Feasibility of Analytical Thinking Test Instrument: An Analysis of Test Quality and Learner Abilities Using the Rasch Model

Urai Nurbaiti, Afandi *, Andi Besse Tenriawaru
Biology Education Study Program, Faculty of Teacher Training and Education, Tanjungpura University
*Email: Afandi@fkip.untan.ac.id

DOI: https://doi.org/10.33369/pendipa.8.3.488-497

ABSTRACT

This study aimed to determine the quality of analytical thinking test instruments and the level of students' abilities through Rasch model analysis. This research used descriptive method with quantitative approach. The sample used amounted to 246 students. The instrument used was an analytical thinking test instrument totaling 9 items in the form of descriptions. Data collection was carried out by testing the test instrument. Data analysis was carried out by Rasch model analysis. The results showed that all items were fit. The level of item difficulty varies consisting of 1 very difficult question, 4 difficult questions, 3 easy questions, and 1 very easy question. Cronbach's alpha value of 0.80 is categorized as good, person reliability value of 0.79 is categorized as sufficient, and item reliability value of 0.98 is categorized as excellent. The level of item difficulty consists of 1 very difficult question, 4 difficult questions, 3 easy questions, and 1 very easy question. Person separation value 1.92 (H = 2.89, rounded to 3) and item separation 6.26 (H = 8.68 rounded to 9). Person measure analysis obtained 114 high ability (46%), 27 medium ability (11%), and 105 low ability (43%). Person fit analysis obtained 178 fits (72%), 2 misfits (1%), and 66 outliers (27%).

Keywords: *Analytical Thinking; Rasch; Test Instrument.*

INTRODUCTION

Education in the 21st century has a crucial role in facing the rapid development in the current era of globalization (Hanipah, 2023). Problems that develop in the era of globalization are not just simple issues, but problems that involve many considerations to find solutions (Gulacar *et al.*, 202; Tipani *et al*,. 2019). This development has triggered the use of current issues in learning to help prepare students for the future (Septiningrum & Fauziah, 2021).

Supported by BSKAP (Badan Standar, Kurikulum, and Asesmen Pendidikan) issued the Regulation of the Minister of Education and Culture of the Republic of Indonesia, Number 36 of 2022 regarding the independent curriculum at the end of phase E, emphasizing that learners must be responsive to global issues and active in solving problems. This means that learners learners are required to be able to understand information and create solutions to problems based on local, national, or global issues. In line with research (Sugiarto & Farid, 2023; Rahmafitri *et al.*, 2024; Widiyono & Millati, 2021) which reveals that through the independent curriculum learners are encouraged to become independent, emphasize critical thinking skills, behave well, innovate, collaborate, appreciate global diversity, and become autonomous individuals, as well as improve their skills to analyze and solve everyday problems.

Someone who thinks at a higher level can affect learning ability, speed and effectiveness of learning, so it is very important in the education process, and they will be better at making decisions, making judgments, and solving problems effectively (Saputra & Sudrajat, 2024). The basic ability and one of the main abilities that must be developed in higher-level thinking is the skill of analyzing (Irawati *et al.*, 2018). Then in line with the statement (Mahyastuti *et al.*, 2021), that analytical thinking is one of the

important skills that must be mastered by students in order to improve students' higher-level thinking skills, so that they can develop their own abilities. Supported by the opinion (Ramos *et al.,* 2018), that analytical thinking skills in the cognitive domain are included in the revised Bloom taxonomy at level C4 and are part of higher order thinking. Therefore, in learning analytical thinking skills need to be mastered by students (Yulina *et al.,* 2019; Yarmalinda, 2020)

The results of research conducted (Ilma *et al.,* 2017; Setiawaty *et al.,* 2019) state that overall students' analytical thinking skills are still in the low category, which is 40%. Only a few students on certain questions showed moderate criteria (Anselmus *et al.,* 2021). Then research Kiong *et al.,* (2012), found that analytical thinking skills have the lowest level of the five elements in Bloom's taxonomic framework. The test instruments used by teachers are also not effective because they have not gone through an adequate validation, reliability, and trial process, so the quality of the test items used to measure and evaluate the competence of students is unknown. Of the several test questions developed, only 1-2 questions lead to analytical thinking skills (C4) (Kusuma *et al* (2021).

One of the efforts to develop students' analytical thinking skills is to familiarize them with analytical thinking and measure the extent of students' skills (Saputra & Sudrajat, 2024; Hetarion *et al.,* 2020). Test questions that are suitable for training students' thinking skills are description test questions, because through description tests can evaluate thinking skills in more depth, and make students free to develop arguments or solutions to problems based on previously acquired knowledge and have the potential to support the development of students' thinking skills (Andriani & Lume, 2023; Blegur *et al.,* 2023; Febriano *et al.,* 2021; Miller, 2003; Rusmayani, 2020).

A good test instrument or measuring instrument is a valid and reliable measuring instrument, so that it can provide information about the ability of students appropriately, and produce credible data to be used as a reference in making policies or decisions (Ramadhan *et al.,* 2024). Validity and reliability of test instruments are the main requirements that must be met by a measurement instrument. This is because if the test used is not reliable or invalid, it will provide less careful information about the ability of a particular individual and actually produce biased conclusions (Laksono *et al.,* 2017). In addition, an analysis is carried out on the level of difficulty of the test, and differentiating power (Yusup, 2018; Dewi *et al.,* 2019).

The test instruments developed in this study were analyzed using item response theory, namely with the Rasch model analysis. The Rasch model is one of the analysis methods used to determine the feasibility of an instrument with the help of the Winsteps application. The Rasch model has the advantage of being able to describe the relationship between subjects and test items which makes the measurement results precise and more objective (Sumintono & Widhiarso, 2015). The advantages of the Rasch model analysis are that it is able to identify wrong answers, identify inappropriate assessments, and can predict missing data based on systematic response patterns (Hamdu *et al.,* 2020). The Rasch model also does not depend on the sample used, can sort in a structured manner from the most difficult questions to the easiest questions and can sort test takers with high to low abilities(Untary & Risdianto, 2020).

Based on the problems that have been described, researchers are interested in analyzing the feasibility of analytical thinking test instruments on biological technology innovation material using Rasch model analysis. This study aims to determine the quality of analytical thinking test instruments and the level of students' abilities through Rasch model analysis.

RESEARCH METHOD

The research method used in this study is a descriptive method with a quantitative approach. This research focuses on analyzing the quality of thinking test instruments analytical thinking test instrument and analyzing the level of students' abilities using Rasch model analysis.

Time and Location of Research

The research was conducted on August 19-29 and was conducted in 3 different schools in Tebas Sub-district, namely, SMA Negeri 1 Tebas, SMA Negeri 2 Tebas, and SMA Negeri 3 Tebas.

Population and Sample

The total population in this study was 638 people (based on observation data in each public high school in Tebas District). Based on the Slovin formula, a sample of 246 people was obtained and each school had 82 students who were sampled.

Data Collection Technique

Data collection techniques in this study used interview guidelines, validation sheets and test instruments that had been developed. The test instrument in this study is a question sheet on biological technology innovation material made by researchers in the form of descriptions and contains 9 questions.

Data Analysis Technique

Data analysis techniques are used to answer research questions that have been formulated. In this study, the analysis was carried out on test instruments that would be validated by experts. If the experts assess the instrument as feasible, then the instrument will be tested on students according to the predetermined sample size. Furthermore, the trial results will be analyzed using the Rasch model with the help of Winstep software (Alfarisa & Purnama, 2020). The items analyzed using the Rasch model are item fit, reliability, difficulty level, and differentiating power.

Table 1. Item Fit and Person Fit Value Criteria

Criteria	Value
Outfit Mean Square (MNSQ)	0,5 < MNSQ < 1,5
Outfit Z-standard (ZSTD)	-2,0 < ZSTD < +2,0
Point Measure Correlation (Pt Mean Cor)	0,4 < *Pt Measure Cor* < 0,85

(Sumintono & Widhiarso, 2015)

Table 2. Reliability of Items

Statistics	Index Value	Interpretation
Person and *Item Reliability*	< 0, 67	Weak
	0,67 - 0,80	Fair
	0,81-0,90	Good
	0,91-0,94	Very good
	> 0,94	Special
Alpha Cronbach	< 0,5	Bad
	0,5 - 0,6	Bad
	0,6-0,7	Fair
	0,7-0,8	Good
	> 0,8	Very good

(Sumintono & Widhiarso, 2015)

Table 3. Item Difficulty Level (Item Measure)

Indikator	Description
0,0 *logit* > 1SD	Very difficult question
0,0 *logit* + 1SD	Difficult question
0,0 *logit* - 1SD	Easy question
0,0 *logit* < 1SD	Very easy question

(Sumintono & Widhiarso, 2015)

Then analyze the differentiating power of the question. Question discriminating power is the ability of the question to distinguish students who are able to answer questions and are unable to answer questions. The question differentiator equation is as follows.

$$H = \frac{[(4\ \mathrm{X\ SEPARATION}) + 1]}{3}$$

The greater the separation value, the better the quality of the instrument in terms of overall respondents and items (Sumintono & Widhiarso, 2015).

The grouping of students' ability levels in the Rasch model uses the standard deviation (SD) and the average logit value (MEAN) generated from the Person Measure output (Lestari *et al.*, 2023; Tyas *et al.*, 2020; Sumintono & Widhiarso, 2015).

RESULTS AND DISCUSSION

The analytical thinking test instrument analyzed consisted of 9 items in the form of descriptions. The answers from the instrument trial on 246 samples were corrected using the scoring rubric that had been designed to obtain raw data. The raw data was then analyzed using the Rasch model with the help of Winsteps software version 3.73. In the Rasch model, the analysis includes item fit, reliability, item difficulty (item measure), distinguishing power (separation), the level of ability of students (person measure), and the suitability of students' answers (person fit), which is explained as follows.

Item Fit Analysis

Items can be categorized as fit if they meet one of the three criteria used. However, if there are items that do not meet all three criteria, then the items have poor quality so they need to be revised, replaced, or discarded (Sumintono & Widhiarso, 2015).

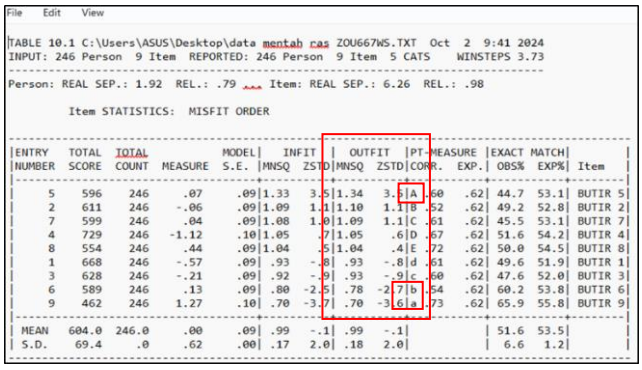


TABLE 10.1 C:\Users\ASUS\Desktop\data mentah cas ZOU667WS.TXT Oct 2 9:41 2024
INPUT: 246 Person 9 Item REPORTED: 246 Person 9 Item 5 CATS WINSTEPS 3.73

Person: REAL SEP.: 1.92 REL.: .79 ... Item: REAL SEP.: 6.26 REL.: .98

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	596	246	.07	.09	1.33	3.5	1.34	3.5	A .60	.62	44.7	53.1	BUTIR 5
2	611	246	-.06	.09	1.09	1.1	1.10	1.1	B .52	.62	49.2	52.8	BUTIR 2
7	599	246	.04	.09	1.08	1.0	1.09	1.1	C .61	.62	45.5	53.1	BUTIR 7
4	729	246	-1.12	.10	1.05	.7	1.05	.6	D .67	.62	51.6	54.2	BUTIR 4
8	554	246	.44	.09	1.04	.5	1.04	.4	E .72	.62	50.0	54.5	BUTIR 8
1	668	246	-.57	.09	.93	-.8	.93	-.8	d .61	.62	49.6	51.9	BUTIR 1
3	628	246	-.21	.09	.92	-.9	.93	-.9	c .60	.62	47.6	52.0	BUTIR 3
6	589	246	.13	.09	.80	-2.5	.78	-2.7	b .54	.62	60.2	53.8	BUTIR 6
9	462	246	1.27	.10	.70	-3.7	.70	-3.6	a .73	.62	65.9	55.8	BUTIR 9
MEAN	604.0	246.0	.00	.09	.99	-.1	.99	-.1			51.6	53.5	
S.D.	69.4	.0	.62	.00	.17	2.0	.18	2.0			6.6	1.2	

Figure 1. Output Tables Item Fit

The results of item fit testing for each item using the Rasch approach can be seen in Table 5.

Table 5. Interpretation of Fit Item Analysis

No *Item*	Outfit		Pt. Measure Corr	Description
	MNSQ	ZSTD		
1	0.93	-0.8	0.61	Accepted
2	1.10	1.1	0.52	Accepted
3	0.93	-0.9	0.60	Accepted
4	1.05	0.6	0.67	Accepted
5	1.34	3.5	0.60	Accepted
6	0.78	-2.7	0.54	Accepted
7	1.09	1.1	0.61	Accepted
8	1.04	0.4	0.72	Accepted
9	0.70	-3.6	0.73	Accepted

Based on the results of the analysis of the analytical thinking test instrument using the Winsteps program, it is known that of the 9 items of items developed, all are declared acceptable. From the Item Fit Order analysis, 3 items werefoundwhose ZSTD values did not meet the criteria, namely question number 5 (ZSTD = 3.5), 6 (ZSTD = -2.7) and question number 9 (ZSTD = -3.6). However, the three items were retained because they only did not meet one criterion, namely outfit ZSTD, while the outfit MNSQ and Pt Measure Corr values on the three items still met the criteria (fit).

Therefore, question numbers 5, 6, and 9 were retained. The question items that meet the three accepted criteria are numbers 1, 2, 3, 4, 7, and 8.

Item Measure Analysis

Analysis of the level of difficulty of the items was carried out on 9 description test questions. The purpose of this analysis is to identify the grouping of items classified as very difficult, difficult, easy, and very easy, so that researchers can ascertain whether the distribution of questions is proportional and includes various levels of difficulty. A good question is a question that has a balance and varies in difficulty, not too easy and not too difficult (Rifana *et al.*, 2024). Determination of the level of difficulty is done by comparing the logit measure value on each item and the standard deviation (SD) value.

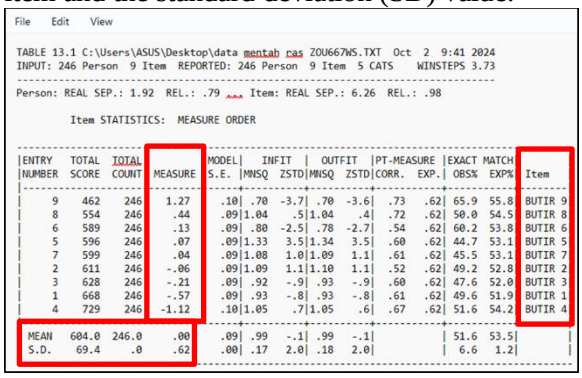


TABLE 13.1 C:\Users\ASUS\Desktop\data mentah cas ZOU667WS.TXT Oct 2 9:41 2024
INPUT: 246 Person 9 Item REPORTED: 246 Person 9 Item 5 CATS WINSTEPS 3.73

Person: REAL SEP.: 1.92 REL.: .79 ... Item: REAL SEP.: 6.26 REL.: .98

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
9	462	246	1.27	.10	.70	-3.7	.70	-3.6	.73	.62	65.9	55.8	BUTIR 9
8	554	246	.44	.09	1.04	.5	1.04	.4	.72	.62	50.0	54.5	BUTIR 8
6	589	246	.13	.09	.80	-2.5	.78	-2.7	.54	.62	60.2	53.8	BUTIR 6
5	596	246	.07	.09	1.33	3.5	1.34	3.5	.60	.62	44.7	53.1	BUTIR 5
7	599	246	.04	.09	1.08	1.0	1.09	1.1	.61	.62	45.5	53.1	BUTIR 7
2	611	246	-.06	.09	1.09	1.1	1.10	1.1	.52	.62	49.2	52.8	BUTIR 2
3	628	246	-.21	.09	.92	-.9	.93	-.9	.60	.62	47.6	52.0	BUTIR 3
1	668	246	-.57	.09	.93	-.8	.93	-.8	.61	.62	49.6	51.9	BUTIR 1
4	729	246	-1.12	.10	1.05	-.7	1.05	.6	.67	.62	51.6	54.2	BUTIR 4
MEAN	604.0	246.0	.00	.09	.99	-.1	.99	-.1			51.6	53.5	
S.D.	69.4	.0	.62	.00	.17	2.0	.18	2.0			6.6	1.2	

Figure 2. Output Tables Item Measure

Based on Figure 2, the standard deviation (SD) value from the instrument trial results is 0.62. The interpretation of the difficulty levels of the 9 items is presented in the following table.

Table 6. The Interpretation Item Measure

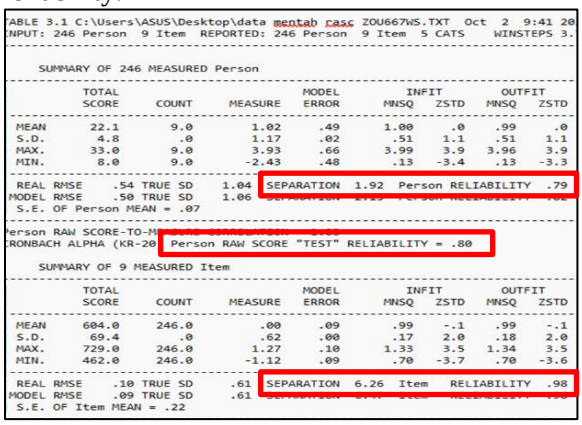
Analytical Thinking Indicator	Question No.	*Measue Logit*	Description
Organizing	1	-0.57	Easy
	6	0.13	Difficult
	9	1.27	Very difficult
Attributing	2	-0.06	Easy
	5	0.07	Difficult
	8	0.44	Difficult
Differentiating	3	-0.21	Easy
	4	-1.12	Very Easy
	7	0.04	Difficult

Based on the table, it is known that there is 1 question (question no. 9) which is included in the very difficult category, 4 questions (5, 6, 7, 8)

are in the difficult category, 3 questions (1, 2, 3) are in the easy category, and 1 question (4) is in the very easy category. The results of this analysis indicate that the analytical thinking test instrument has a good level of difficulty. According to (Hambleton & Swaminathan (2017) the level of test difficulty can be said to be good if the test has a varying level of difficulty. In line with the views Ishak (2019) and Rusiyah *et al.*, (2020), a good test item is one that has a proportional level of difficulty. Then supported by Arifin (2017), opinion, the ideal instrument must include a balanced proportion between these difficulty levels. Therefore, it is important to maintain this balance in the preparation of questions (Fiska *et al.*, 2021; Halik *et al.*, 2019; Rahmaini & Taufiq, 2018).

Summary Statistics Analysis Results

The results of the analysis of the summary statistics output are used to see the value of reliability and distinguishing power (separation). Reliability is used to determine the consistency of measurement results, both in repetition of items against other samples (person), as well as measurement of samples (person) against other conditions (Aryadoust *et al.*, 2021). Reliability in Rasch modeling is shown through Cronbach's alpha value, person reliability, and item reliability.



```
TABLE 3.1 C:\Users\ASUS\Desktop\data mentah rasc ZOU667WS.TXT  Oct  2  9:41 20
INPUT: 246 Person  9 Item  REPORTED: 246 Person  9 Item  5 CATS    WINSTEPS 3.

    SUMMARY OF 246 MEASURED Person

          TOTAL                          MODEL        INFIT        OUTFIT
          SCORE     COUNT     MEASURE    ERROR     MNSQ   ZSTD   MNSQ   ZSTD
 MEAN      22.1       9.0        1.02      .49     1.00     .0    .99     .0
 S.D.       4.8        .0        1.17      .02      .51    1.1    .51    1.1
 MAX.      33.0       9.0        3.93      .66     3.99    3.9   3.96    3.9
 MIN.       8.0       9.0       -2.43      .48      .13   -3.4    .13   -3.3

 REAL RMSE    .54 TRUE SD    1.04  SEPARATION  1.92  Person RELIABILITY  .79
MODEL RMSE    .50 TRUE SD    1.06  SEPARATION  2.13  Person RELIABILITY  .82
 S.E. OF Person MEAN = .07

Person RAW SCORE-TO-M
CRONBACH ALPHA (KR-20  Person RAW SCORE "TEST" RELIABILITY = .80

    SUMMARY OF 9 MEASURED Item

          TOTAL                          MODEL        INFIT        OUTFIT
          SCORE     COUNT     MEASURE    ERROR     MNSQ   ZSTD   MNSQ   ZSTD
 MEAN     604.0     246.0         .00      .09      .99    -.1    .99    -.1
 S.D.      69.4        .0         .62      .00      .17    2.0    .18    2.0
 MAX.     729.0     246.0        1.27      .10     1.33    3.5   1.34    3.5
 MIN.     462.0     246.0       -1.12      .09      .70   -3.7    .70   -3.6

 REAL RMSE    .10 TRUE SD     .61  SEPARATION  6.26  Item   RELIABILITY  .98
MODEL RMSE    .09 TRUE SD     .61  SEPARATION  6.77  Item   RELIABILITY  .98
 S.E. OF Item MEAN = .22
```

Figure 3. Output Tables Summary Statistic

Based on Figure 3, the values obtained on Cronbach's alpha, person reliability, and item reliability show adequate results. The Cronbach alpha value obtained of 0.80 is in the range of 0.7-0.8 which is included in the good category, meaning that the interaction between students (person) and items (item) as a whole is good. The person reliability value obtained of 0.79 is in the range of 0.67-0.80 which is included in the sufficient category, meaning that the consistency of students' answers is sufficient. While the item reliability value of 0.98 indicates that the quality of the items is excellent.

Differentiation analysis aims to assess the ability of items to distinguish students who master the material well and students who do not master the material well. In line with the statements of Uno & Koni (2012) and Rusmayani (2020), the analysis of differentiating power is intended to examine the ability of questions to distinguish between students who have high achievement and low achievement.

Based on the results of the analysis, it is known that the person separation value is 1.92, then the value of the differentiating power is H = 2.89 rounded to 3, indicating that the respondent group can be divided into three groups. While the item separation is 6.26, the value of the differentiating power is H = 8.68 rounded to 9, so there are nine groups of items. According to Linacre, (2010), that the separation index that exceeds 2 can be said to have a good value. In line with the opinion, saying that the greater the separation value, the better the quality of the instrument in terms of overall respondents and items, because it can identify respondent groups and item groups.

Person Measure Analysis

The analysis of the level of individual ability or person measure aims to analyze the level of ability of individual students in solving problems. (Rohmah *et al.*, 2022; Apipatunnisa *et al.*, 2022)

Table 7. Person Measure Analysis Results

Number of Learners	Percentage	Category
114	46%	High
27	11%	Medium
105	43%	Low

Based on Table 7, students with high ability levels have a percentage of 46% with 114 students. Learners with a medium level of ability have a percentage of 11% with 27 students, while students with a low level of ability have a percentage of 43% with 105 students.

Learners with person code PA069 have the highest logit score with a value of 3.93 this

learner is almost close to a perfect score of 36 and a raw score of 33 points. The highest logit value obtained by learner PA069 is due to his ability to answer very difficult items correctly. Conversely, learners with the lowest level of ability are owned by learners with the person code LA082 with a logit value of (-2.43) who make many mistakes in filling in answers or cannot complete the entire item and only get a total raw score of 8 points. The lowest logit value obtained by LA082 learners is due to the low ability to answer question items, so that the points obtained on each item are low, coupled with 2 questions not answered. In line with research (Rohmah *et al.,* 2022) which states that a high logit value indicates that the ability of students to solve or answer items correctly is also high.

Person Fit Analysis

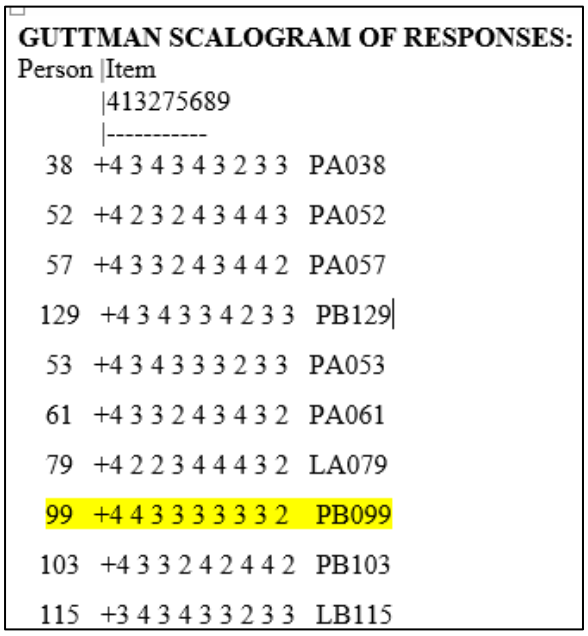
The level of individual fit (Person fit) with the Rasch model can identify individuals with inappropriate response patterns. Inappropriate response patterns mean thatthereis a mismatch between the learners' abilities and the answer patterns given in answering the items (Kurniawan & Andriyani, 2018).

Table 8. Person Fit Analysis Results

Category	Amount	Percentage (%)
Fit	178	72%
Mifit	2	1%
Outlier	66	27%

Based on Table 8, it is known that students who are included in the fit category have a percentage of 72% with a total of 178 students. Learners who fall into the category of not fit have a percentage of 1% with a total of 2 people, while students who fall into the outlier category have a percentage of 27% with a total of 66 students. Learners who are in the unfit category are thought to provide answer responses that are not in accordance with the ideal model (Lestari *et al.,* 2023). Rasch model theory has characteristics that rank items from the easiest to the most difficult, as well as the ability of students from the highest to the lowest, which is displayed based on a Guttman matrix or scalogram (Subando & Wahid, 2022).

Further information on the causes of learners' misfit can be seen from the output scalograms shown in Figure 4.



GUTTMAN SCALOGRAM OF RESPONSES:

```
Person |Item
       |413275689
       |----------
  38  +4 3 4 3 4 3 2 3 3  PA038
  52  +4 2 3 2 4 3 4 4 3  PA052
  57  +4 3 3 2 4 3 4 4 2  PA057
 129  +4 3 4 3 3 4 2 3 3  PB129
  53  +4 3 4 3 3 3 2 3 3  PA053
  61  +4 3 3 2 4 3 4 3 2  PA061
  79  +4 2 2 3 4 4 4 3 2  LA079
  99  +4 4 3 3 3 3 3 3 2  PB099
 103  +4 3 3 2 4 2 4 4 2  PB103
 115  +3 4 3 4 3 3 2 3 3  LB115
```

Figure 4. Output scalograms

Learners with the person code PB099 were identified by the Rasch model as a person not fit because they showed a creative responding pattern. PB099 tended to answer easier questions inconsistently, but managed to answer more difficult questions in an unusual way. This is in line with the definition of creative responding described by (Meijer *et al.,* 1996) and (Karabatsos, 2003) where high ability participants can give unexpected responses to easier problems due to creative interpretation. This causes learners LC168 and PB099 to fall into the category of learners who do not fit.

In the results of the person fit analysis, outlier data were detected, which can be seen from the disappearance of data for 77 students in the person fit output. This outlier data is data that is significantly different from other data (Febriyansyah *et al.,* 2020). The existence of outlier data can be caused by errors in entering data, measurement errors, analysis, or other errors. The loss of person data is caused by significant differences in data patterns with other data or incompatibility with the Rasch model, so it cannot be read in the analysis process. Sari & Mahmudi (2024) also reported that the results of the person fit analysis showed that there were 110 outlier data, equivalent to 50% of the total 216 persons, which disturbed the data. However, the presence of person data categorized as

outliers does not affect the quality of the items. This is in line with the statement (Widodo & Chotimah, 2023) which states that in the Rasch model analysis, the assessment of person ability is not affected by the item, and the quality of the item is not affected by person ability.

CONCLUSION

Based on the results that have been described, it can be concluded that the test instrument developed is able to identify the quality of the analytical thinking test instrument and the level of students' abilities through Rasch model analysis. The analysis results show that the instrument developed has good quality with items that meet the criteria of validity and reliability. Learners' ability levels are diversely distributed, reflecting the variation in ability among them. The Rasch model proved effective in providing a detailed description of item characteristics and learner ability profiles, so that it can be used as a credible evaluation tool to measure analytical thinking skills.

DAFTAR PUSTAKA

Alfarisa, F., & Purnama, D. N. (2019). Analisis butir soal ulangan akhir semester mata pelajaran ekonomi SMA menggunakan Rasch model. *Jurnal Pendidikan Ekonomi Undiksha*, 11(2), 366-374.

Andriani, R., & Lume, L. (2023). Pengembangan Instrumen Penilaian Untuk Mengukur Kemampuan Berpikir Analitis Pada Mahasiswa Fakultas Pertanian. *Jurnal Jendela Pendidikan*, *3*(03), 304–311. https://doi.org/10.57008/jjp.v3i03.555

Anselmus, A., Risalah, D., & Sandie, S. (2021). Pengembangan Instrumen Tes HOTS dalam Mengukur Kemampuan Berpikir Analitis Siswa pada Materi Program Linear di Kelas XI SMA Sungai Kehidupan. *JURING (Journal for Research in Mathematics Learning)*, *4*(4), 371. https://doi.org/10.24014/juring.v4i4.14295

Apipatunnisa, I., Hamdu, G., & Giyartini, R. (2022). Eksplorasi kemampuan literasi dan numerasi siswa sekolah dasar dengan pemodelan rasch. *COLLASE (Creative of Learning Students Elementary Education)*, 5(4), 668-680.

Arifin, Z. (2017). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya, hlm. 23-27

Aryadoust, Vahid, Li Ying Ng, and H. S. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40.

Blegur, J., Rajagukguk, C. P. M., Sjioen, A. E., & Souisa, M. (2023). Innovation of Analytical Thinking Skills Instrument for Throwing and Catching Game Activities for Elementary School Students. *International Journal of Instruction*, *16*(1), 723–740. https://doi.org/10.29333/iji.2023.16140a

Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). Analisis Tingkat Kesukaran Dan Daya Pembeda Soal Olimpiade Matematika (Omi) Tingkat Smp Tahun 2018. *Transformasi : Jurnal Pendidikan Matematika Dan Matematika*, *3*(1), 15–26. https://doi.org/10.36526/tr.v3i1.388

Febriano, R., Tandililing, E., & Enawaty, E. (2021). Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Matematis Dengan Menggunakan Analisis Model Rasch Pada Siswa SMP. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, *10*(9), 1–12.

Febriyansyah, I. A., Fadhil, R., & Zulfahrizal, Z. (2020). Pengembangan Metode Klasifikasi Biji Kopi Beras Arabika Gayo dan Robusta Gayo dengan Metode PCA (Principal Component Analysis) Berdasarkan Pengolahannya. *Jurnal Ilmiah Mahasiswa Pertanian*, *4*(4), 472–481. https://doi.org/10.17969/jimfp.v4i4.12776

Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, *4*(1), 65–76. https://doi.org/10.21107/nser.v4i1.8133

Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of ThirtySix Person-Fit Statistics. *Applied Measurement in Education*, *16*(4), 277–298.

Gulacar, O., Zowada, C., Burke, S., Nabavizadeh, A., Bernardo, A., & Eilks, I. (2020). Integration of a sustainability-

oriented socio-scientific issue into the general chemistry curriculum: Examining the effects on student motivation and self-efficacy. *Sustainable Chemistry and Pharmacy*, *15*, 100232. https://doi.org/https://doi.org/10.1016/j.scp.2020.100232

Halik, A. S., Mania, S., & Nur, F. (2019). Analisis Butir Soal Ujian Akhir Sekolah (Uas) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 Smp Negeri 36 Makassar. *Al Asma : Journal of Islamic Education*, *1*(1), 11. https://doi.org/10.24252/asma.v1i1.11249

Hambleton, Ronald K., and H. S. (2013). *Item response theory: Principles and applications*. Dordrecht: Springer Science & Business Media.

Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). *Item*s quality analysis using rasch model to measure elementary school students' critical thinking skill on stem learning. *JPI (Jurnal Pendidikan Indonesia)*, *9*(1), 61–74.

Hanipah Sri. (2023). Analisis Kurikulum Merdeka Belajar Dalam Memfasilitasi Pembelajaran Abad Ke-21 Pada Siswa Menengah Atas. *Jurnal Bintang Pendidikan Indonesia*, *1*(2), 264–275. https://doi.org/10.55606/jubpi.v1i2.1860

Hetarion, Bety DS, Yosina Hetarion, and V. M. (2020). Implementasi pendidikan karakter berbasis kearifan lokal cuci negeri dalam pembelajaran IPS. *JTP-Jurnal Teknologi Pendidikan*, *22*(1), 1–12.

Ilma, R., Hamdani, A. S., & Lailiyah, S. (2017). *Ilma 2017 Penelitian Kemampuan Analitis Rendah*. *2*(1), 1–14.

Irawati, T. N. and M. M. (2018). Pengembangan instrumen kemampuan berpikir analisis siswa smp dalam menyelesaikan soal pemecahan masalah matematika. *Kadikma*, *9*(2), 1–11.

Kiong Tze Tee, Jailani, M. Y., Razali, H., Yee, M. H., Atan, H. H., & Mimi, M. M. (2012). Thinking Skills for Secondary School Students in Malaysia. *Journal of Research, Policy & Practice of Teachers & Teacher Education*, *2*(2), 12–23.

Kurniawan, & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, *6*(1), 34–39.

Kusuma, A. S. K. M. D. U. R. (2021). The Development of Higher Order-Thinking Skills (HOTS) Instrument Assessment in Physics Study. *Journal of Physics: Conference Series*, *1899*(1), 1–7. https://doi.org/10.1088/1742-6596/1899/1/012140

Laksono, E. W., Rohaeti, E., Suyanta, S., & Irwanto, I. (2017). The Evaluation Instrument of Analytical Thinking and Science Process Skill in Chemistry Subject. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, 1(1), 100-110.

Lestari, T. D. H. R. S. (2023). Analisis Soal Literasi Numerasi Menggunakan Pemodelan Rasch Konteks Pemanasan Global Berbasis ESD Untuk Sekolah Dasar Tiara. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *8*(2), 2489-2503.

Linacre, J. M. (2010). *A user's guide to Winsteps Ministep. Rasch-model computer programs*.

Mahyastuti, I., Dwiyana, D., & Hidayanto, E. (2021). Kemampuan Berpikir Analitis Siswa dalam Memecahkan Masalah Matematis. *Jurnal Pendidikan Matematika Dan Sains*, *8*(1), 1–6. https://doi.org/10.21831/jpms.v8i1.19644

Meijer, R. R., Muijtjens, A. M. M., & van der Vlueten, C. P. M. (1996). Nonparametric Person-Fit Research: Some Theoretical issues an Empirical Example. *Applied Measurement in Education*, *9*(1), 77–89.

Miller, T. (2003). Essay assessment with latent semantic analysis. *Journal of Educational Computing Research*, *29*(4), 495–512.

Rahmafitri, F., Deswita Sekolah Menengah Atas Negeri, E., & Trisoni, R. (2024). Analisis Kebijakan Kurikulum Merdeka dan Implikasinya Terhadap Kualitas Pendidikan. *Dirasah : Jurnal Studi Ilmu Dan Manajemen Pendidikan Islam*, *7*(1), 45–55. https://ejournal.iaifa.ac.id/index.php/dirasah/article/view/1050

Rahmaini, A., & Taufiq, A. N. (2018). Analisis Butir Soal Pendidikan Agama Islam Di SMK N 1 Sedayu Tahun Ajaran 2017/2018 (Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor pada Soal Pilihan Ganda Kelas XI). *Jurnal*

Mudarrisuna, *8*(1), 1–24.

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975. https://doi.org/10.31004/joe.v6i2.4885

Ramos, Dolipas, & V. (2018). Higher Order Thinking Skillss and Academic Performance in Physics of College Students: A Regression Analysis. *International Journal of Innovative Interdisciplinary Research, Issue 4, p: 48-60.*, *4*, 48–60.

Rifana, F., Ramadhan, S., & Putro, K. Z. (2024). Analisis Butir Soal Ulangan Harian Siswa Mata Pelajaran PPKn Kelas IV Menggunakan Rach Model di Madrasah Ibtidaiyah Negeri. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *7*(1), 99–110. https://doi.org/10.54069/attadrib.v7i1.424

Rohmah, F. N., Susilaningsih, E., Haryani, S., & Kasmui. (2022). Desain Asesmen Kompetensi Minimum Literasi Membaca Bermuatan High Order Thinking Skills untuk Menganalisis Kompetensi Minimum Siswa Materi Asam-Basa. *Chemistry in Education*, *11*(2), 117–125. http://journal.unnes.ac.id/sju/index.php/chemined

Rusiyah, R., Eraku, S. S., & Supadmi, S. (2020). Analisis Soal Ujian Akhir Semester Mata Pelajaran Geografi Dengan Menggunakan Pemodelan Rasch. *Jurnal Swarnabhumi : Jurnal Geografi Dan Pembelajaran Geografi*, *5*(1), 11. https://doi.org/10.31851/swarnabhumi.v5i1.4136

Rusmayani. (2020). Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam Di SMP Bintang Persada Tabanan-Bali. *Widya Balina*, *5*(1), 41–49.

Saputra, O. A., & Sudrajat, A. (2024). Analisis Instrumen Penilaian Hasil Belajar Kimia untuk Mengukur High Order Thinking Skill (Hots) Semester Genap Kelas X Berdasarkan Permodelan Rasch. *Didaktika: Jurnal Kependidikan*, *13*(2), 1967–1984.

Sari, E. D. K., & Mahmudi, I. (2024). Analisis pemodelan Rasch pada assessment pendidikan (analisis dengan menggunakan aplikasi Winstep). Jawa Tengah: Pena Persada Kerta Utama.

Septiningrum Wahyu, A., & Fauziah, H. N. (2021). Investigasi Kebiasaan Berpikir Analitis Melalui Pengembangan Instrumen Penilaian Berbasis Isu Sosiosaintifik. *Jurnal Tadris IPA Indonesia*, *1*(3), 269–281. https://doi.org/10.21154/jtii.v1i3.156

Setiawaty, Bintang Titik, Widha Sunarno, and S. S. (2019). Profil kemampuan berpikir analisis siswa sekolah menengah pertama di Surakarta. *Prosiding SNPS (Seminar Nasional Pendidikan Sains)*, 234–238.

Sugiarto, & Farid, A. (2023). Literasi Digital Sebagai Jalan Penguatan Pendidikan Karakter Di Era Society 5.0. *Cetta: Jurnal Ilmu Pendidikan*, *6*(3), 580–597. https://doi.org/10.37329/cetta.v6i3.2603

Sumintono, B, & Widhiarso, W (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Trim Komunikata.

Tipani, A., Toto, Yulisma, L. (2019). Implementasi model PjBL berbasis STEM untuk meningkatkan penguasaan konsep dan kemampuan berpikir analitis siswa. *Bio Educatio : (The Journal of Science and Biology Education)*, *4*(2), 70–76. http://www.jurnal.unma.ac.id/index.php/BE/article/view/1700

Tyas, E. H., Hamdu, G., & Haki Pranata, O. (2020). Analisis Soal Pilihan Ganda dengan Menggunakan Pemodelan RASCH untuk Mengukur Kemampuan Siswa dalam Mengurutkan Bilangan Pecahan di Sekolah Dasar. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *7*(2), 1–12. https://doi.org/10.17509/pedadidaktika.v7i2.24773

Subando, J & Wahid, N. (2022). *Evaluasi Hasil Belajar Pendidikan Agama Islam*. Jawa Tengah: Lakeisha.

Uno, B. H dan Koni, S. (2012). *Assessment Pembelajaran*. Jakarta: Bumi Aksara.

Untary, H., & Risdianto, E. (2020). *Kusen, Analisis Data Penelitian Dengan Model Rasch Dan Winstep*. Bogor:

Widiyono, A., & Millati, I. (2021). The Role of Educational Technology in the Perspective of Independent Learning in Era 4.0. *Journal of Education and Teaching*, *2*(1), 1–9.

Widodo, W., & Chotimah, C. (2023). Adaptasi dan Analisis Psikometri Skala Kompetensi Multikultural Calon Guru Menggunakan Pemodelan Rasch. *Jurnal Pendidikan Dan Kebudayaan*, *8*(2), 153–172. https://doi.org/10.24832/jpnk.v8i2.4228

Yarmalinda, D, & S. (2020). Biolearning Journal. *Biolearning Journal*, *7*(1), 39–43.

Yulina, I. K., Permanasari, A., Hernani, H., & Setiawan, W. (2019). Analytical thinking skill profile and perception of pre service chemistry teachers in analytical chemistry learning. *Journal of Physics: Conference Series*, *1157*(4). https://doi.org/10.1088/1742-6596/1157/4/042046

Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan*, *7*(1). http://103.180.95.17/index.php/jtjik/article/view/2100