

An Analysis of Elementary School Thematic Question Points Using Iteman Software and Winstep Rasch Model

Difa Maulidya*

Madrasah Ibtidaiyah Teacher Education, UIN Sunan Kalijaga, Yogyakarta, Indonesia

Shaleh Shaleh

Madrasah Ibtidaiyah Teacher Education, UIN Sunan Kalijaga, Yogyakarta, Indonesia

Puspo Rohmi

Physics Education, UIN Sunan Kalijaga, Yogyakarta, Indonesia

*Correspondence author: 21204082040@student.uin-suka.ac.id

Abstract:

This study analyzed thematic question items in elementary schools using Iteman software and the Winsteps Rasch Model. The research employed a quantitative method with a descriptive approach, collecting data from 25 multiple-choice questions and 42 student respondents. Reliability scores were 0.641 (classical tests) and 0.68 (modern tests), falling below the desired alpha coefficient of 0.70. Two questions did not meet validity requirements based on MNSQ, ZSTD, and PT Measure. Difficulty levels revealed four very difficult, three too easy, and 18 difficult questions. Classical test analysis showed 10 poor, 12 sufficient, and three very poor discriminating power questions. Modern test analysis identified two students with poor differentiation. Overall, 23 out of 25 questions had well-functioning distractors. Implications include improving measurement quality, curriculum development, student understanding, and learning evaluation at the elementary level.

Keywords: Question item; Iteman; Rasch Model; Winstep

DOI: [10.18952/aladzkapgmi.v13i1.9415](https://doi.org/10.18952/aladzkapgmi.v13i1.9415)

Received: 29 Mei 2023 ; **Accepted:** 19 Juni 2023 ; **Published:** 31 Juli 2023

How to cite : Maulidya, D., Shaleh. S., Rohmi, P. (2023). An Analysis of Elementary School Thematic Question Points Using Iteman Software and Winstep Rasch Model, *Al-Adzka: Jurnal Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, 13(1), 30-51



This is an open access article under Creative Commons Attribution 4.0 International License

Introduction

Evaluating student learning is something that teachers must do, making it a very important task. Evaluation is a type of assessment that involves collecting, analyzing, and interpreting data to determine whether or not a student's learning objectives have been achieved (Magdalena, Fauzi, & Putri, 2020). Many ways to assess learning, changes in technology, and teaching methods have led to interest in different methods so guidelines are needed (Neto, Neto, & Furnham, 2022). Guidelines for assessing student learning outcomes are regulated in Article 3 of the Regulation of the Minister of Education and Culture Number 21 of 2022 for use in elementary, junior high, and secondary schools. Determining assessment objectives, selecting or creating assessment tools, conducting assessments, processing assessment findings, and reporting assessment results are some of the topics discussed (Heryanti, 2022).

A number of factors must be taken into account in order for the valuation of the item to be considered accurate. First, evaluation criteria can measure things accordingly. Secondly, it is important to pay attention to the consistency of the test as indicated by the dependency of the scoring item. Third, it is very important to assess the difficulty of the question by comparing the proportion of students who answered correctly against the entire student population. Fourth, it is important to focus on the differentiating power of the evaluation items, or their ability to distinguish between students who have understood the subject matter and those who have not. Fifth, it is important to consider the extent to which each evaluation item may hinder students' ability to think clearly when choosing solutions. Finally, when evaluating scoring items, it is important to determine how successful the question tricker is (Tarmizi, Setiono, Amaliyah, & Agrian, 2021).

Effective assessment tools are reliable and valid. Evaluation tools will allow teachers to examine the teaching strategies used and determine whether or not learning objectives are successfully achieved (Widiastuti, Putrayasa, & ..., 2022). The whole set of questions and each component of the assessment questions are examined. The degree of complexity of the object, the ability to distinguish, and deception are all considered in the analysis. To determine whether the questions are too easy or difficult for students to do, as well as to recognize the ability of the questions to distinguish between students who have mastered and those who have not, a question analysis is carried out. This indicates the need to look at items through the lens of validity, reliability, discriminating power, and difficulty (Santosa & Badawi, 2022).

One of the most well-known theories of modern analysis is the Rasch modeling. The ability to predict missing data is an advantage of Rasch modeling over other methods, particularly classical test theory. In research, Rasch's modeling will also produce accurate statistical analysis. By applying the Rasch Modeling method that uses equal distances to compare scores, information can be obtained about the test taker and the quality of the evaluation items answered. The disadvantage of Rasch's modeling is that mathematical equations in IRT are more difficult to understand than in CCT, thus requiring the use of computers (Harmuni, 2019).

The results of an initial interview with a grade VI thematic teacher at SD Mantuil 2 Kota Banjarmasin revealed that the results of student evaluations were still low. This is due to the teacher's insensitivity in looking at the test questions given to students. Teachers are not sure the questions they use in tests or assessments are acceptable to gauge students' abilities. As a result, teachers cannot determine whether each assessment item is adequate and good. This evaluation question is used to measure the level of difficulty, validity, dependability, discriminating power, and other factors. The Rasch Model in Winstep software and

Iteman software were used in this study to examine thematic problems in elementary schools. The purpose of this study is to assess the validity, reliability, level of difficulty, and differentiating power of thematic questions. The study also sought to pinpoint the most effective distractors in this question. An innovative aspect of the research is how the Iteman and Rasch software was used to conduct thematic problem analysis in elementary schools, which offers more thorough details about reliability, validity, difficulty, and discriminating power. The results of this study have important consequences for improving measurement standards, curriculum creation, student understanding, and assessment of learning at the primary school level. Findings from a previous study by Fitri al-farisa et al. entitled "Analysis of Class V Social Studies Questions Conducted by Teachers in Serang City Show Good Quality" was published. In this study, it was found that 90% of the discriminating questions were in a good category, 80% of the difficulty questions were in the medium category, and 90% of the deceivers worked well. In addition, the teacher's estimate of the dependence of the instrument, which has a coefficient of 0.75, is also good (Alfarisa, Chudari, & Robiansyah, 2019).

The previous research "Analysis of Physical Assessment Using the Rasch Model with the R Program" by Made Rai Suci Shanti et al. is relevant to this research because they both use the Rasch Model in their research. The results of 16 physics questions tested on 99 junior high school students and analyzed by the R program had an average difficulty level (brata2) of -0.331 and a standard deviation of 1 (Ayub et al., 2020).

Research Method

This research is quantitative research that uses descriptive data analysis methods. Without attempting to test a particular hypothesis, quantitative descriptive research is research that describes, analyzes, and explains a phenomenon using data (numbers) as it is (Sulistiyawati, Wahyudi, & Trinuryono, 2022). According to Sudjana and Ibrahim, descriptive research aims to characterize a phenomenon, event, or ongoing event (Jayusman & Shavab, 2020).

So that the purpose of this study is used in order to be able to witness carefully, evaluate, and provide a detailed description of the numerical object under study as it is, as well as describe the characteristics of elementary school thematic questions classically and modern. A total of 42 students from SDN Mantuil 2 in Banjarmasin City participated in this study. This study adopted a non-interactive methodology. In other words, research is carried out through document analysis. The midterm assessment (PTS) thematic question sheet is used as the basis for research instruments in the 2022–2023 school year at SDN Mantuil 2 Banjarmasin City. A multiple-choice exam with four possible answers was used to collect data for this investigation. The thematic teacher made a total of 25 exam questions. After student testing, Winstep and Iteman software were used to analyze the data using the Rasch model and classical theory, respectively.

Results and Discussion

Iteman is a computer program that functions to perform classical analysis on items, consisting of a number of programs issued by the Assessment Systems Corporation in 1982 and commonly called MicroCATn. The program is capable of several tasks, including (1) analyzing ASCII-formatted data files containing answers to questions, which can be entered manually or scanned by machines; (2) In this study, scores were calculated and data on the Likert scale and multiple-choice questions were analyzed using a sample of 30,000 students and 250 items. In addition, the study will evaluate the test, which has ten subtests (or scales), and offer details regarding the reliability, validity, and distinguishing power of each item,

as well as the level of difficulty and proportion of correct answers for each item (Nuryanti, Masykuri, & Susilowati, 2018). The Rasch model, often known as the 1PL (one logistics parameter) technique, was created by Georg Rasch as an analytical model for item response theory (IRT). Rasch modeling analysis can be used to describe item quality (Fauziana & Dessy Wulansari, 2021).

The two software tools listed above can evaluate questions teachers use to gauge students' understanding of what they have learned, regardless of whether a semester or an academic year has passed since the lesson was delivered. Because evaluation is a process of managing information from the results of interactions between teachers and students in the learning process to identify deficiencies and then make changes towards improvement, the results of information management are taken into consideration by decision-makers when developing further programs (Adji et al., 2018). Another evaluation theory in the book Educational Evaluation explains that evaluation is a process that involves collecting information or data about an object and is carried out systematically and continuously to assess the quality (value or meaning) of something to make decisions (Widodo, 2021).

Boone et al in Sumintono provide criteria for evaluating the quality of item fit in the Rasch analysis, referred to as item fit. These criteria include: (1) the mean square (MNSQ) value of Accepted Outfit should be between 0.5 to 1.5; (2) the Z-standard Accepted Clothing (ZSTD) value should be between -2.0 to +2.0; and (3) the Accepted Measure Correlation value (Pr Mean Corr) should be between 0.4 to 0.85 based on the Pt Measure Corr. In the analysis using the Rasch Model and Winsteps software, the mean square value should be 1.0 and the Z-standardized value should be 0.0. In addition, the analysis can also identify very easy or very difficult questions, as well as identify outliers in respondents (Nuryanti et al., 2018).

The resulting data will be broken down into several statistical summaries that will provide information on the reliability and validity of questions and respondents.

Reliability

An indicator of how trustworthy or reliable measurement equipment is called reliability. The consistency of the measuring instrument, or whether it maintains consistency after repeated measurements, can be assessed using reliability. If the measuring device produces the same finding after several measurements, it is considered reliable (Janna & Herianto, 2021). Nunnally states that the instrument is trustworthy because the value of the Cronbach Alpha reliability coefficient is higher than 0.70 ($r_i > 0.70$), while Streiner insists that the result should not be higher than 0.90 ($r_i < 0.90$). If Cronbach's Alpha reliability coefficient is less than 0.70 ($r_i < 0.70$), Tavakol and Dennick suggest changing or removing elements with poor correlation. The use of computer programs can help to determine the Cronbach Alpha reliability coefficient quickly and it is recommended to reduce the number of questions if the reliability coefficient is too high, although the questions have the same criteria but with different sentences (Febrianawati, 2018).

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First- Last	S-B Odd- Even
Scored items	0,641	2,234	0,469	0,439	0,422	0,638	0,611	0,593

Figure 1. Reliability of Iteman Analysis

In data analysis using Iteman and Winsteps software, it was found that the Cronbach Alpha reliability coefficients were 0.641 and 0.68, respectively. According

to general standards, the desired reliability coefficient is at least 0.70. Therefore, these results show that the level of reliability of the questions is still low. In this regard, Tavakol and Dennick recommend several actions that can be taken to improve the reliability of the problem. One action that can be taken is to revise or delete items with low correlation (Gadermann, Guhn, & Zumbo, 2019).

The process of revising or deleting items with low correlation aims to improve the reliability of measurement instruments. The correlation between the items in the measuring instrument should be high, indicating consistency in the measurements made. By deleting or revising items with low correlation, we can increase the level of reliability of the instrument.

SUMMARY OF 42 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	12.5	25.0	.01	.46	1.00	.0	1.05	.0
S.D.	3.7	.0	.81	.05	.17	1.0	.48	1.0
MAX.	21.0	25.0	1.95	.65	1.32	1.7	3.63	2.4
MIN.	3.0	25.0	-2.32	.44	.64	-2.6	.45	-2.2
REAL RMSE	.48	TRUE SD	.65	SEPARATION	1.36	Person RELIABILITY	.65	
MODEL RMSE	.47	TRUE SD	.66	SEPARATION	1.43	Person RELIABILITY	.67	
S.E. OF Person MEAN = .13								
erson RAW SCORE-TO-MEASURE CORRELATION = 1.00								
RONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .63								
SUMMARY OF 25 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	21.1	42.0	.00	.36	.99	.0	1.05	.1
S.D.	8.0	.0	.98	.04	.14	1.0	.43	1.2
MAX.	37.0	42.0	1.81	.50	1.38	2.5	2.96	3.4
MIN.	7.0	42.0	-2.21	.33	.84	-1.3	.68	-1.3
REAL RMSE	.37	TRUE SD	.91	SEPARATION	2.44	Item RELIABILITY	.86	
MODEL RMSE	.36	TRUE SD	.91	SEPARATION	2.52	Item RELIABILITY	.86	
S.E. OF Item MEAN = .20								
MEAN=.0000 USCALE=1.0000								

Figure 2. Winstep Analysis Reliability

The Rasch model can provide a comprehensive analysis of instrument quality, student responses, and the relationship between respondents and question items. Cronbach's Alpha values can indicate the reliability of items on an instrument, where values of 0.50 (less), 0.50-0.60 (less), 0.60-0.70 (sufficient), 0.70-0.80 (good), and >0.80 indicate that the items are reliable (excellent). According to Sumintono, dependency can be assessed using item and person reliability scores, and the following reliability criteria can be applied: A score of 0.67 is considered poor, followed by 0.67-0.8, 0.8-0.9, 0.91-0.94, and > 0.94 (Widhiarso & Sumintono, B., 2019). Analysis of the Rasch model in this study gave personal reliability results of 0.65 to 0.67, which can be considered sufficient. Personal reliability measures the consistency or reliability of individual responses to questions in a measurement instrument. Although personal reliability is in a fairly low range, it is acceptable to measure relative comparisons between individuals.

However, the reliability of the items in this study reached 0.86, which can be categorized as excellent. Item reliability measures the consistency or reliability of questions in a measurement instrument. A high item reliability value indicates that

the questions consistently measure the desired construct (Airasian & Cruikshank, 2018). In addition, Cronbach's alpha (KR-20) value of 0.63 was also obtained. Cronbach's alpha is one of the commonly used methods for measuring the internal reliability of a measurement instrument assuming that all items in the instrument measure the same construct. The resulting Cronbach's alpha value (0.63) indicates a sufficient level of reliability (Suskie, 2018).

Based on these results, it can be concluded that the reliability of the questions in this study is good enough and does not need to be replaced. This indicates that the measurement instruments used have shown consistency and reliability in measuring the constructs studied. Relationship with learning evaluation theory can be done by looking at the influence of the reliability of measurement instruments on learning evaluation results. Good reliability on measurement instruments ensures that the data obtained is accurate and reliable. It is important in learning evaluations to ensure that decisions or judgments made are based on valid and reliable information.

Validity

The degree of suitability (validity) of items is used to describe whether or not they function normally for measurement. Validity is a property that indicates the relationship between measurement and the intent or purpose of a standard or learning behavior (A. Suryadi, n.d.). Boone has established three criteria for assessing the suitability of inadequate items in Rasch's analysis: 1) the received MNSQ Outfit has a value between 0.5 to 1.5, 2) the accepted Z-Standard Outfit has a value between -2.0 to +2.0, and 3) the received PT-Measure Corr has a value between 0.4 to 0.85 (Yulianto & Widodo, 2020).

When an item in the Rasch model analysis does not meet the criteria of MNSQ (Mean Square Fit) and PT Measure Corr (Person Measure Correlation), but still has a ZSTD (Standardized Residual) value that meets the requirements, then the item requires further examination to determine whether it should be removed or modified. The ZSTD value is used to identify items that stand out, meaning that they have a response that does not match the common response patterns in measuring instruments (Engelhard Jr, 2018). In the context of learning evaluation, further examination of the item needs to be done to ensure the validity and reliability of the measurement. Items that do not meet the criteria of MNSQ and PT Measure Corr indicate a mismatch between student responses and the characteristics measured by those items. This may indicate a problem in the formulation or selection of items (DeMars, 2019).

Further examination may include qualitative analysis of the item, involving the use of experts or experts in the field of evaluation or the material measured by the item. Consideration may also be given to the context and objectives of the evaluation, as well as the characteristics of the population being measured.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBSN	EXP%		
24	7	42	1.81	.44	1.38	1.3	2.96	3.4	A	.30	.29	76.2	83.8	524
6	16	42	.55	.34	1.32	2.5	1.50	2.5	B	-.08	.33	57.3	67.2	56
17	15	42	.67	.34	.87	-1.0	1.31	1.6	C	.39	.33	78.6	68.9	517
14	15	42	.67	.34	1.21	1.5	1.20	1.1	D	.10	.33	59.5	68.9	514
18	22	42	-.10	.33	1.13	1.5	1.19	1.3	E	.16	.33	52.4	62.4	518
19	21	42	.01	.33	1.13	1.5	1.12	.9	F	.18	.33	50.0	62.6	519
8	28	42	-.77	.35	1.07	.6	1.06	.3	G	.23	.32	69.0	69.9	58
23	21	42	.01	.33	1.07	.8	1.02	.2	H	.27	.33	54.8	62.6	523
10	19	42	.22	.33	1.05	.5	1.02	.2	I	.28	.33	59.5	63.3	510
5	9	42	1.47	.40	1.02	.1	.98	.0	J	.29	.31	81.0	80.2	95
12	8	42	1.63	.41	1.01	.1	1.00	.1	K	.28	.30	83.5	82.8	512
16	15	42	.67	.34	.99	.0	.97	-.1	L	.35	.33	69.0	68.9	516
21	28	42	-.77	.35	.96	-.3	.91	-.4	M	.37	.32	73.8	69.9	521
13	32	42	-1.29	.38	.95	-.2	.93	-.1	N	.35	.30	76.2	77.7	513
25	23	42	-.21	.33	.93	-.8	.89	-.7	O	.42	.33	73.8	62.9	525
22	23	42	-.21	.33	.91	-.9	.85	-1.0	P	.44	.33	69.0	62.9	522
11	22	42	-.10	.33	.98	-1.1	.84	-1.1	Q	.46	.33	66.7	62.4	511
15	37	42	-2.21	.50	.89	-.2	.68	-.5	R	.39	.25	90.5	88.2	515
1	34	42	-1.60	.41	.89	-.4	.88	-.5	S	.41	.28	85.7	81.8	51
7	24	42	-.31	.33	.88	-1.2	.83	-1.1	T	.47	.33	73.8	63.6	57
20	9	42	1.47	.40	.86	-.5	.88	-.3	U	.44	.31	85.7	80.2	520
4	28	42	-.77	.35	.86	-1.0	.82	-.9	V	.48	.32	73.8	69.9	54
3	17	42	.44	.33	.86	-1.3	.83	-1.0	W	.50	.33	76.2	65.6	53
2	26	42	-.54	.34	.86	-1.3	.78	-1.3	X	.50	.32	71.4	66.5	52
9	28	42	-.77	.35	.84	-1.2	.78	-1.1	Y	.51	.32	73.8	69.9	59
MEAN	21.1	42.0	.80	.36	.99	.0	1.05	.1				73.2	70.5	
S.D.	8.8	.0	.98	.04	.14	1.0	.43	1.2				10.5	7.8	

Figure 3. Winstep Misfit Order (Validity) Results

In the results of the Misfit Order analysis, it was found that items number 24 and 6 did not meet the three question validity criteria, namely MNSQ, ZSTD, and PT Measure Corr. Both of these items can be considered invalid in measuring the desired construct. This indicates a misunderstanding of the students towards both items, as they do not function normally in the measurement process. Therefore, these two items need to be revised before the exam can be conducted. Revision can be done by changing or correcting the formulation, context, or answer choice of both items, resulting in a more valid measurement instrument (Tarigan, R., 2020). However, other items still meet the criteria of MNSQ and ZSTD, which indicate that the validity level of questions in the Final Semester Assessment (PAS) at Mantuil 2 Elementary School reaches 92%. That is, most items on the measuring instrument can be considered valid in measuring the desired construct.

In the context of learning evaluation, it is important to ensure the validity of measurement instruments in order to obtain accurate and reliable information about student achievement. The validity of an item can be affected by a variety of factors, including the suitability of the item to the learning context, the language used, and the appropriateness of the instructions given to students (Suherman, E., 2018).

Difficulty Level (P value)

An item will be considered a good question if the difficulty level is not too difficult or too easy, but rather moderate enough. Robert L. Thorndike has developed a formula for calculating the difficulty index of an item. To categorize the difficulty of an item, the following classifications are used: 0.00-0.30 (difficult), 0.31-0.70 (medium), and 0.71-1.00 (easy) (Ngalimun, 2018).

The difficulty level is divided into three categories by Nana Sujana, with a good category between 0.25 and 0.75. The question is said to be challenging if the difficulty level is less than 0.25 and too easy if it is greater than 0.75 (Nugroho, 2021).

Score	Frequency
0,0 to 0,1	0
0,1 to 0,2	2
0,2 to 0,3	2
0,3 to 0,4	4
0,4 to 0,5	2
0,5 to 0,6	7
0,6 to 0,7	5
0,7 to 0,8	1
0,8 to 0,9	2
0,9 to 1,0	0

Figure 4. Frequency Distribution for the P values

In the analysis conducted using Iteman software, there are four questions that have a P value or difficulty level below 0.25, indicating that these questions are very difficult. On the contrary, there are three questions with scores above 0.75, which indicates that these questions are too easy. In addition, there are 18 questions with scores between 0.25-0.75, which indicates a sufficient level of difficulty. In the context of learning evaluation, it is important to maintain a balance of difficulty levels in the formulation of questions. Problems that are too easy or too difficult can reduce the ability of measuring instruments to distinguish between different levels of student comprehension or achievement. Therefore, it is necessary to reassess the questions that have a level of difficulty that is not ideal (Kurniawan, S., & Yuliani, 2020).

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT MATCH EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
24	7	42	1.81	.44	1.38	1.3	2.96	3.4	-.30	.29	76.2	83.8	S24
12	8	42	1.63	.41	1.01	.1	1.00	.1	.28	.30	83.3	82.0	S12
5	9	42	1.47	.40	1.02	.1	.98	.0	.29	.31	81.0	80.2	S5
20	9	42	1.47	.40	.86	-.5	.88	-.3	.44	.31	85.7	80.2	S20
14	15	42	.67	.34	1.21	1.5	1.20	1.1	.10	.33	59.5	68.9	S14
16	15	42	.67	.34	.99	.0	.97	-.1	.35	.33	69.0	68.9	S16
17	15	42	.67	.34	.87	-1.0	1.31	1.6	.39	.33	78.6	68.9	S17
6	16	42	.55	.34	1.32	2.5	1.50	2.5	-.08	.33	57.1	67.2	S6
3	17	42	.44	.33	.86	-1.3	.83	-1.0	.50	.33	76.2	65.6	S3
10	19	42	.22	.33	1.05	.5	1.02	.2	.28	.33	59.5	63.3	S10
19	21	42	.01	.33	1.13	1.5	1.12	.9	.18	.33	50.0	62.6	S19
23	21	42	.01	.33	1.07	.8	1.02	.2	.27	.33	54.8	62.6	S23
11	22	42	-.10	.33	.90	-1.1	.84	-1.1	.46	.33	66.7	62.4	S11
18	22	42	-.10	.33	1.13	1.5	1.19	1.3	.16	.33	52.4	62.4	S18
22	23	42	-.21	.33	.91	-.9	.85	-1.0	.44	.33	69.0	62.9	S22
25	23	42	-.21	.33	.93	-.8	.89	-.7	.42	.33	73.8	62.9	S25
7	24	42	-.31	.33	.88	-1.2	.83	-1.1	.47	.33	73.8	63.6	S7
2	26	42	-.54	.34	.86	-1.3	.78	-1.3	.50	.32	71.4	66.5	S2
4	28	42	-.77	.35	.86	-1.0	.82	-.9	.48	.32	73.8	69.9	S4
8	28	42	-.77	.35	1.07	.6	1.06	.3	.23	.32	69.0	69.9	S8
9	28	42	-.77	.35	.84	-1.2	.78	-1.1	.51	.32	73.8	69.9	S9
21	28	42	-.77	.35	.96	-.3	.91	-.4	.37	.32	73.8	69.9	S21
13	32	42	-1.29	.38	.95	-.2	.93	-.1	.35	.30	76.2	77.7	S13
1	34	42	-1.60	.41	.89	-.4	.80	-.5	.41	.28	85.7	81.8	S1
15	37	42	-2.21	.50	.89	-.2	.68	-.5	.39	.25	90.5	88.2	S15
MEAN	21.1	42.0	.00	.36	.99	.0	1.05	.1			71.2	70.5	
S.D.	8.0	.0	.98	.04	.14	1.0	.43	1.2			10.5	7.8	

Figure 5. Measure Order (Rasch model)

The level of difficulty of an item in this study can be measured using the logit value contained in the "measure" column. A high logit number indicates a higher level of difficulty in answering the item. In the Rasch model analysis, item difficulty can be classified into three categories based on the log value, namely: "Very difficult" category (logit value > 0.98): The questions that fall into this category have a very high level of difficulty and are considered difficult tasks for respondents to answer correctly. "Good" category (logit score 0.0-0.98): Questions that fall into this category have a good level of difficulty, where respondents have a fairly balanced probability of answering correctly. "Easy" category (logit score < 0.98): Questions that fall into this category are considered easy by respondents, with a lower level of difficulty (Sumarni, W., 2019).

Based on the measure or logit value, there are four questions (numbers 24, 12, 5, and 20) with a value greater than 0.98, indicating that the questions have a high level of difficulty or are very difficult. Conversely, there are three questions (numbers 13, 1, and 15) with a measured value of less than 0.98, indicating that these questions fall into the easy category. In addition, the fact that the measured value ranges from 0.0 to 0.98 indicates that there are 18 questions (numbers 21, 9, 8, 4, 2, 7, 25, 22, 18, 11, 23, 19, 10, 3, 6, 17, 16, and 14) that belong to the medium difficulty category. It is important to ensure a balance of difficulty in the evaluation instrument. This will allow us to get a more comprehensive picture of student's understanding and abilities. If there is a significant imbalance, it is necessary to revise or adjust the questions that have a level of difficulty that is not ideal (D. Suryadi, 2018).

The Power difference (Rpbis) Iteman

The distinguishing power of an item is the ability to select between students in whom there is expertise in answering questions in a material. If the item differentiating power index is high, then the item is effective in distinguishing students' abilities. However, if the discriminating power index of the question items is low, it means that more students do not understand the topic. The basic assumption when applying any evaluation score is that strange actions should have no impact on the examinee's answers. However, other aberrant reactions are often observed in testing (Wan & Keller, 2023). This means that any action of the student will not affect the difference in the question items if the question is analyzed. There are many criteria that can be used to assess the level of differentiating power of a question item, including 0.00-0.20 (less), 0.21-0.40 (sufficient), 0.41-0.70 (very good), and 0.71-1.00 (very good). The low distinguishing power for the item is indicated by a negative value (Arikunto, 2020).

Score	Frequency
-0,4 to -0,3	1
-0,3 to -0,2	1
-0,2 to -0,1	0
-0,1 to 0,0	1
0,0 to 0,1	2
0,1 to 0,2	5
0,2 to 0,3	3
0,3 to 0,4	9
0,4 to 0,5	3
0,5 to 0,6	0
0,6 to 0,7	0
0,7 to 0,8	0
0,8 to 0,9	0
0,9 to 1,0	0

Figure 6. Frequency Distributor For The Rqbis (Daya Beda) Iteman

Based on the statistical data in the table, there are 10 out of 25 items that have low distinguishing power, with scores between 0.0 to 0.20. A total of 12 items have a sufficient level of distinguishing power, with a score between 0.21 to 0.40, and there are 3 items with a score between 0.41 to 0.70. Unfortunately, none of the items showed strong discriminating power, with scores ranging from 0.71 to 1.00. In fact, there are 3 items with negative scores, which indicates weak discrimination power. Basically, discriminating power (also known as discrimination power) measures the extent to which an item can distinguish between students who have high ability and students who have low ability in what the item measures. High discriminating power indicates that items can effectively distinguish between students' ability levels (Pramono, R., 2019).

In the context of learning evaluation, it is important to have items that have high discriminating power. This allows us to obtain more accurate information about students' abilities and distinguish different ability levels. If there are items with low or negative discriminating power, this may indicate that they are not effective in distinguishing between students who have different ability levels (Samsiah, S., 2018). To increase the distinguishing power of items, it is necessary to revise or develop these items. Some strategies that can be done include: Pay attention to the selection of keywords or distractors that are more effective to distinguish between students with high and low abilities. Pay attention to the difficulty level of the item so that the item is not too easy or too difficult. Use more authentic and relevant contexts or situations to increase the distinguishing power of items. Involve experts or experts in the relevant field to review and advise on items (Azizah & Sutrisno, 2022).

Different Power Rasch Model

Rasch model analysis can be used to distinguish the individual skills of students who can answer questions with low ability. The same criteria are used to determine individual items and fits. The Individual fit was evaluated using mean square outfit (MNSQ) values ranging from 0.5 to 1.5, Z-standard outfit (ZSTD) scores ranging from -2.0 to +2.0, and point measurement correlation values ranging from 0.4 to 0.85 (Point Measure Correlation/Pt Ukur Kor) (Nuryanti et al., 2018).

The difference in the Rasch model can be seen from the person measure as follows:

Person STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS%	EXP%	Person
14	21	25	1.95	.58	.97	.0	.69	-.3	.39	.32	84.0	83.9	14PK
16	21	25	1.95	.58	.70	-.8	.45	-.9	.61	.32	84.0	83.9	16PK
17	21	25	1.95	.58	1.02	.2	.83	-.1	.33	.32	84.0	83.9	17PK
34	17	25	.90	.47	1.03	.2	1.10	.4	.35	.39	76.0	73.2	34PK
36	17	25	.90	.47	.82	-.9	.79	-.6	.55	.39	84.0	73.2	36LK
25	16	25	.69	.46	.75	-1.4	.71	-1.1	.63	.40	80.0	70.7	25PK
2	15	25	.49	.45	.76	-1.5	.67	-1.4	.64	.41	76.0	69.2	02PK
31	15	25	.49	.45	1.15	.9	1.08	.4	.29	.41	60.0	69.2	31LK
40	15	25	.49	.45	1.04	.3	.97	.0	.39	.41	68.0	69.2	40LK
1	14	25	.29	.44	1.09	.6	1.04	.2	.34	.41	60.0	68.2	01PK
4	14	25	.29	.44	.71	-1.9	.65	-1.6	.68	.41	84.0	68.2	04PK
9	14	25	.29	.44	1.07	.5	1.08	.4	.34	.41	76.0	68.2	09LK
18	14	25	.29	.44	.88	-.7	.80	-.8	.54	.41	68.0	68.2	18PK
28	14	25	.29	.44	1.00	.1	.95	-.1	.42	.41	68.0	68.2	28PK
26	14	25	.29	.44	1.04	.3	1.10	.5	.36	.41	68.0	68.2	26PK
42	14	25	.29	.44	1.14	.9	1.14	.7	.28	.41	52.0	68.2	42PK
5	13	25	.10	.44	.64	-2.6	.58	-2.2	.75	.41	84.0	67.3	05LK
6	13	25	.10	.44	1.01	.1	.92	-.3	.43	.41	60.0	67.3	06PK
13	13	25	.10	.44	.96	-.2	.95	-.2	.45	.41	68.0	67.3	13LK
35	13	25	.10	.44	.68	-2.2	.61	-2.0	.71	.41	84.0	67.3	35LK
11	12	25	-.10	.44	.89	-.7	.83	-.7	.52	.41	72.0	66.7	11PK
21	12	25	-.10	.44	.89	-.7	.81	-.8	.53	.41	72.0	66.7	21PK
23	12	25	-.10	.44	.85	-.9	.77	-1.1	.56	.41	72.0	66.7	23PK
28	12	25	-.10	.44	1.01	.1	1.07	.4	.39	.41	72.0	66.7	28PK
29	12	25	-.10	.44	1.05	.4	1.08	.4	.36	.41	64.0	66.7	29LK
41	12	25	-.10	.44	1.19	1.2	1.29	1.3	.21	.41	64.0	66.7	41LK
3	11	25	-.29	.44	.86	-.8	.87	-.5	.53	.41	68.0	67.4	03PK
7	11	25	-.29	.44	1.06	.4	1.11	.6	.34	.41	60.0	67.4	07PK
12	11	25	-.29	.44	1.28	1.7	1.58	2.2	.09	.41	60.0	67.4	12PK
27	11	25	-.29	.44	.91	-.5	.83	-.7	.50	.41	68.0	67.4	27LK
30	11	25	-.29	.44	1.17	1.1	1.32	1.4	.22	.41	68.0	67.4	30PK
32	11	25	-.29	.44	.95	-.3	.88	-.5	.47	.41	76.0	67.4	32PK
39	11	25	-.29	.44	1.09	.6	1.09	.5	.33	.41	68.0	67.4	39PK
19	10	25	-.49	.45	.81	-1.1	.72	-1.1	.59	.40	76.0	68.6	19PK
24	10	25	-.49	.45	1.29	1.6	1.44	1.7	.11	.40	60.0	68.6	24LK
37	10	25	-.49	.45	1.20	1.2	1.35	1.3	.20	.40	60.0	68.6	37LK
15	9	25	-.69	.45	1.13	.8	1.18	.7	.27	.40	60.0	70.0	15PK
8	8	25	-.90	.47	1.23	1.1	1.09	.4	.22	.39	56.0	72.4	08PK
33	8	25	-.90	.47	1.09	.5	1.06	.3	.31	.39	72.0	72.4	33PK
22	7	25	-1.12	.48	1.05	.3	1.53	1.4	.26	.38	84.0	75.2	22LK
38	5	25	-1.64	.54	1.07	.3	1.75	.6	.23	.34	84.0	81.0	38LK
10	3	25	-2.32	.65	1.32	.8	1.63	2.4	.22	.29	88.0	87.9	10LK
MEAN	12.5	25.0	.01	.46	1.00	.0	1.05	.0			71.2	70.5	
S.D.	3.7	.0	.81	.05	.17	1.0	.48	1.0			9.5	5.4	

Figure 7. Measure Order

In Rasch's model, we can see the discriminating power in terms of students' ability to answer items. The table on students 12 and 10 shows that the two students did not meet the three-component fit requirement. This indicates that the student's response pattern has not been ideal or is inconsistent with the level of ability that should be.

It is important to understand the pattern of student responses to certain items. When a student's response pattern is not as expected, it can provide an indication of problems in student understanding, interpretation of instruction, or other factors that affect their performance. Analysis of student response patterns can provide insight into certain aspects that need attention in the development and improvement of evaluation instruments (Herawati & Adnyana, 2021).

Distractor

The Iteman program version 3.00 is used to evaluate the quality of answer options on a test and can display the effectiveness index of answer options through the results of Iteman Prop Endorsing. If the percentage of respondents who choose one of the biser answer choices and the biser alternative value of the negative

answer option exceeds 5%, it indicates that the majority of voters from that answer option come from the group of students with low ability, and the answer option is effective as a distraction. The effectiveness of the distractor is determined by analyzing the item using the Rasch model with the help of the Winsteps program; If the value is greater than 5%, the distractor operates. Furthermore, the distractor problem is said to be effective if the Average Ability value is negative, which indicates that the distractor was chosen by a low-ability participant (Purniasari, Masykuri, & Ariani, 2021).

A distractor is said to function if all of them have voters, and it is said to be dysfunctional if of them no one votes (Himawan & Nurgiyantoro, 2022). As stated in Annex 1 distractors use Iteman software and Annex 2 uses Winstep Rasch Model software.

Based on the analysis of distraction questions conducted using Iteman software and Winstep Rasch Model software for 25 class VI Thematic PAS questions at SDN Mantuil 2 Banjarmasin City, there are 23 questions whose distractors function properly, namely at numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25. Questions 11 and 15 are examples of distractors failing to function properly. If all or three distractors are active, then item distractors are active. If you have one, two, or three malfunctioning distractors, they will not work (Himawan & Nurgiyantoro, 2022).

Conclusion

The level of validity of the 25 questions is that two questions are declared invalid because they do not meet the three validity requirements, namely the value of MNSQ, ZSTD, and PT Measure. The results of the analysis based on classical test theory get a score of 0.641 and using modern tests get a score of 0.68 shows that the reliability of the question items does not reach an alpha coefficient of 0.70 which means the reliability of the questions is still low. The spread of difficulty levels using classical tests and modern tests gets the same result, that is, there are 4 questions very difficult, 3 questions too easy, 18 questions difficult. Differentiating power uses classical tests of 10 points of bad questions, 12 points of discriminating power are enough, and 3 points of discriminating power questions are very bad. The difference with modern tests is that there are two sides with the worst difference, meaning that student response patterns are not ideal. Based on the analysis of distraction questions conducted using classical tests and modern tests for 25 questions, there are 23 questions for which the distractor works properly.

References

- Adji, W. S., Ramadhan, B., Studi, P., Guru, P., Ibtidaiyah, M., Islam, U., & Antasari, N. (2018). Evaluasi Mata Kuliah Pendidikan Karakter. *Al-Adzka*, VIII(September), 85–106.
- Airasian, P. W., & Cruikshank, K. A. (2018). *Assessment of student learning: A comprehensive guide to the analysis of student learning*. Pearson.
- Alfarisa, F., Chudari, I. N., & Robiansyah, F. (2019). Analisis Butir Soal IPS Kelas V Sekolah Dasar Menggunakan Software ITEMAN. *EduBasic Journal: Jurnal Pendidikan Dasar*, 1(2), 100–106. <https://doi.org/10.17509/ej.v1i2.26474>
- Arikunto, S. (2020). *Dasar-dasar evaluasi pendidikan (2nd ed.)*. Jakarta: Bumiaksara.
- Ayub, M. R. S. S. N., Istiyono, E., Munadi, S., Permadi, C., Pattiserlihun, A., & Sudjito, D. N. (2020). Analisa Penilaian Soal Fisika Menggunakan Model Rasch Dengan Program R. *Jurnal Sains Dan Edukasi Sains*, 3(2), 46–52.

- <https://doi.org/10.24246/juses.v3i2p46-52>
- Azizah, R., & Sutrisno, A. (2022). Analisis Kesulitan dan Daya Pembeda Soal Ujian Sekolah Berstandar Nasional Bahasa Inggris SMA. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 7(2), 186-.
- DeMars, C. E. (2019). *Item response theory*. Oxford University Press.
- Engelhard Jr, G. (2018). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.
- Fauziana, A., & Dessy Wulansari, A. (2021). Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch. *Ibriez: Jurnal Kependidikan Dasar Islam Berbasis Sains*, 6, 10-19. <https://doi.org/10.21154/ibriez.v6i1.112>
- Febrianawati. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jorpres (Jurnal Olahraga Prestasi)*, 13(1), 53-59. <https://doi.org/10.21831/jorpres.v13i1.12884>
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2019). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 24(1), 1-1.
- Harmuni, L. (2019). *Instrumen Penilaian dan Validasinya*. Ponogoro: Uwais Inspirasi Indonesia.
- Herawati, D., & Adnyana, I. G. A. A. (2021). Pengaruh Strategi Pembelajaran dan Penilaian Pembelajaran Terhadap Pemahaman Konsep dan Keterampilan Berpikir Kritis Siswa. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 6(2), 178-.
- Heryanti, ai denti. (2022). Standar Penilaian Pendidikan Permendikbudristek No 21 tahun 2022. *Gurusumedang.Com*. Retrieved from <https://www.gurusumedang.com/2022/06/standar-penilaian-pendidikan.html>
- Himawan, R., & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN (Analysis of exercise items for odd semester end of semester Indonesian language subjects class. *Kembara: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 8(1), 160-180.
- Janna, N. M., & Herianto. (2021). KONSEP UJI VALIDITAS DAN RELIABILITAS DENGAN MENGGUNAKAN SPSS. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, (18210047), 1-12.
- Jayusman, I., & Shavab, O. A. K. (2020). Aktivitas Belajar Mahasiswa Dengan Menggunakan Media Pembelajaran Learning Management System (Lms) Berbasis Edmodo Dalam Pembelajaran Sejarah. *Jurnal Artefak*, 7(1), 13. <https://doi.org/10.25157/ja.v7i1.3180>
- Kurniawan, S., & Yuliani, F. R. (2020). Analisis Kesulitan Soal Ujian Sekolah Berstandar Nasional (USBN) Matematika Sekolah Menengah Pertama. *Jurnal Matematika*, 8(1), 9-18.
- Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya Evaluasi Dalam Pembelajaran Dan Akibat Memanipulasinya. *Jurnal Pendidikan Dan Sains*, 2(2), 244-257. Retrieved from <https://ejournal.stitpn.ac.id/index.php/bintang>
- Neto, J., Neto, F., & Furnham, A. (2022). Predictors of students' preferences for assessment methods. *Assessment and Evaluation in Higher Education*, 48(4), 556-565. <https://doi.org/10.1080/02602938.2022.2087860>
- Ngalimun. (2018). *Evaluasi dan penilaian pembelajaran*. Yogyakarta: Parama Ilmu.
- Nugroho, W. (2021). Pendekatan Problem Based Learning Model Diskusi Kelompok Berbantuan Video YouTube untuk Meningkatkan Aktivitas dan Hasil Belajar Statistika. *Jurnal Pendidikan Matematika (Kudus)*, 4(2), 211. <https://doi.org/10.21043/jmtk.v4i2.12259>

- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan Iteman and rasch model analysis on the development of critical thinking instruments for vocational school students. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233. Retrieved from <http://journal.uny.ac.id/index.php/jipi>
- Pramono, R., et al. (2019). Analisis Kesulitan Butir Soal dan Daya Pembeda pada Ujian Akhir Semester Fisika. *Jurnal Pembelajaran Fisika*, 7(2), 165-.
- Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia Sma N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman Dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214. Retrieved from <https://jurnal.uns.ac.id/jpkim>
- Samsiah, S., et al. (2018). Analisis Butir Soal Tes Kemampuan Berpikir Kritis. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 3(10), 131.
- Santosa, S., & Badawi, J. A. (2022). Analisis Butir Soal Pilihan Ganda Tema Pertumbuhan dan Perkembangan Makhluk Hidup Kelas III Madrasah Ibtidaiyah. *Jurnal Basicedu*, 6(2), 1678–1686. <https://doi.org/10.31004/basicedu.v6i2.2206>
- Suherman, E., et al. (2018). Validitas Isi Instrumen Evaluasi Pembelajaran. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 3(8), 1001.
- Sulistiyawati, W., Wahyudi, & Trinuryono, S. (2022). Analisis (Deskriptif Kuantitatif) Motivasi Belajar Siswa Dengan Model Blended Learning Di Masa Pandemi Covid19. *Kadikma*, 13, No. 1, 68–73.
- Sumarni, W., et al. (2019). Analisis Kualitas Soal Ulangan Harian Matematika Berdasarkan Analisis Kualitas Butir Soal dan Analisis Kesalahan Jawaban Peserta Didik. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 4(11), 150.
- Suryadi, A. (n.d.). *Evaluasi Pembelajaran*. Jawa Barat: CV. Jejak.
- Suryadi, D. (2018). Pengembangan Instrumen Evaluasi Pembelajaran dan Penilaian Hasil Belajar. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 3(11), 150.
- Suskie, L. (2018). *Assessing student learning: A common sense guide*. John Wiley & Sons.
- Tarigan, R., et al. (2020). Validitas dan Reliabilitas Instrumen Penilaian Pembelajaran IPA Kelas VII di SMP. *Jurnal Pendidikan IPA Indonesia*, 9(1), 11–1.
- Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu. *ELSE (Elementary School Education Journal): Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, 4(2), 124. <https://doi.org/10.30651/else.v4i2.7090>
- Wan, S., & Keller, L. A. (2023). Using Cumulative Sum Control Chart to Detect Aberrant Responses in Educational Assessments. *Practical Assessment, Research and Evaluation*, 28(February).
- Widhiarso, W., & Sumintono, B. (2019). *Aplikasi pemodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunikata Publishing House.
- Widiastuti, N. P. K., Putrayasa, I. B., & ... (2022). Menyusun Instrument Penilaian Keterampilan Menulis Siswa Sekolah Dasar. ... *Dan Pembelajaran*, 5(1), 50–56. Retrieved from <https://ejournal.undiksha.ac.id/index.php/JP2/article/view/42473%0Ahttps://ejournal.undiksha.ac.id/index.php/JP2/article/download/42473/22326>
- Widodo, H. (2021). *Evaluasi Pendidikan*. Yogyakarta: UAD Press.
- Yulianto, A., & Widodo, A. (2020). Disclosure of Difficulty Distribution of HOTS-

Based Test Questions through Rasch Modeling. *Indonesian Journal of Primary Education*, 4(2), 197–203. <https://doi.org/10.17509/ijpe.v4i2.29318>

Appendices

Appendix 1. Results of evaluation of the effectiveness of the distractor (Iteman)

No.	Distraktor (Prop)	N	Category
1.	A. .070	3	Effective Distraktor
	B. .048	2	Effective Distraktor
	C. .081	34	The Right Answer
	D. .071	3	Effective Distraktor
2.	A. .619	26	The Right Answer
	B. .024	1	Effective Distraktor
	C. .310	13	Effective Distraktor
	D. .048	2	Effective Distraktor
3.	A. .238	10	Effective Distraktor
	B. .405	17	The Right Answer
	C. .071	3	Effective Distraktor
	D. .286	12	Effective Distraktor
4.	A. .095	4	Effective Distraktor
	B. .119	5	Effective Distraktor
	C. .119	5	Effective Distraktor
	D. .667	28	The Right Answer
5.	A. .333	14	Effective Distraktor
	B. .214	9	The Right Answer
	C. .143	6	Effective Distraktor
	D. .310	13	Effective Distraktor
6.	A. .333	14	Effective Distraktor
	B. .095	4	Effective Distraktor
	C. .190	8	Effective Distraktor
	D. .381	16	The Right Answer
7.	A. .167	7	Effective Distraktor
	B. .190	8	Effective Distraktor
	C. .571	24	The Right Answer
	D. .071	3	Effective Distraktor
8.	A. .667	28	The Right Answer
	B. .048	2	Effective Distraktor
	C. .095	4	Effective Distraktor
	D. .190	8	Effective Distraktor
9.	A. .667	28	The Right Answer
	B. .095	4	Effective Distraktor
	C. .119	5	Effective Distraktor
	D. .119	5	Effective Distraktor
10.	A. .143	6	Effective Distraktor
	B. .214	9	Effective Distraktor
	C. .190	8	Effective Distraktor
	D. .452	19	The Right Answer
11.	A. .333	14	Effective Distraktor
	B. .524	22	The Right Answer
	C. .00	0	Ineffective Distraktor
	D. .143	6	Effective Distraktor
12.	A. .333	14	Effective Distraktor
	B. .190	8	Effective Distraktor
	C. .190	8	The Right Answer
	D. .286	12	Effective Distraktor

13.	A. .190 B. .762 C. .024 D. .024	8 32 1 1	Effective Distraktor The Right Answer Effective Distraktor Effective Distraktor
14.	A. .476 B. .095 C. .357 D. .071	20 4 15 3	Effective Distraktor Effective Distraktor The Right Answer Effective Distraktor
15.	A. .024 B. .095 C. .00 D. .881	1 4 0 37	Effective Distraktor Effective Distraktor Ineffective Distraktor The Right Answer
16.	A. .381 B. .357 C. .024 D. .238	16 15 1 10	Effective Distraktor The Right Answer Effective Distraktor Effective Distraktor
17.	A. .357 B. .071 C. .262 D. .310	15 3 11 13	The Right Answer Effective Distraktor Effective Distraktor Effective Distraktor
18.	A. .214 B. .143 C. .119 D. .524	9 6 5 22	Effective Distraktor Effective Distraktor Effective Distraktor The Right Answer
19.	A. .500 B. .119 C. .310 D. .071	21 5 13 3	The Right Answer Effective Distraktor Effective Distraktor Effective Distraktor
20.	A. .143 B. .405 C. .214 D. .238	6 17 9 10	Effective Distraktor Effective Distraktor The Right Answer Effective Distraktor
21.	A. .238 B. .071 C. .024 D. .667	10 3 1 28	The Right Answer Effective Distraktor Effective Distraktor Effective Distraktor
22.	A. .190 B. .071 C. .548 D. .190	8 3 23 8	Effective Distraktor Effective Distraktor The Right Answer Effective Distraktor
23.	A. .024 B. .405 C. .071 D. .500	1 17 3 21	Effective Distraktor Effective Distraktor Effective Distraktor The Right Answer
24.	A. .286 B. .167 C. .452 D. .095	12 7 19 4	Effective Distraktor The Right Answer Effective Distraktor Effective Distraktor
25.	A. .143 B. .071 C. .238 D. .548	6 3 10 23	Effective Distraktor Effective Distraktor Effective Distraktor The Right Answer

Appendix 2. Results of deceptive function analysis (Wistep Rasch Model)

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MEASURE ORDER									
ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	Item
24	A	0	12	29	-.12	.26	1.1	-.11	S24
	C	0	19	45	.25	.16	1.4	.26	
	D	0	4	10	.26	.37	1.3	-.10	
	B	1	7	17	-.53*	.35	3.3	-.30	
12	B	0	8	19	-.25	.32	.8	-.16	S12
	A	0	14	33	-.22	.19	.8	-.21	
	D	0	12	29	.16	.21	1.3	.11	
	C	1	8	19	.49	.34	1.0	.28	
5	D	0	13	31	-.29	.21	.8	-.25	S5
	A	0	14	33	-.11	.18	.9	-.11	
	C	0	6	14	.27	.41	1.7	.13	
	B	1	9	21	.47	.30	1.0	.29	
20	D	0	10	24	-.34	.25	.7	-.25	S20
	A	0	6	14	-.33	.16	.6	-.17	
	B	0	17	40	-.02	.15	1.0	-.03	
	C	1	9	21	.70	.35	.9	.44	
14	D	0	3	7	-.42	.63	.8	-.15	S14
	B	0	4	10	-.05	.09	.9	-.02	
	A	0	20	48	.01	.22	1.5	.00	
	C	1	15	36	.12	.17	1.1	.10	
16	D	0	10	24	-.46	.31	.8	-.33	S16
	A	0	16	38	-.10	.08	.9	-.11	
	C	0	1	2	.90		2.3	.17	
	B	1	15	36	.39*	.24	1.0	.35	

17	C	0	11	26	-.33	.10	.7	-.25	S17
	D	0	13	31	-.22	.11	.8	-.19	
	B	0	3	7	.16	.23	1.1	.05	
	A	1	15	36	.44	.30	1.6	.39	
6	B	0	4	10	-.24	.09	.7	-.10	S6
	A	0	14	33	-.06	.25	1.3	-.06	
	C	0	8	19	.44	.37	2.4	.25	
	D	1	16	38	-.07*	.17	1.5	-.08	
3	C	0	3	7	-.43	.24	.6	-.15	S3
	A	0	10	24	-.35	.16	.7	-.25	
	D	0	12	29	-.27	.30	1.2	-.22	
	B	1	17	40	.50	.16	.7	.50	
10	C	0	8	19	-.28	.31	.9	-.17	S10
	A	0	6	14	-.26	.29	.9	-.14	
	B	0	9	21	-.08	.19	1.0	-.06	
	D	1	19	45	.27	.21	1.1	.28	
19	D	0	3	7	-.84	.74	.6	-.29	S19
	B	0	5	12	-.38	.21	.7	-.18	
	C	0	13	31	.13	.09	1.2	.09	
	A	1	21	50	.16	.21	1.3	.18	
23	B	0	17	40	-.24	.19	1.0	-.26	S23
	C	0	3	7	-.10	.22	.9	-.04	
	A	0	1	2	.10		1.1	.02	
	D	1	21	50	.23	.19	1.1	.27	
11	D	0	6	14	-.47	.39	.8	-.24	S11
	A	0	14	33	-.34	.17	.9	-.31	
	B	1	22	52	.37	.16	.8	.46	
18	C	0	5	12	-.40	.31	.8	-.19	S18
	A	0	9	21	-.09	.40	1.7	-.07	
	B	0	6	14	.06	.15	1.1	-.03	
	D	1	22	52	.14	.16	1.1	.16	
22	A	0	8	19	-.67	.34	.7	-.41	S22
	B	0	3	7	-.49	.20	.6	-.17	
	D	0	8	19	-.05	.15	1.0	-.04	
	C	1	23	55	.34	.16	.9	.44	
25	C	0	10	24	-.58	.31	.8	-.41	S25
	A	0	6	14	-.13	.09	.9	-.07	
	B	0	3	7	-.10	.53	1.2	-.04	
	D	1	23	55	.32	.15	.9	.42	
7	D	0	3	7	-1.25	.59	.4	-.43	S7
	A	0	7	17	-.51	.24	.7	-.29	
	B	0	8	19	-.05	.13	1.0	-.04	
	C	1	24	57	.35	.15	.8	.47	
2	C	0	13	31	-.56	.21	.7	-.47	S2
	B	0	1	2	-.49		.6	-.10	
	D	0	2	5	-.19	.10	.9	-.06	
	A	1	26	62	.34	.14	.8	.50	
4	A	0	4	10	-1.29	.44	.4	-.52	S4
	B	0	5	12	-.38	.34	1.0	-.18	
	C	0	5	12	-.10	.06	1.0	-.05	
	D	1	28	67	.29	.13	.8	.48	
8	B	0	2	5	-.71	.42	.6	-.20	S8
	C	0	4	10	-.25	.27	.9	-.11	
	D	0	8	19	-.14	.25	1.1	-.09	
	A	1	28	67	.15	.17	1.2	.23	

9	B	0	4	10	-1,14	.51	.5	-.46	S9
	D	0	5	12	-.58	.19	.6	-.27	
	C	0	5	12	-.10	.21	1.1	-.05	
	A	1	28	67	.30	.13	.8	.51	
21	B	0	3	7	-.68	.56	.7	-.24	S21
	A	0	10	24	-.37	.28	1.0	-.26	
	C	0	1	2	-.10		1.0	-.02	
	D	1	28	67	.23	.14	.9	.37	
13	D	0	1	2	-.69		.6	-.14	S13
	A	0	8	19	-.52	.33	1.0	-.32	
	C	0	1	2	-.10		1.0	-.02	
	B	1	32	76	.17	.13	1.0	.35	
1	B	0	2	5	-1.02	1.31	.8	-.28	S1
	A	0	3	7	-.94	.39	.5	-.33	
	D	0	3	7	-.16	.06	1.0	-.06	
	C	1	34	81	.17	.12	.9	.41	
15	B	0	4	10	-1.00	.46	.6	-.41	S15
	A	0	1	2	-.29		.9	-.06	
	D	1	37	88	.13	.12	.9	.39	

Lampiran 3. Instrumen Penelitian



PENILAIAN TENGAH SEMESTER
SDN MANTUIL 2
TAHUN AJARAN 2022/2023



Tema : I Nama :
Kelas / Semester : VI / I

I. Berilah tanda silang (X) huruf a, b, c, atau d pada jawaban yang paling tepat!

PPKN KD 3.1

1. Memajukan pergaulan demi persatuan dan kesatuan bangsa yang ber-Bhineka Tunggal Ika merupakan perilaku yang mencerminkan Pancasila sila ke....
a. 1 b. 2 c. 3 d. 4
2. Menghormati teman yang sedang melakukan ibadah merupakan penerapan Pancasila sila....
a. Ketuhanan yang maha esa c. Kemanusiaan yang adil dan beradab
b. Persatuan Indonesia d. Keadilan sosial bagi seluruh rakyat Indonesia
3. Sikap positif yang tidak termasuk dalam sila kelima adalah....
a. Menghargai hasil karya orang lain c. Menyayangi semua manusia
b. melaksanakan hidup sederhana d. Mengupayakan kesejahteraan rakyat
4. Pemilihan kepala daerah di Kalimantan Selatan kemarin, merupakan bentuk penerapan Pancasila sila...
a. Pertama b. Kedua c. Ketiga d. Keempat
5. Beni dan kawan-kawan sedang bergotong royong membersihkan halaman sekolah. Prilaku Beni dan kawan-kawan tersebut merupakan pengamalan Pancasila sila....
a. Kedua b. Ketiga c. Keempat d. Kelima

BAHASA INDONESIA KD. 3.1

Balach teks di bawah ini untuk soal nomor 6 sampai dengan 10!

Tumbuhan Bagi Manusia

Tumbuhan merupakan jenis makhluk hidup yang berdampingan dan bermanfaat bagi manusia. Beragam manfaat dari tumbuhan dapat dimanfaatkan manusia dalam kehidupannya. Misalnya sayuran dan buah-buahan yang dapat dijadikan sumber vitamin dan mineral bagi tubuh manusia. Bahkan ada beberapa jenis tanaman yang diambil bagian tubuhnya sebagai bahan obat-obatan. Misalnya pohon kayu putih, jahe, dan mengkudu.

Beragam manfaat tumbuhan tersebut terkadang tidak diikuti dengan upaya menjaganya. Hilangnya habitat tempat tumbuh menjadi sebab utamanya. Perubahan tempat hidup ini disebabkan ulah manusia yang memperluas lahan tempat tinggal. Sehingga tumbuhan harus ditebang. Bahkan yang paling menyedihkan adalah sistem penebangan liar yang sangat merusak ekosistem hutan kita saat ini.

6. Jumlah kalimat yang terdapat dalam paragraf pertama di atas adalah....
a. dua b. tiga c. empat d. lima
7. Kalimat utama paragraf kedua dari teks di atas adalah....
a. Perubahan tempat hidup tumbuhan karena perluasan lahan
b. Penebangan liar karena memperluas lahan tempat tinggal
c. Hilangnya habitat tumbuhan karena penebangan liar
d. Rusaknya habitat tumbuhan karena tidak dijaga
8. Di bawah ini, informasi yang sesuai dengan isi paragraf pertama adalah....
a. Pohon Kayu Putih termasuk contoh tumbuhan yang bermanfaat bagi manusia
b. Semangka termasuk jenis tumbuhan yang banyak mengandung mineral
c. Pucuk Pohon Jambu dapat digunakan untuk obat tradisional bagi diare
d. Beragam manfaat tumbuhan tidak diikuti dengan upaya menjaganya

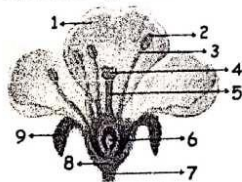
9. Ide pokok dari paragraf pertama teks di atas adalah....
- Tumbuhan merupakan makhluk hidup yang berdampingan dan bermanfaat bagi manusia
 - Ada beberapa jenis tanaman yang diambil bagian tubuhnya sebagai bahan obat-obatan
 - Beragam manfaat dari tumbuhan dapat dimanfaatkan manusia dalam kehidupannya
 - Pohon kayu putih, jahe, dan mengkudu adalah contoh tumbuhan yang bermanfaat
10. Di bawah ini pertanyaan yang berkaitan dengan paragraf kedua teks di atas adalah...
- Kapan tumbuhan dimanfaatkan manusia untuk kehidupannya?
 - Mengapa tumbuhan dapat memberi manfaat bagi manusia?
 - Bagaimana cara tumbuhan dapat kehilangan habitatnya?
 - Apa yang menyebabkan tumbuhan kehilangan habitatnya?

IPA KD. 3.1

11. Setiap makhluk hidup mempunyai kemampuan untuk berkembang biak. Berkembang biak adalah....
- kemampuan untuk berfotosintesis
 - kemampuan menambah keturunan
 - kemampuan membesarkan keturunan
 - Kemampuan menyesuaikan diri dengan lingkungan

12. Menempelnnya serbuk sari bunga mangga golek ke kepala putik bunga mangga gadung adalah contoh penyerbukkan
- bastar
 - sendiri
 - silang
 - tetangga
13. Menyusui merupakan ciri-ciri hewan yang berkembang biak secara....
- ovipar
 - vivipar
 - amphibi
 - reptile
14. Teknik perkembangbiakan tumbuhan dimana mata tunas yang memiliki sifat unggul dari tumbuhan berbeda digabungkan pada batang tumbuhan lain, disebut
- cangkok
 - enten
 - okulasi
 - runduk

15. Perhatikan gambar di bawah ini!



Nama bagian yang ditunjukkan oleh nomor 6 (enam) pada gambar di samping adalah....

- mahkota
- kepala putik
- benang sari
- bakal biji

IPS KD. 3.1


16. Mayoritas penduduk Myanmar adalah etnis....
- Manila
 - Burma
 - Bushman
 - Melayu
17. Berikut ini adalah persamaan keadaan sosial di negara Kawasan di Asia Tenggara adalah....
- sebagian besar penduduknya bekerja dalam bidang agraris
 - seluruh negara yang ada pernah dijajah oleh negara barat
 - sebagian besar penduduknya berasal dari rumpun suku melayu
 - semua negara yang berada di kawasan ini merupakan negara maju
18. Negara yang memiliki julukan Negara Gajah Putih, adalah....
- Laos
 - Vietnam
 - Indonesia
 - Thailand
19. Negara yang tidak memiliki garis pantai adalah....
- Brunei Darussalam
 - Malaysia
 - Laos
 - Indonesia
20. Negara yang hampir sama dengan Indonesia, yaitu seluruh wilayahnya berupa kepulauan adalah...
- Vietnam
 - Malaysia
 - Singapura
 - Filiphina


SBdP KD 3.4

21. Alat untuk menghaluskan permukaan ukiran kayu pada patung adalah....
a. pahat b. gergaji c. palu d. amplas

22. Teknik yang digunakan untuk membuat patung dari bahan tanah liat dan gips adalah....
a. cor b. tempa c. butsir d. pilin

23. Kegiatan awal dalam membuat patung dari pasir dan semen adalah membuat kerangka objek menggunakan....
a. benang b. kayu c. ranting d. kawat

24. Perhatikan gambar di bawah ini!
 Gambar patung di samping berasal dari
a. Hulu Sungai Tengah
b. Hulu Sungai Utara
c. Tanah Bumbu
d. Tanah Laut

25. Perhatikan gambar di bawah ini!
 Gambar patung di samping berasal dari....
a. Hulu Sungai Selatan b. Hulu Sungai Tengah
c. Banjarbaru
d. Kotabaru