STUDENTS' ABILITY TO WORK ON MULTIPLE CHOICE SCIENCE QUESTIONS IN CLASS VII USING THE RASCH MODEL

Anenda Bagus Satrya Ganesha [1*], Tina Hayati Dahlan [2], Salila Prasidya Hidayati [3]

[1.2] *Educational Psychology Study Program, Universitas Pendidikan Indonesia*
[3] *Applied English Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi (KMUTT) Bangkok, Thailand*

*Email : anendabagus@upi.edu

Abstract: This study aim to assess the ability of seventh-grade student to solve multiple choice science questions using the Rasch Model. A descriptive quantitative approach was employed, utilizing an instrument to consisting of 30 science questions aligned with the Kurikulum Merdeka. The sample comprised 75 junior high school students in Bandung City, selected through non probability sampling. The results indicated that participant 49 demonstrated the highest ability, whereas participants 15 and 22 as the lowest abilities. Question number 21 was identified as the most difficult, while question number 5 was the easiest. The Rasch Model revealed a high reliability of the questions, with a coefficient of 0.91. furthermore, the analysis indicated that question number 4,11,12,13,14,15, and 18 had low validity and should be discarded. Additionally, the Rasch Model was able to identify students who were potentially cheating during the test. These findings confirm that the Rasch Model is viable tool for evaluating the quality of test questions before they are administrated to students.

Keywords: Multiple Choice Questions, Rasch Model, Student Abilities

INTRODUCTION

Natural science is one of the subjects taught in junior high school. The aim is to improve students' understanding, ability, and scientific attitude about natural phenomena and the environment. To determine the ability of students to understand science, a test is carried out, or more precisely, an evaluation. Ideally, the tools or inquiries utilized to gauge student learning achievements should be meticulously crafted, developed according to question-writing guidelines, assessed, and scrutinized for excellence (Syahroni et al., 2024). An exam or test is an evaluation procedure commonly conducted by a teacher of a student's knowledge and skills to determine their performance using certain instruments (Sumintono, 2016). Evaluation is the process of assessing student learning outcomes in which there is also feedback for students and teachers (Kurniasi et al., 2020). The evaluation of learning outcomes assessment must include systematic steps to get more accurate results (Amrianto et al., 2019). Evaluation is one of the important components that must be present in the learning process as it is an activity to collect data to measure

the extent to which learning objectives have been achieved (Erfan et al., 2020). The results of the measurements are then interpreted and taken into account in the form of an assessment (Zainal, 2020).

The most popular type of test is multiple choice questions (MCQs). MCQs are a commonly used tool for evaluating students in education, and they typically necessitate selecting the accurate answer from a range of options (McKenna, 2018). The multiple-choice and True or False exams are especially suitable for situations where there are many test-takers, limited grading time, and extensive material coverage (Kusumawati et al., 2018). An item in Multiple-choice Questions called a Single or One Best Answer comprises a stem accompanied by several options, typically ranging from three to five. Among these options, one is correct, while the others serve as distractors (Elgadal & Mariod, 2021). Multiple-choice, fill-in-the-blank, practical, or descriptive tests are some of the ways to measure students' science competencies. This is because it has several advantages, such as being easy to prepare, present, and assess, and can measure various affective, cognitive, and psychomotor aspects (Sudjana, 1995). A good test is a test that is objective, valid, and also reliable (Valen & T. G. Satria, 2021). Tests are methods that, in measurement or methods, can be used to evaluate educational aspects (Septiliana, 2023). However, there are some disadvantages of multiple choice. These include being susceptible to guesswork, it is difficult to know why students choose certain answers, and it depends on the items' quality. The smallest unit of an exam, which consists of questions or statements that must be answered, is called a question item or what the student must complete. Factors such as validity, reliability, difficulty, differentiation, and distractors determine the quality of the items (Arikunto, 2021). A quality item is valid, reliable, has an appropriate level of difficulty and differentiation, and has effective distractors. Better tests can result from high-quality items, which in turn can improve student learning outcomes.

An instrument test is considered good if it provide precise information regarding student abilities in the competencies being assessed (Azizah & Sapti Wahyuningsih, 2020). The Rasch model is one way to measure item quality. Rasch model is one of the analysis models that can be used to determine the feasibility of an instrument (Maulana et al., 2023). The Rasch model is a psychometric model used to analyze categorical data such as observational behavior, questionnaire responses, or test answers based on respondent ability and item difficulty. The Rasch model analysis operates within a stochastic or probabilistic framework, where measurements involve considering two factors: the difficulty of test items and the abilities of individuals (Mahmud et al., 2013). Rasch model has many advantages, including the ability to estimate ability and difficulty parameters objectively and accurately, the ability to test the fit of the data to the model, the

ability to create interval scales from ordinal data, and the ability to conduct equating tests across groups or time (Bond & Fox, 2013).

The Rasch model has been widely used in various fields, including education. In several studies, the Rasch model has been used to analyze the quality of items in various fields, such as mathematics, English, Indonesian, and (Ginting et al., 2020). The results show that the Rasch model can be used to evaluate and improve the quality of question parts. In addition, it can provide feedback to students, teachers, and curriculum developers. The Rasch Measurement Models illustrate the correlation between an individual and an item, considering a shared latent trait. It enables the prediction of the probability that an individual with a specific ability will respond correctly to an item of a particular difficulty level (Abdullah et al., 2017).

Kurikulum Merdeka is one context in which the Rasch model can be applied. Kurikulum Merdeka is an emergency curriculum implemented in Indonesia. Kurikulum Merdeka was implemented due to a learning crisis caused by a virus that spread throughout the world starting in 2019 (Zakso, 2022). Kurikulum Merdeka has a variety of intracurricular learning, so students have enough time to learn concepts and strengthen their skills. Kurikulum Merdeka is a curriculum with diverse intracurricular learning, content will be optimized so that learners have enough time to explore concepts and strengthen (Khoirurrijal et al., 2022). Kurikulum Merdeka allows teachers to create good lessons that suit students' needs and learning environment (Kemendikbudristek, 2022). Kurikulum Merdeka, launched in 2022, includes grade VII science as an optional subject. Grade VII science is one of the subjects in the Kurikulum Merdeka.

According to the explanation above, the following research question arises: What is the quality of grade VII science multiple-choice items in the Kurikulum Merdeka based on the Rasch model? This study aimed to determine the quality of grade VII science multiple choice items in the independent curriculum using the Rasch model. This research provides information about the quality of grade VII science multiple choice items on the independent curriculum so teachers and schools can use it to assess and improve. Contribute to the development of science, especially in the field of educational measurement and science education.

METHOD

The research used a quantitative approach with descriptive research type through the use of secondary data, namely the instrument of seventh-grade science questions consisting of Chapter I material (Nature of Science and Scientific Method), Chapter II - Substances and Their Changes, and Chapter III (Temperature, Heat, and Expansion). The 30 multiple-choice questions with 2 answer options (correct and incorrect) through the Google form platform were given to 75 seventh-

grade students from various junior high schools in Bandung City with non probability sampling. The research data collection used techniques in the form of question instruments. Data collection through tests is used to determine the quality of questions based on analysis using the Rasch Model through the Winsteps software application. The Rasch Model offers several benefits, including the ability to detect incorrect responses, predict scores for missing data, distinguish between respondents with identical raw scores, analyze both dichotomous and polytomous data as well as their combinations, and identify signs of guessing and cheating (Linacre, 2016) The Rasch model has widely used by many researcher to analyze test questions for student, both in secondary schools and even in universities. The Rasch can detect student's misconceptions in the concentration of mechanics material in Physics courses, with data results showing a match between the students and test instruments used (Ibnu et al., 2019). The Rasch Model is an evaluation tool recommended for educators to measure and assess student learning outcomes accurately, revealing student's true abilities (Darmana et al., 2021).

The logit value (log odds unit) reveals student ability and the difficulty level a question. This value stems from the mathematical advancements by George Rasch, a Danish mathematician, who developed a method to measure the probability relationship between individual abilities and question difficulty using logarithmic functions, resulting in measurements with equal intervals (Sumintono & Widhiarso (2015). For question dichotomous answer choices, a combination of algorithms is used to express the expected probability results for item (i) and respondents (n) with the following formulation:

$$P_{ni}(X_{ni} = 1b_n, d_i) = \frac{e^{(b_n - d_i)}}{1 + e^{(b_n - d_i)}}$$

$P_{ni}(X_{ni} = 1b_n, d_i)$ represents the probability that respondent (n) will correctly answer item (i) (where x=1). A student's ability is determined by their success in answering question correctly, which is calculated as the difference between the respondent's ability and the question diffiulty level. The criteria determining the fit of items or identifying outliers or misfits are as follows: (1) the acceptable Outfit Means Square (MNSQ) value ranges from 0.5 to 1.5; (2) the acceptable Z-Standard Outfit (ZTSD) value ranges from -2.0 to +2.0; and (3) the acceptable Point Measure Correlation (Pt Measure Corr) value ranges from 0.4 to 0.85 (Boone et al., 2014).

RESULTS AND DISCUSSION

Question Item Analysis

The question difficulty level can be determined by looking at the data from the item measure output. Table 1 shows the information for each item. The table is sorted according to the question's difficulty level based on the logit value. The logit

value at the top is the question with the highest difficulty level. Conversely, the logit value at the bottom is the question with the lowest difficulty level. For example, question 21 (S21), with a logit value of 3.91, is the most difficult question for students because only 11 students could answer correctly.

Table 1. Problem Difficulty Level

Entry Number	Total Score	Measure	Item
21	11	3.91	S21
7	25	2.59	S7
4	44	1.32	S4
19	50	.91	S19
24	50	.91	S24
3	52	.77	S3
11	52	.77	S11
13	53	.69	S13
15	57	.38	S15
25	58	.29	S25
18	60	.11	S18
22	62	-.08	S22
17	63	-.19	S17
6	64	-.30	S6
8	65	-.42	S8
10	65	-.42	S10
23	65	-.42	S23
9	66	-.55	S9
16	67	-.70	S16
12	69	-1.03	S12
14	69	-1.03	S14
2	70	-1.24	S2
20	71	-1.49	S20
1	72	-1.80	S1
5	74	-2.95	S5

The question difficulty level is indicated by the question column (item). Table 1 shows that question 21 (S21) has the highest difficulty level, and question 5 (S5) is the question with the lowest difficulty level. Based on the difficulty level, S21 has 3.91 logits, and S5 has -2.95 logits. Therefore, S21 has almost seven times the difficulty level of S5. Based on the scores according to table 2, there were 11 students who answered question 21 correctly, while question 5 was answered correctly by 74 students. This means that almost all students answered question 5 correctly, or only one student answered incorrectly. According to the analysis of the difficulty level based on table 2, there were various factors that can affect a student's ability to correctly answer science questions which were not examined in this study. Several factors may influence a student's ability to answer these

questions such as misconceptions, working memory limitations, lack of knowledge of construction or learning experiences, or even psychological aspects like anxiety or confidence. Further research is needed to uncover the factors that influence students in answering the science exam. There were several factors that may affect the learning outcome and scientific literacy such as psychological factors, family factors, school factors, teachers, school or class facility, teaching materials (media), and also learning activity outside (Jufrida et al., 2019).

Table 2. Question Item Fit Level

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Soal
15	57	75	.38	.29	1.10	.77	1.93	2.92	A .15	.32	78.4	76.6	S15
1	72	75	-1.80	.60	1.05	.26	1.61	.90	B .05	.15	95.9	95.9	S1
12	69	75	-1.03	.44	1.13	.47	1.44	.86	C .04	.20	91.9	91.9	S12
21	11	75	3.91	.36	1.08	.40	1.28	.78	D .32	.41	87.8	87.0	S21
4	44	75	1.32	.26	1.10	1.10	1.24	1.57	E .29	.39	60.8	67.4	S4
23	65	75	-.42	.35	1.05	.28	1.22	.63	F .20	.25	86.5	86.5	S23
11	52	75	.77	.27	1.17	1.50	1.12	.66	G .22	.36	64.9	71.7	S11
13	53	75	.69	.27	1.11	.97	1.04	.26	H .27	.35	66.2	72.4	S13
18	60	75	.11	.31	1.11	.70	1.02	.17	I .22	.30	77.0	80.1	S18
2	70	75	-1.24	.47	1.04	.23	.94	.13	J .15	.19	93.2	93.2	S2
7	25	75	2.59	.27	.99	-.07	1.03	.26	K .43	.43	68.9	73.0	S7
24	50	75	.91	.27	1.01	.12	.89	-.53	L .38	.37	67.6	70.2	S24
19	50	75	.91	.27	1.00	.05	.98	-.07	M .37	.37	67.6	70.2	S19
20	71	75	-1.49	.52	1.00	.15	.83	.01	l .18	.17	94.6	94.6	S20
16	67	75	-.70	.39	.99	.05	.68	-.58	k .28	.23	89.2	89.2	S16
25	58	75	.29	.29	.95	-.27	.99	.06	j .35	.32	79.7	77.7	S25
10	65	75	-.42	.35	.94	-.18	.89	-.13	i .30	.25	86.5	86.5	S10
5	74	75	-2.95	1.01	.93	.25	.23	-.53	h .21	.09	98.6	98.6	S5
17	63	75	-.19	.33	.93	-.27	.81	-.44	g .34	.28	83.8	83.8	S17
9	66	75	-.55	.37	.90	-.33	.73	-.52	f .34	.24	87.8	87.8	S9
8	65	75	-.42	.35	.89	-.41	.60	-1.00	e .39	.25	86.5	86.5	S8
14	69	75	-1.03	.44	.89	-.24	.54	-.77	d .34	.20	91.9	91.9	S14
6	64	75	-.30	.34	.88	-.52	.65	-.91	c .40	.27	85.1	85.1	S6
22	62	75	-.08	.32	.85	-.79	.70	-.87	b .42	.28	85.1	82.5	S22
3	52	75	.77	.27	.83	-1.53	.72	-1.47	a .51	.36	75.7	71.7	S3
MEAN	58.2	75.0	.00	.38	1.00	.1	.97	.1			82.1	82.9	
P.SD	14.3	.0	1.37	.16	.09	.6	.35	.9			10.6	9.0	

The quality of the items with the model or item fit is presented in Table 2. This item fit determines whether or not the items used function normally when making measurements. When questions do not fit, it suggests that there might be misconceptions among students regarding the questions being asked. Based on data in table 2, the average logit value of the items is 0.0 logit which indicates that the instruments holistically can measure and if not 0.0 the item is not good as a whole (Misbach & Sumintono, 2014) based on table 2, the logit number of items from the average and standard deviation is 1.00 + 0.09 = +1.09, so based on this criterion, there are 6 items that have a greater INFIT MNSQ value. They are items number 15, 12, 4, 11, 13 and 18 (S15, S12, S4, S11, S13 and S18). These items need to be discarded.

The information about item fit is very important and useful for teachers. Teachers can improve the quality of their teaching so that these misconceptions can be minimized or even avoided. Enhancing or even replacing questions is necessary if questions do not meet all the criteria. However, if the question does not meet only one of the criteria, then the question can still be used without improvement.

One of the requirements for a measurement to be valid is to look at the "bias" of the instrument or question items used. Instruments or questions are said to be biased if one respondent is "more favored" compared to other respondents. For example, items that are easier to answer by male students compared to female students or vice versa. This is called a question that contains gender bias (Sumintono & Widhiarso, 2015). The instrument is biased if the item's probability value is below 5% or 0.05.

Table 3. Detection of Item Bias through Differential Item Functioning Data

Prob.	Stem Number	Name	Prob.	Stem Number	Name
0,6222	1	S1	0,5978	1	S1
0,3893	2	S2	0,3804	2	S2
0,8504	3	S3	0,8582	3	S3
0,6481	4	S4	0,6763	4	S4
0,5402	5	S5	0,8170	5	S5
0,5416	6	S6	0,5423	6	S6
0,5349	7	S7	0,5911	7	S7
0,3822	8	S8	0,3816	8	S8
0,2475	9	S9	0,2496	9	S9
0,9114	10	S10	0,9118	10	S10
0,8504	11	S11	0,8582	11	S11
0,2212	12	S12	0,1843	12	S12
0,6791	13	S13	0,6947	13	S13
0,6225	14	S14	0,6114	14	S14
0,3804	15	S15	0,3959	15	S15
0,5188	16	S16	0,5084	16	S16
0,5722	17	S17	0,5731	17	S17
0,8946	18	S18	0,8958	18	S18
0,8626	19	S19	0,8690	19	S19
0,9207	20	S20	0,9103	20	S20
0,9123	21	S21	0,9200	21	S21
0,5612	22	S22	0,5664	22	S22
0,9114	23	S23	0,9118	23	S23
0,8324	24	S24	0,8426	24	S24
0,5772	25	S25	0,5904	25	S25

Based on the data in Table 3, it can be explained that the item probability data shows a number above 5%. This indicates that all items in the instrument do not contain gender bias.

Student Abilities Analysis

Analyzing students' abilities in answering the given questions helps teachers assists the student learning process more effectively (Sumintono & Widhiarso, 2015). The level of individual ability can be seen from the response pattern. This

can help teachers determine the consistency of their students' thinking (Sumintono & Wahyu Widhiarso, 2014). The results of analyzing students' abilities give teachers insight that enable them to support the teaching and learning process more effectively.

Table 4. *Person Measure*

Entry Number	Total Score	Measure	Person	Entry Number	Total Score	Measure	Person
49	25	5,59	49P7e	2	19	1,45	02L7b
8	24	4,16	08P7b	3	19	1,45	03L7b
42	24	4,16	42L7d	13	19	1,45	13P7c
52	24	4,16	52P7e	14	19	1,45	14P7c
17	23	3,20	17P7c	19	19	1,45	19L7c
36	23	3,20	36P7d	20	19	1,45	20L7c
44	23	3,20	44P7d	25	19	1,45	25L7d
53	23	3,20	53L7e	40	19	1,45	40L7d
54	23	3,20	54L7e	45	19	1,45	45L7d
71	23	3,20	71P7k	58	19	1,45	58L7e
74	23	3,20	74P7k	61	19	1,45	61L7k
4	22	2,59	04P7b	66	19	1,45	66L7k
11	22	2,59	11P7c	72	19	1,45	72L7k
41	22	2,59	41L7d	16	18	1,17	16L7c
59	22	2,59	59L7e	32	18	1,17	32P7d
60	22	2,59	60P7e	48	18	1,17	48L7e
67	22	2,59	67P7k	10	17	0,92	10L7c
75	22	2,59	75P7k	23	17	0,92	23L7d
6	21	2,13	06P7b	29	17	0,92	29P7d
12	21	2,13	12P7c	35	17	0,92	35P7d
18	21	2,13	18L7c	47	17	0,92	47L7e
24	21	2,13	24L7d	57	17	0,92	57L7e
27	21	2,13	27L7d	63	17	0,92	63L7k
31	21	2,13	31P7d	5	16	0,69	05L7b
43	21	2,13	43P7d	26	16	0,69	26P7d
50	21	2,13	50L7e	28	16	0,69	28P7d
51	21	2,13	51P7e	37	16	0,69	37P7d
55	21	2,13	55L7e	38	16	0,69	38P7d
56	21	2,13	56L7e	70	16	0,69	70L7k
73	21	2,13	73L7k	39	15	0,47	39P7d
7	20	1,76	07P7b	68	15	0,47	68L7k
21	20	1,76	21P7d	69	15	0,47	69P7k
30	20	1,76	30P7d	9	14	0,25	09P7c
34	20	1,76	34P7d	33	14	0,25	33P7d
62	20	1,76	62P7k	15	12	-0,16	15P7c
64	20	1,76	64P7k	22	12	-0,16	22P7d
65	20	1,76	65P7k	MEAN	19,4	1,78	

Entry Number	Total Score	Measure	Person
1	19	1,45	01P7b
46	18	1,17	46P7d

Entry Number	Total Score	Measure	Person
P.SD	2,9	1,06	

Student abilities can be seen in the person column sorted from the highest to the lowest. Based on Table 4, the student with the highest ability is 49P7e, with a logit value of 5.59. while the student with the lowest ability is 22P7d with a logit value of -1.6. A high logit value indicates a strong problem-solving ability while a low logit value indicates a weaker problem-solving ability. For example, student 49P7e has the highest ability to do all questions correctly. Students who have the same logit value have the same measure value. The measured column is each student's logit value, which can be used to compare students' abilities in answering questions.

Based on the data in Table 4, student 49P7e has a very unfit or unusual response compared to other students. Information on unusual response patterns can be known more deeply through the data shown by the scalogram in Figure 1.

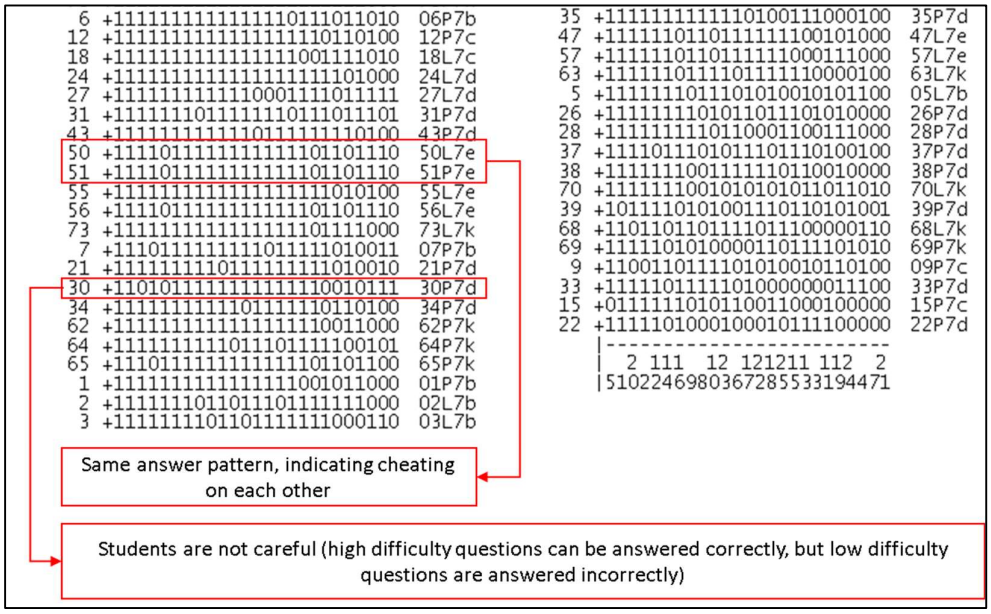


Figure 1. Scalogram

Based on Figure 1 (Scalogram), it can be explained that 2 respondents indicated cheating, namely respondents with code 50L7e (student with sequence number 50 is male from class 7E) and respondents with code 51P7e (student with sequence number 51 is female from class 7E). The two students most likely worked together by sitting close to each other because they had the same answers. In addition, the respondent with the code 30P7d (student with a serial number 30 of female gender from class 7D) is a respondent who is not careful. 30P7d could

answer questions that other respondents categorized as very difficult to answer correctly, while those considered the easiest could not be answered correctly.

Wright Map Analysis

The analysis of item validity using the Rasch Model was conducted through the Winsteps application, with the Wright Person Item Map output results shown in Figure 2.

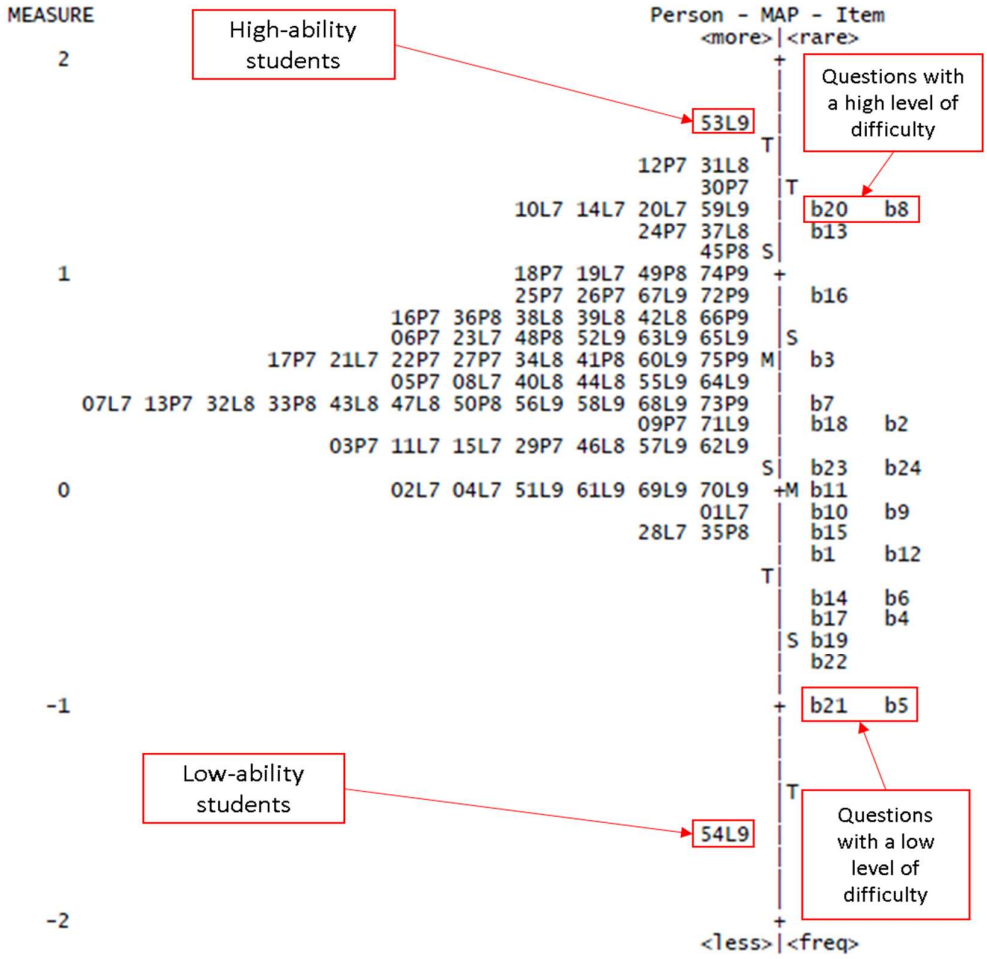


Figure 2. Wright Person Item Map

The Wright map based on Figure 2 illustrates the distribution of student abilities and the distribution of the difficulty level of the questions using the same scale. The Wright map is a comprehensive person-item map. The left map depicts the student ability map as seen from the presence of one student with high ability, namely 49P. The logit value of this student is more than +4 logits. 49P is also outside the standard deviation limit (T), indicating a distinct high intelligence or outlier. On the other hand, the least able students were 15P and 22P with logits below 0, indicating very low ability (outliers).

Summary Statistics

Rach modeling presents data on person, item, and instrument reliability. Person reliability values are utilized to assess the consistency of respondent's answers. Item reliability is employed to evaluate the quality of statement items within the instrument. Meanwhile, instrument reliability, or test reliability, is used to understand the interaction between respondents and the statement items. Summary statistics in the Rasch Model provide concise descriptions of key aspects of the data analyzed using this model. These statistics typically include measures such as person reliability, item reliability, and instrument reliability.

SUMMARY OF 25 MEASURED (NON-EXTREME) Soal

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	58.2	75.0	.00	.38	1.00	.11	.97	.06
SEM	2.9	.0	.28	.03	.02	.13	.07	.19
P.SD	14.3	.0	1.37	.16	.09	.62	.35	.91
S.SD	14.6	.0	1.39	.16	.10	.63	.36	.93
MAX.	74.0	75.0	3.91	1.01	1.17	1.50	1.93	2.92
MIN.	11.0	75.0	-2.95	.26	.83	-1.53	.23	-1.47
REAL RMSE	.41	TRUE SD	1.30	SEPARATION	3.15	Soal	RELIABILITY	.91
MODEL RMSE	.41	TRUE SD	1.30	SEPARATION	3.19	Soal	RELIABILITY	.91
S.E. OF Soal MEAN = .28								

Soal RAW SCORE-TO-MEASURE CORRELATION = -.96
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 3. Summary Statistics

According to figure 3, it can be explained that the reliability value of the statement items (butir) is obtained at 0.91, indicating that the statement items fall into the category of excellent quality. Hence, the instrument can be utilized for seventh-grade odd semester final examinations.

CONCLUSION

Multiple-choice questions analyzed using the Rasch Model can provide a measurement scale with equal interval. Raw scores certainly cannot be used directly to interpret student ability. This Rasch modeling uses respondent and item score data together in an interaction. Respondent and item scores form the basis for estimating the original score, which indicates the respondents' abilities and difficulty levels of the items. The results of the Rasch Model analysis show that students with the highest ability are 49P to the lowest student abilities, namely 15P and 22P. The results of the Rasch analysis showed the most difficult questions and the easiest questions. The difficulty level of the question from the most difficult is S21, and the easiest question is S5. Further research is needed to find out what factors affect the difficulty or ease of working on these questions. According to Rasch analysis, question number 4,11,12,13,14,15, and 18 had low validity and

should be discarded. We recommend using the Rasch model in analyzing science test items to evaluate the quality of the questions provided to students. Items identified infit can be promptly revised or replaced. Additionally, the use of the Rasch model can determine the possibility of students cheating during tests. Thus, teachers can improve the quality of questions used for daily tests, midterm exams, or end-of-semester exams.

REFERENCE

Abdullah, N., Noranee, S., & Mohd Rahim Khamis. (2017). The Use of Rasch Wright Map in Assessing Conceptual Understanding of Electricity. *Pertanika Journal of Social Science & Humanity*, *25*(5), 81–88.

Amrianto, Z., Zein, A., & Darussyamsu, R. (2019). Analisis Soal Ujian Tengah Semester II Mata Pelajaran IPA Kelas VIII SMP Pembangunan Laboratorium Universitas Negeri Padang dan SMP Pertiwi Siteba Padang Tahun Pelajaran 2015/2016. *Center for Open Science*, 1–9. https://doi.org/http://dx.doi.org/10.31227/osf.io/c9hg8

Arikunto, S. (2021). *Dasar-dasar Evaluasi Pendidikan* (3rd ed.). Bumi Aksara.

Azizah, & Sapti Wahyuningsih. (2020). Penggunaan Model Rasch untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika,* , *3*(1), 45–50.

Bond, T. G., & Fox, C. M. (2013). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Psychology Press.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Science+Business Media Dordrecht.

Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, & Nurhaswinda. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *JPSI: Jurnal Pendidikan Sains Indonesia*, *329*(345).

Elgadal, A., & Mariod, A. A. (2021). Assessment Tool For Quality Assurance Measures. *Sudan Journal of Medical Sciences*, *16*(3), 334–346.

Erfan, M., Maulyda, M. A., & Hidayati, V. R. (2020). Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Tes Klasik dan Model Rasch. *Indonesian Journal of Educational Research and Review*, *3*(10), 11–19.

Ginting, C. R., Ds, D. B., Y. Miranti N., & N. Nurmasyitah. (2020). Analisis Kualitas Butir Soal UTS Mata Pelajaran IPA Tahun Pelajaran 2019/2020 Kelas VII di SMPN 6 Kota Langsa. *Jurnal Pendidikan Fisika Dan Sains, 3*(1), 29–32.

Ibnu, M., Indriyani, B., Inayatullah, H., & Guntara, Y. (2019). Aplikasi Rasch Model: Pengembangan Instrumen Tes untuk Mengukur Miskonsepsi Mahasiswa pada Materi Mekanika. *Prosiding Seminar Nasional Pendidikan FKIP*, 205–2010.

Jufrida, Basuki, F. R., Pangestu, M. D., & Prasetya, N. A. D. (2019). Analisis Faktor yang Mempengaruhi Hasil Belajar IPA dan Literasi Sains di SMP Negeri 1 Muaro Jambi. *Edu Fisika: Jurnal Pendidikan Fisika*, *4*(2), 31–37.

Kemendikbudristek. (2022). *Kurikulum Merdeka*. Kemendikbudristek.

Khoirurrijal, Fadriati, Sofia, Makrufi, A. D., Gandi, S., Muin, A., Tajeri, Fakhrudin, A., Hamdani, & Suprapno. (2022). *Pengembangan Kurikulum Merdeka*. Literasi Nusantara Abadi.

Kurniasi, E. R., Y, Y., & Karennisa, F. (2020). Analisis soal ulangan harian matematika kelas IX SMP Negeri 1 Toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara,*, *12*(1), 43–52.

Kusumawati, Mutiara, & Hadi, S. (2018). An analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from Mathematics Assessments in Senior High School. *ReiD: Research and Evaluation in Education*, *4*(1), 70–78.

Linacre, J. M. (2016). *A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*. IL: Winsteps.Com, Chicago.

Mahmud, Z., Ghani, N. A. M., & Rahim, R. A. (2013). Assessing Students' Learning Ability In A Postgraduate Statistical Course: A Rasch Analysis. *Procedia - Social and Behavioral Sciences*, *89*, 890–894.

Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsi, E. (2023). Implementasi Rasch Model dalam Pengembangan Instrumen Tes Diagnostik. *Prosiding Seminar Nasional Pascasarjana*, 748–757.

McKenna, P. (2018). Multiple Choice Questions: Answering Correctly and Knowing the Answer. *Proceedings of the IADIS International Conference E-Learning*, 105–114.

Misbach, I. H., & Sumintono, B. (2014). Pengembangan dan Validasi Instrumen "Persepsi Siswa Tehadap Karakter Moral Guru" di Indonesia dengan Model Rasch. *Seminar Nasional Psikometri, UMS*, 148–172. http://hdl.handle.net/11617/6409

Septiliana, L. (2023). Analisis Item Soal dengan Menggunakan Rasch Model sebagai Ukuran Kualitas Madrasah Ibtidaiyah Pada Mata Pelajaran IPA Lilla Septiliana. *Jurnal Pendidikan*, *12*(2), 1–12.

Sudjana, N. (1995). *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya.

Sumintono, B. (2016). Penilaian Keterampilan Berpikir Tingkat Tinggi: Aplikasi Pemodelan Rasch pada Asesmen Pendidikan. *Seminar Nasional Pendidikan IPA*.

Sumintono, B., & Wahyu Widhiarso. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Penerbit Trim Komunika.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Penerbit Trim Komunika.

Syahroni, M., Nurhasanah, A., & Putra, A. P. (2024). Analysis of Multiple Choice Questions as an Evaluation Tool for the end of Year Assessment (PAT) in the Subject of Indonesian History Class XI SMA. *Historia: Jurnal Program Studi Pendidikan Sejarah*, *12*(1), 143–154. https://doi.org/https://doi.org/10.24127/hj.v12i1.8640

Valen, A., & T. G. Satria. (2021). Analisis Tingkat Kesulitan Soal PAS (Penilaian Akhir Semester) Mata Pelajaran IPS di Sekolah Dasar. *Jurnal Basicedu*, *5*(4), 2199–2208.

Zainal, N. F. (2020). Pengukuran, Asesmen, dan Evalusi dalam Pembelajaran Matematika. *Laplace: Jurnal Pendidikan Matematika*, *3*(1), 8–26.

Zakso, A. (2022). Implementasi Kurikulum Merdeka Belajar di Indonesia. . . *J-PSH: Jurnal Pendidikan Sosiologi Dan Humaniora*, *13*(2), 916–9622.

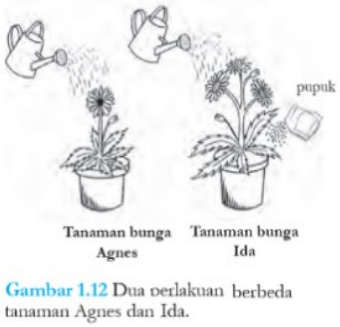
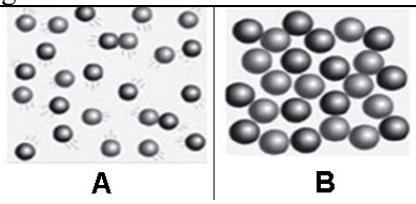
APPENDIX
SUMMATIVE PRE-ASSESSMENT QUESTIONS
(SCIENCE GRADE 7TH)

Instructions:

- Fill in your personal data completely
- Put a **check mark** (✓) on the statement that you think is true or false.

Full Name	
Class	
School	
Date of Test	
Time Start of test	

No	Item Tests	Correct	Incorrect
1	I am the science on which other sciences are based. I am called the "Mother of Science". I am the philosophy or parent of science.		
2	One of Prof. B.J. Habibie's findings is the Crack Propagation Theory, which is how mechanics detect damage to aircraft construction.		
3	 The diagram shows an Erlenmeyer flask.		
4	To determine the aroma of a substance, you can directly inhale the substance in a test tube or beaker.		
5	A science laboratory is a place to conduct experiments, investigations, measurements, or scientific training.		
6	A hypothesis is a temporary answer or conjecture in an experiment.		
7	The dependent variable is the variable that affects the independent variable.		
8	The final stage of an experiment is concluding.		
9	Everything that is measured is called a magnitude. At the same time, everything that is used as a comparison of a magnitude is a unit.		
10	To measure mass, you can use a tape measure.		

No	Item Tests	Correct	Incorrect
11	 The graph is illustrate data from the table.		
12	 Gambar 1.12 Dua perlakuan berbeda tanaman Agnes dan Ida. The hypothesis of the experiment was that plants given fertilizer would tend to be more fertile than plants given water only.		
13	Liquid substances have a very high particle density compared to solids.		
14	Perfumed oil smelled from a distance is an example of gaseous diffusion.		
15	 A is the shape of a liquid particle, while B is a gaseous particle.		
16	Water flows from low to high places.		
17	Solid particles will stretch when a temperature rise occurs.		
18	The higher the temperature, the stronger and more stretchable the motion of the particles.		
19	Iron with a melting point of 1535^{O}C requires a higher temperature to transform than gold at 1064^{O}C.		

No	Item Tests	Correct	Incorrect
20	All things that burn undergo chemical changes.		
21	The volume of water in the world increases during the rainy season and decreases during the dry season.		
22	The presence of bubbles and sediment is one of the characteristics of chemical changes.		
23	Mr. Joko turning teak wood into a guest table is an example of a physical change.		
24	If the density (ρ) of the object is 100 g/cm^3 and the volume is 20 ml, then the mass of the object is 2 kg.		
25	An object will float in a liquid when it has the same density as the liquid.		

*Item number 4, 11, 12, 13, 15, and 18 need to be discarded