rasch Model is considered good.

Keywords: items characteristics; integrated science; rasch model

How to Cite: Yamtinah, S., Ashadi, A., Mulyani, B., Utami, B., & Kristyasari, M. L. (2023). Characteristics analysis of integrated science question items with the rasch model. *Journal of Environment and Sustainability Education*, *1*(1), 7-19.

Introduction

The learning process is a teaching and learning activity that involves patterns and processes of interaction between teachers and students in the context of implementing educational programs (Rooijakkers et al, 1991). There are several elements supporting learning activities, one of which is an evaluation tool. Evaluation of learning outcomes is an important element and the process cannot be separated from learning activities, because it aims to monitor the learning process, progress, and improve learning outcomes as well as to assess and measure the level of competency achievement of students (BNSP, 2009).

The exam is a procedure of learning evaluation carried out by the teacher on the knowledge and skills of students to find out their performance achievements. In exams, there are many types of questions that can be used by teachers to measure students' abilities, one of the most frequently used is multiple choice questions. However, it turns out that the usual multiple-choice questions still have some weaknesses. For this reason, the researcher modified this multiple-choice question to become a Two-Tier Multiple-Choice question. The questions used as evaluation tools also need to be analyzed to see the quality of the questions and to identify deficiencies in each item made by the teacher (Arikunto, 2012).

Item analysis is an activity carried out to examine each item through gathering information from students' answers in order to obtain quality questions before the questions are used (Ratumanan &

Laurens, 2011). Item analysis can be done qualitatively (content and form) and quantitatively (statistics). Quantitative analysis includes analysis of validity, reliability, level of difficulty, discriminating power, and distractor index of questions (Rahayu et al, 2014). Valid test items indicate that the instrument can be used to measure students' abilities. A reliable instrument is an instrument which when tested on the same object at different times will show the same results (Indrawan & Jalilah, 2021). Meanwhile, the level of difficulty and discriminating power of questions relates to the ability of students to answer questions properly and correctly.

This study aims to analyze Integrated Science question items using the Rasch Model. Rasch modeling can be used to measure quantitative analysis directly based on the probability principle that exists through the logarithmic function (Tennant et al, 2004). This aims to produce measurements with the same intervals called logit. Rasch modeling was chosen for the item analysis tool because it has several advantages that are not shared by other item analysis. The advantages possessed by the Rasch Model include: 1) being able to predict missing data, which is based on a systematic response pattern; 2) capable of producing standard error measurement values for the instruments used which can increase the accuracy of calculations; and 3) calibration which is carried out simultaneously in three aspects, namely measurement scale, person, and item (Sumintono, & Widhiarso, 2015).

The Rasch Model is very suitable for use in this study because it can evaluate the ability of Integrated Science item questions on students. The use of the Rasch Model is more effectively applied than the classical analysis (Fisher, 1993). Based on the explanation above, this paper aims to analyze the characteristics of the Integrated Science test items in one junior high school in Surakarta using the Rasch Model in terms of validity, reliability, level of difficulty, discriminating power, and the distractor index of the items.

Method

Data Source

The research method uses a quantitative descriptive method. The focus of the research was on Junior High School Integrated Science questions which totaled 40 questions consisting of 20 tier 1 questions and 20 tier 2 questions. The research subjects were 90 students of grade 8 junior high school in Surakarta City.

Analysis Method

Data analysis used quantitative descriptive analysis techniques. The data studied were then analyzed using the Rasch Model with the Winstep® application version 3.73. The analysis of the items produced by the Rasch Model is in the form of an analysis of validity, reliability, level of difficulty, and the discriminating power of the items as well as the distractor index.

Results and Discussion

The results and discussion section contains research findings obtained from the research data and hypotheses, the discussion of research results and comparison with similar theories and/or similar research. The results and discussion section can be divided into several sub-sections.

Validity

There are two types of validity analysis generated by Winstep®, namely construct and content validity. Content validity itself consists of the suitability level of the item which serves to see the quality of the item's suitability level with the model. The suitability level of the item is determined from the criterion value [8] as follows:

1. Accepted *Outfit Mean Square* (MNSQ) score: 0,5 < MNSQ < 1,5.
2. Accepted *Outfit Z-Standard* (ZSTD) score: -2,0 < ZSTD < +0,2.

3. Accepted *Point Measure Correlation* (Pt Measure Corr) score: 0,4 < Pt Measure Corr < 0,85.

An item can be said to be a fit question if it meets at least 2 of the criteria above and is corrected if it only meets 1 criterion and must be discarded if none of the criteria can be met by the item. The suitability value of the item is strongly influenced by the amount of data. The larger the sample used, the better the suitability level. The suitability level of the items in the current data analysis is shown in the output tables of item fit presented in Table 1.

Table 1. The results of the validity analysis of the suitability level of the items

No	Item	*Outfit MNSQ*	*Outfit ZSTD*	*Pt Mean Corr*	Note
1.	S7B	1.25	2.0	0.21	
2.	S15A	1.23	1.9	0.17	
3.	S12A	1.20	1.6	0.24	
4.	S8A	1.13	0.5	0.23	
5.	S20B	1.13	1.0	0.27	
6.	S2B	1.01	0.1	0.22	
7.	S15B	1.12	1.1	0.27	
8.	S1B	1.08	0.7	0.27	
9.	S11B	1.10	1.0	0.28	
10.	S8B	1.09	0.4	0.22	
11.	S17A	1.09	0.8	0.31	
12.	S16B	1.05	0.4	0.30	
13.	S13B	1.04	0.4	0.31	
14.	S9A	0.97	-0.1	0.32	
15.	S12B	0.97	-0.2	0.37	
16.	S6A	1.01	0.2	0.40	
17.	S16A	1.00	0.1	0.28	
18.	S4A	0.83	-0.1	0.22	
19.	S9B	0.91	-0.5	0.39	Infit Item
20.	S18A	0.87	-0.6	0.39	
21.	S10A	0.80	-0.5	0.35	
22.	S10B	0.67	-0.9	0.35	
23.	S1A	0.93	-0.6	0.43	
24.	S14B	0.84	-1.1	0.45	
25.	S2A	0.79	-0.4	0.33	
26.	S13A	0.81	-1.2	0.46	
27.	S7A	0.90	-0.8	0.48	
28.	S6B	0.88	-1.1	0.51	
29.	S5B	0.64	-1.1	0.43	
30.	S19B	0.61	-1.1	0.42	
31.	S17B	0.86	-0.7	0.51	
32.	S14A	0.76	-1.6	0.52	
33.	S5A	0.74	-1.4	0.50	
34.	S18B	0.74	-1.6	0.53	
35.	S19A	0.56	-1.3	0.47	
36.	S4B	0.70	-2.4	0.62	
37.	S3B	1.41	2.2	0.18	
38.	S3A	1.39	2.2	0.15	
39.	S11A	1.33	2.7	0.09	Outfit Item
40.	S20A	1.30	2.4	0.23	

Based on the criteria for the suitability level of the items, Table 1 shows that the results of the validity analysis of the suitability level of the items from 40 Integrated Science questions consisting of 20 tier 1 questions and 20 tier 2 questions obtained as many as 37 questions declared infit because they met at least 2 criteria for item suitability levels be it the MNSQ outfit, ZSTD outfit, or Pt Mean Corr. Questions that have been declared infit above can be assessed as appropriate and quality items

because they can guarantee the level of understanding of students. Meanwhile, 3 questions were declared as unfit questions because they only met 1 good item suitability level criteria. Even so, the 3 questions that were declared outfit could still be used for further trials, if they had been corrected first.

Reliability

Rasch modeling can also carry out analysis into three aspects, namely aspects of the item, person, and instrument in more detail. This can guide and facilitate teachers in the process of making questions and in taking appropriate, logical and scientific policies based on the guidelines for instrument quality criteria [8] as follows:

1. *Person* dan *item measure* shows the average value of students and item questions. The lowest value indicates the level of ability of students and the lowest questions are questions that are close to a logit value of 0.0.
2. *Alpha Cronbach Score* is the value used to measure the reliability between the person and the question item. The criteria used:
   < 0,5 : very poor.
   0,5 – 0,6 : poor.
   0,6 – 0,7 : moderate.
   0,7 – 0,8 : good.
   > 0,8 : very good.
3. *Person reliability* and *item reliability Scores* is the value of the reliability of students and the value of the item reliability *questions*. Category used:
   < 0,67 : weak.
   0,67 – 0,8 : moderate.
   0,81– 0,9 : good.
   0,91 – 0,94: very good.
   > 0,94 : outstanding.
4. *Separation* is a grouping of students and items. The greater the separation value, the better the quality of the instrument in all respondents (students) and items. This is because these questions can identify groups of respondents and groups of items.

The reliability of the items using the Winstep® program is shown through a statistical summary. The results of the summary statistics can be seen in Table 2.

Table 2. Student reliability value and question items

Variable	SD	Reliability	*α Cronbach*
Students	0.80	0.79	0.81
Items	0.98	0.93	

Table 2. Shows the results of the student's reliability value of 0.79 which implies that the student's reliability is in the moderate category because it is in the range of 0.67 – 0.80. Student reliability aims to measure the consistency of students' answers. So that when viewed from the results of the reliability of students, the consistency of students' answers is considered moderate. Meanwhile, the item reliability value was 0.93, which means that the item reliability was very good because it was in the range of 0.91 – 0.94. Cronbach's α value is 0.81. The value of 0.81 indicates that the reliability value of Cronbach's α is included in the good category, because the range of values is in the range of 0.80 – 0.90. Thus, based on Cronbach's α value, it can be stated that there is good interaction between students and the items. The results of the reliability analysis can be concluded that, although the consistency of students' answers is considered sufficient, the quality of the items is very good, thus creating good interaction between students and the items.

Level of Difficulty

The item difficulty level (item measure) provides a detailed display of the logit value of each item. The output results from the table generated from Winstep® provide information about the items whose

results are sorted from those with the highest logit measure value to the lowest one. This shows that the questions are sorted from the most difficult item to the easiest item. The grouping of the difficulty level of the questions can be determined based on the average logit value added to the standard deviation (SD) value obtained. This value is useful for identifying groups of items (separation). The item separation value obtained is 2.61. The grouping of item items can be determined through Equation (1) below.

$$H = \frac{[(4\ x\ \text{SEPARATION})+1]}{3} \quad (1)$$

Note:
H: Level score (grouping)
Separation: Respondent separation value

From the above equation, it is obtained:

$$H_{item} = \frac{[(4\ x 2{,}61)+1]}{3} = \frac{11{,}44}{3} = 3{,}81$$

The H_{item} value is 3.81 which is rounded to 4. This implies that the item separation is of good value because it can divide the questions into 4 groups, namely, very high, high, easy, and very easy.

The distribution of item difficulty levels in detail can be seen in the item measure output. From the output of this measure item, the average logit measure value and the SD value are obtained. In the Rasch model, there are four categories of measure (SD) values as determinants of the difficulty level of the items [8], namely:

< -1 : very easy item
-1 sampai 0 : easy item
0,1 sampai 1 : difficult item
> 1 : very difficult item

The average value of the logit measure and SD values were 0.00 and 1. From the SD results obtained, it can be concluded that the difficulty level of the items is in the difficult category. A summary of the results of the item measure against the difficulty level of the item is presented in Table 3.

Table 3. The results of the item measure the difficulty analysis of the items

Item Statistic	Measure	Note
Mean	0.00	Difficult Items
SD	1.00	

While the grouping of item difficulty levels in detail can be seen in Table 4.

Table 4. Item difficulty level group

Logit value	Item	Category	%
< -1	S4A, S2A, S19B, S19A, S10B, S8A, S16A, S10A, S8B, S5B	Very Easy	25
-1 – 0	S2B, S18A, S5A, S9A, S18B, S3A, S14A, S13A, S9B, S14B	Easy	25
0.1 – 1	S20B, S7A, S20A, S16B, S7B, S12B, S1A, S1B, S17A, S11A, S13B, S11B, S15B, S6A,	Difficult	35
> 1	S6B, S15A, S12A, S4B, S3B, S17B	Very Difficult	15
	Total		100

The results of grouping the level of difficulty of the items in Table 4 confirm the validity of the results obtained in Table 3 which states that the level of difficulty of the items is in the difficult category. According to the Rasch model, the level of difficulty of items like this is included in the poor category because the items tend to lead to the difficult category even though the percentage shown is only 35% of questions, around 14 questions. Meanwhile, 25% had very easy and easy category questions with 10 questions each. 15% of the questions were in the very difficult category with a total of 6 questions. With this Rasch modeling, information about the level of difficulty of the items can be provided in detail (Curtis, & Boman, 2007). In addition, through the Rasch model it is also highly recommended as a data analysis tool in the field of Education, especially in analyzing items to obtain good problem instruments and can measure the level of understanding of students in detail (Ismail et al, 2010).

In addition to grouping categories, at the level of difficulty of the items, the Rasch modeling can also analyze between students and the items. The grouping of students is determined by the results of separation through the calculation of equation (1). Student separation value is 1.97. From equation (1), it is obtained:

$$H_{student} = \frac{[(4\ x1{,}97)+1]}{3} = \frac{8{,}88}{3} = 2{,}96$$

The $H_{student}$ value is 2.96 which is rounded to 3. This implies that the separation of students is of good value because it can divide students into 3 abilities, namely high, middle, and low abilities. The distribution of students' abilities in detail can be seen on the variable maps.

The results of the analysis between students and the items are presented in Table 5.

Table 5. Results of analysis on students and items

Category	Student	Total	%
Low ability	85P, 79L, 64L, 44P, 69L, 68L, 32P, 88P, 84P, 86P, dan 65P	11	12
Middle ability	75P, 17L, 70L, 36P, 15P, 6P, 80L, 78P, 73L, 72L, 45L, 41P, 77L, 66L, 60L, 89L, 82L, 71P, 63P, 58L, 55L, 53L, 34P, 61P, 74L, 57P, 56L, 48P, 39P, 38L, 35L, 10L, 83P, 81P, 62P, 49L, 40P, 33L, 50L, 47P, 43P, 31P, 16P, 76P, 59P, 46L, 22L, 19L, 27P, 14L, 2P, 1L, 87P, 52P, 37P, 26P, 23P, 20L, 5P, 90P, 67P, 51P, 28P, 24P, 18P, 11P, 54P, dan 4P	69	77
High ability	7P, 30L, 21P, 8P, 42P, 9L, 3P, 13L, 29P, 12L	10	11
	Total		100

The results of the analysis between students and the items above are proof of confirmation of the results of the reliability of students and implies that the abilities possessed by students are dominated by middle abilities with a percentage of 77% while the rest are spread over students with high abilities of 11% and low abilities. 10%. Thus, stating that the ability of students to answer questions correctly is quite good because students have understood the concept of the material intended in the questions correctly. This is one of the advantages of using the Rasch Model to analyze the items, namely that we can measure the level of differences in students' ability to answer the given item (Andersen, 1973).

Discriminating Power

The determination of the discriminating power of the items in Rasch is seen from the value of the S.E Model and to find out the distribution of the discriminating power of the items can be determined through the value of the Point Measure Correlation (Pt Measure Corr). The discriminating power of the items is considered good if the Model S.E value is < 1 logit. The Point Measure Correlation (Pt Measure Corr) value is accepted if the value ranges from 0.4 < Pt Measure Corr < 0.85. The results of the S.E model and Point Measure Correlation (Pt Measure Corr) [8] can be seen in Table 6.

Table 6. Result of S.E model and point measure correlation

No.	S.E Model	PT-Measure Corr.	Item
1.	0.25	0.51	S17B
2.	0.25	0.18	S3B
3.	0.24	0.62	S4B
4.	0.23	0.24	S12A
5.	0.23	0.17	S15A
6.	0.23	0.51	S6B
7.	0.23	0.40	S6A
8.	0.23	0.27	S15B
9.	0.23	0.28	S11B
10.	0.23	0.31	S13B
11.	0.23	0.09	S11A
12.	0.23	0.31	S17A

No.	S.E Model	PT-Measure Corr.	Item
13.	0.23	0.27	S1B
14.	0.23	0.43	S1A
15.	0.23	0.37	S12B
16.	0.23	0.21	S7B
17.	0.23	0.30	S16B
18.	0.23	0.23	S20A
19.	0,23	0,48	S7A
20.	0,23	0,27	S20B
21.	0,24	0,45	S14B
22.	0,24	0,39	S9B
23.	0,24	0,46	S13A
24.	0,24	0.52	S14A
25.	0,24	0,15	S3A
26.	0,25	0,53	S18B
27.	0,25	0,32	S9A
28.	0,25	0,50	S5A
29.	0,25	0,39	S18A
30.	0,26	0,22	S2B
31.	0,31	0,43	S5B
32.	0,31	0,22	S8B
33.	0,31	0,35	S10A
34.	0,31	0,28	S16A
35.	0,33	0,23	S8A
36.	0,33	0,35	S10B
37.	0,33	0,47	S19A
38.	0,33	0,42	S19B
39.	0,36	0,33	S2A
40.	0,43	0,22	S4A

The results in Table 6. show that the S.E model has a value of < 1 logit. This explains that the discriminating power of the items is stated to be good. Meanwhile, the distribution of the discriminating power of the items can be seen from the various Point Measure Correlation values. The classification of Point Measure Correlation values is divided into 4 groups (Alagumalai et al, 2005). namely:

1. > 4 : very good
2. 0,3 – 0,39 : good
3. 0,2 – 0,29 : moderate
4. 0,0 – 0,19 : unable to discriminate
5. < 0 : need item inspection

The distribution of discriminating power items in more detail is presented in Table 7.

Table 7. The distribution of the items discriminating power

PT-Measure Corr Value	Number of Items	Category	%
> 0,4	**14 questions** (S17B, S4B, S6B, S6A, S1A, S7A, S14B, S13A, S14A, S18B, S5A, S5B, S19A, dan S19B)	Very good	35
0,3 – 0,39	**10 questions** (S2A, S10B, S10A, S18A, S9A, S9B, S16B, S12B, S17A, dan S13B)	Good	25
0,2 – 0,29	**12 questions** (S12A, S15B, S11B, S1B, S7B, S20A, S20B, S2B, S8B, S16A, S8A, dan S4A)	Moderate	30

0,0 – 0,19	**4 questions** (S3B, S15A, S11A, dan S3A)	Unable to discriminate	10
	Total		100

From the results of Table 7, the distribution of the discriminating power of the items is clearer and more detailed. The results from Table 7. show that 35% of the item's discriminating power is in the very good category, 25% is good, 30% is moderate, and 10% is unable to discriminate. Based on the data analysis, it appears that the percentage results for the good category are lower than the moderate category. This is because the ability of the majority of students is in the moderate group category while the difficulty level of the items is in the difficult category. So that the interaction between students and the items is also in the moderate category, and causes the greater percentage of discriminating power in the moderate category.

Distractor Index

The index of distractors or distractor index in the Rasch Model can be determined through the value of average ability. The distractor index is declared to function properly if the score value is 0 and the average ability is positive, and there is no (*) sign (Sumintono & Widhiarso, 2015). Based on the analysis using the Rasch Model, the results of the distractor index can be seen in Table 8.

Table 8. Result of distractor index

Item	Data Code	Score Value	Average Ability
S3B	A	0	0.31
	B	0	0.73
	D	0	0.74
	C	1	1.05
S3A	C	0	0.16
	A	0	0.73
	D	0	0.74
	B	1	0.89
S11A	**D**	**0**	**-0.49**
	C	0	0.34
	B	0	0.97
	A	***1***	***0.87****
S20A	A	0	0.47
	D	0	0.48
	B	0	0.80
	C	1	0.98
S7B	**D**	**0**	**-0.19**
	C	0	0.25
	B	0	0.81
	A	1	0.96
S15A	D	0	0.43
	C	0	0.55
	A	0	0.79
	D	1	0.99
S12A	D	0	0.36
	C	0	0.47
	A	0	0.87
	B	1	1.06
S8A	**C**	**0**	**-0.81**
	A	0	0.39
	D	0	0.56
	B	1	0.87
S20B	A	0	0.09

Item	Data Code	Score Value	Average Ability
	C	0	0.67
	B	0	0.88
	D	1	0.98
S2B	D	0	0.14
	A	0	0.17
	C	0	0.61
	B	1	0.91
S15B		0	0.02
	A	0	0.27
	B	0	0.28
	D	0	0.86
	C	1	1.06
S1B	D	0	0.19
	C	0	0.38
	B	0	0.75
	A	1	1.01
S11B	B	0	0.01
	D	0	0.14
	C	0	0.86
	A	1	1.04
S8B	**B**	**0**	**-0.24**
	A	0	0.17
	D	0	0.70
	C	1	0.80
S17A	A	0	0.23
	C	0	0.67
	B	0	0.67
	D	1	1.05
S16B	A	0	0.03
	B	0	0.62
	C	0	0.68
	D	1	1.03
S13B	B	0	0.24
	C	0	0.68
	D	0	0.98
	A	1	1.06
S9A	**D**	**0**	**-0.14**
	B	0	0.19
	A	0	0.55
	C	1	0.98
S12B	**C**	**0**	**-0.48**
	D	0	0.23
	A	0	0.47
	B	1	1.08
S6A	**A**	**0**	**-0.41**
	B	0	0.52
	D	0	0.57
	C	1	1.19
S16A	**C**	**0**	**-0.30**
	D	0	0.37
	B	0	0.46
	A	1	0.90
S4A	**C**	**0**	**∞**
	B	**0**	**-0.19**
	D	0	0.20

Item	Data Code	Score Value	Average Ability
	A	1	0.85
S9B	B	0	01.4
	D	0	0.30
	A	0	0.31
	C	1	1.04
S18A	**D**	**0**	**-0.41**
	B	0	0.14
	C	0	0.45
	A	1	1.01
S10A	**D**	**0**	**-0.34**
	B	**0**	**-0.05**
	C	0	0.41
	A	1	0.92
S10B	**A**	**0**	**-0.15**
	B	**0**	**-0.11**
	C	0	0.64
	D	1	0.91
S1A	B	0	0.30
	A	0	0.38
	C	0	0.48
	D	1	1.14
S14B	B	0	0.03
	A	0	0.15
	D	0	0.45
	C	1	1.09
S2A	**A**	0	**-0.33**
	B	0	**-0.27**
	D	0	1.49
	C	***1***	***0.90****
S13A	**A**	**0**	**-1.07**
	C	0	0.18
	D	0	0.38
	B	1	1.08
S7A	A	0	0.11
	C	0	0.28
	B	0	0.54
	D	1	1.14
S6B	**A**	**0**	**-0.07**
	B	0	0.28
	C	0	0.58
	D	1	1.34
S5B	**C**	**0**	**-0.41**
	A	**0**	**-0.05**
	B	0	0.16
	D	1	0.95
S19B	**B**	**0**	**-0.40**
	A	**0**	**-0.23**
	C	**0**	**-0.06**
	D	1	0.94
S17B	A	0	0.44
	C	0	0.49
	D	0	0.58
	B	1	1.53
S14A	**A**	**0**	**-0.58**
	C	**0**	**-0.13**

Item	Data Code	Score Value	Average Ability
	B	0	0.54
	D	1	1.11
S5A	**D**	**0**	**-0.30**
	A	0	0.08
	B	0	0.09
	C	1	1.07
S18B	**B**	**0**	**-0.11**
	A	0	0.17
	C	0	0.19
	D	1	1.11
S19A	**A**	0	**-0.62**
	C	0	**-0.29**
	D	0	**-0.19**
	B	1	0.95
S4B	**C**	**0**	**-0.04**
	B	0	0.07
	A	0	0.49
	D	1	1.54

Table 8 provides the results of the analysis of the distractor index. The answer option with a score value of 1 means that it is in accordance with the answer key, while average ability is used to see the performance of the answer options and the answer key to the item with answers that also do not work well. The answer options are said to have performance as an effective distractor index if the average ability is not negative, and answer keys can also work well if the average ability is marked (*). From the data presented in Table 8, the results obtained are several answer options that are still not functioning properly as distractor index and there are two answer keys that are also not working optimally. The performance of the distractor index and the answer key is summarized in Table 9.

Table 9. Summary of the performance of the distractor index and answer keys

Distractor Index			
No	Item	Answer	Note
1.	S11A	D	Not an effective distractor index
2.	S7B	D	
3.	S8A	C	
4.	S8B	B	
5.	S9A	D	
6.	S12B	C	
7.	S6A	A	
8.	S16A	C	
9.	S4A	B	
10.	S18A	D	
11.	S10A	A, B	
12.	S10B	C, A	
13.	S2A	A, B	
14.	S13A	A	
15.	S6B	A	
16.	S5B	C, A	
17.	S19B	B, A, C	
18.	S14A	A, C	
19.	S5A	D	
20.	S18B	B	
21.	S19A	A, C, D	
22.	S4B	C	
Key Answer			
No.	Item	Answer	Note

1.	S11A	A	Not an effective distractor index
2.	S2A	C	

Based on the results of the performance summary of the distractor index and the answer key in Table 9, it can be interpreted that there are still many answer options that do not function properly as a distractor index. There are 22 questions whose answer options still do not function effectively as a distractor index, both items at tier 1 and tier 2. This can be seen from Table 8. The data shows that the answer options for the 22 items have a negative average ability. This can happen because the group of students who have middle and high abilities choose more of these answer options compared to students who have low abilities. In this analysis, there were 2 questions whose answer keys also did not work well, because many students were distracted by other answer options. In question 2A, many students chose option D rather than the answer key itself contained in option C. The same thing was also found in question 11A, option A (as the answer key) had a lower average ability compared to option B. The theory of average ability in answer key options is worth more than the average ability of other options. However, the results of the analysis obtained state that the value of the average ability of the other answer options is greater than the value of the average ability of the answer options that serve as the answer key. This is what causes the average ability value of the answer keys for the two questions to be marked (*) which means that the performance ability of the answer options in these two questions is not optimal as an answer key.

Conclusion

Analysis of the Integrated Science items in Grade 8 Middle School obtained the results of the validity of the suitability level of the items as many as 37 infit questions and 3 outfit questions. Student reliability 0.79 means moderate, item reliability 0.93 means very good, and Alpha Cronbach reliability 0.81 means very good. So, the general interaction between students and the items is very good. The distribution of the difficulty level of the questions is as much as 5% very difficult, 15% difficult, 55% moderate, 22.5% easy, and 2.5 very easy. Therefore, overall, this question is included in the moderate category. The discriminating power of the items had various results, 35% of the items had very good discriminating power, 25% of the items were good, 30% of the items were moderate, and 10% of the items are unable to discriminate. Overall, the discriminating power of this item is in the very good category. The distractor index also works fine. The conclusion that can be drawn is that the analysis of the characteristics of the Integrated Science items using the Rasch Model is stated to be valid, reliable, the level of difficulty and the discriminating power of the items is very good, and the distractor index also functions well. For this reason, the Integrated Science questions meet the criteria as good questions.

Acknowledgment

The author would like to thank the Institute for Research and Community Service of Universitas Sebelas Maret, Surakarta, Indonesia for providing grant funding for this research through the HGR research group in 2020.

References

Alagumalai, N. S., Curti, D.D., & Hungi. (2005). *Applied Rasch measurement: Book of exemplars.papers in honour of John P. Keeves*.Dordrecth: Springer.

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38(1), pp. 123–140.

Arikunto, S. (2012). Dasar-dasar Evaluasi Pendidikan. *Jakarta: Bumi Aksara*.

BNSP. (2009). Peraturan Pemerintah Republik Indonesia nomor 19 tahun 2005 tentang standar nasional pendidikan. Departemen Pendidikan Nasional Republik Indonesia.

Curtis, D. D., & Boman, P. (2007). X-Ray Your Data with Rasch. *International Education Journal*, 8(2), 249-259.

Fisher, A. G. (1993). The assessment of IADL motor skills: an application of many-faceted Rasch analysis. *American Jornal of Occupational Therapy*, 47(4), pp. 319–329.

Indrawan, D., & Jalilah, S. R. (2021). Metode kombinasi/campuran bentuk integrasi dalam penelitian. Jurnal Studi Guru Dan Pembelajaran, 4(3), 735-739.

Ismail,I. Idrus, R. M. & Johari, S. S. M. (2010). Acceptance on Mobile Learning via SMS: A Rasch Model Analysis. *International Journal of Interactive Mobile Technologies,* 4(2), pp. 10–16.

Rahayu, T. D. Purnomo, B. H. & Sukidin, S. (2014). Analisis tingkat kesukaran dan daya beda pada soal ujian tengah semester ganjil bentuk pilihan ganda mata pelajaran ekonomi kelas X di SMA Negeri 5 Jember tahun ajaran 2012-2013. *Jurnal Edukasi UNEJ*, 1(1), pp. 39–43.

Ratumanan, T. G., & Laurens, T. (2011). Penilaian Hasil Belajar pada Tingkat Satuan Pendidikan. *Ambon: Unesa University Press*.

Rooijakkers, A., Riberu, J., & Sornoro. (1991). Mengajar dengan sukses. *Gramedia*.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi pemodelan rasch pada assessment pendidikan. *Trim komunikata*.

Tennant, A. McKenna, S. P. & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments," *Value in Health*, 7(1), pp. S22–S26.