Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 9, No. 2: November 2025
Hal. 208-233

Design of Assessment Instrument for Mathematical Critical Thinking Indicators-Based Essay Questions for Vocational and High School Students Using the Rasch Model

Jarot Rudi Hartato[1)*], Nizaruddin[2)], Lukman Harun[3)]

Postgraduate Program in Mathematics Education, Universitas PGRI Semarang, Semarang, Indonesia[1,2,3)]

*Email Corresponding: jarotrudih@gmail.com

Abstract. This research aims to create a measurement tool for Critical Thinking Skills (CTS) in mathematics. The instrument was analysed using the Rasch model. The analysis was conducted to determine the results of the content validity test (item suitability), reliability test, dimensionality test (construct validity), person separation test, and item separation, and to analyse the items' difficulty level. This research uses the R&D model with the 4D model. Data collection methods through interviews with two teachers, literature studies to determine CTS indicators, questionnaire validation of CTS test instruments, and CTS test questions given to students. The sampling technique used was saturated sampling. The indicators of CTS are Interpretation, analysis, evaluation, and inference. The results for each instrument item are valid; construct validity is good, reliability is sufficient, and Cronbach's alpha is 0.68. The results of the person separation analysis show that students are divided into two groups, and item separation shows that the questions are divided into five groups. The measure (logit value) analysis results show that the interpretation indicator question is easy, the analysis indicator question is very difficult, the evaluation indicator question is moderate, and the inference indicator question is difficult. This research fills the literature gap by developing CTS indicator-based instruments validated by the RASCH approach, which has not been widely applied in SMA / SMK.

Keywords: Critical Thinking, Test Instrument, Rasch Model

INTRODUCTION

Mathematics is a field of study in Indonesia's education system. Learning and understanding mathematics starts at the most basic level up to university (Anderha dan Maskar, 2021). Even at the level of PAUD and kindergarten education, it has been equipped with CTS, objective, logical, and careful through learning mathematics at school (Jannah et al. 2023; Rosita et al. 2019). The description shows that mathematics is one of the lessons that must be learned at all levels of education.

Education is an important aspect of improving the nation's quality. This statement aligns with the opinion that education's role is to realise and guide humans so that they can think critically and ideally. If education is carried out properly, it will produce an advanced nation. This statement shows that education has a crucial role in the life of a country; the progress or decline of a country is closely related to the quality of education provided. An educator plays a significant role in teaching and learning activities. In addition to planning and implementing learning, the task of an educator is also to determine the assessment process of learning outcomes through evaluation activities.

Evaluation is an important component in the learning process. Evaluation is the process of assessing something. In education, evaluation can be done on student learning outcomes. While learning outcomes are the results obtained by students after the learning process. So it can be concluded that the evaluation of learning outcomes is a series of assessment activities that aim to determine students' level of understanding after participating in the learning process. This activity is carried out as an effort to achieve the learning objectives that have been set. According to Hadiyanti et al. (2024), evaluating learning outcomes benefits students and guides them during the learning process. One of the benefits of evaluation for educators is to plan and implement the next learning program in line with the objectives to be achieved.

One of the evaluations that educators need to do on learning outcomes is students' mathematical competence level. This statement illustrates the importance of learners having various skills, as expressed by Sri Hanipah (2023), including problem solving, creativity, critical thinking, communication, collaboration, digital literacy, and social-emotional skills. The purpose of mastering these skills is to help learners adapt to the rapid development of technology and information. The opinion of (2020) supports this, stating that the learning paradigm has also changed along with the times. In this context, learners need to be equipped with relevant skills to face the challenges of the modern world. Learners must be creative, innovative, think critically, collaborate, and understand technology well. Therefore, mathematical critical thinking is one of the abilities that need to be mastered by students (Rizti & Prihatnani, 2021; Pratama & Mardiani, 2022; Irfiani et al., 2023).

In connection with this, an educator needs to evaluate students' learning outcomes towards the level of students' mathematical CTS.

The definition of CTS revealed by Muliana (2021) is crucial in everyday life. These skills help us solve problems by analysing, interpreting, and evaluating information to make reliable decisions. Another opinion expressed by Ajizah & Putu Artayasa (2022) states that critical thinking involves formulating solutions to problems that involve deep understanding. The mathematical CTS defined by Lutfiah et al. (2023) is an ability that uses logical reasoning to always remember the process and results of learning to collect, analyse, and use data effectively and efficiently. Another definition is also expressed by Zain & Marhayati (2024), which is that CTS refers to how well a person can think deeply to solve problems. Based on the explanations from experts, it can be concluded that CTS in mathematics is the ability to apply logic in collecting, analysing, and utilising information. This process aims to reach accurate decisions in solving various mathematical problems.

A tool is needed to evaluate students' mathematical CTS in the form of an assessment instrument. An instrument assesses a specific object or collects information about a variable, provided the instrument follows predetermined academic criteria. According to Ramadhan et al. (2024), a measuring instrument can be considered valid if the tool can accurately measure what should be measured. Therefore, to get a quality measuring instrument, it is important to test its validity and reliability.

One of the measuring tools that can be used to measure the level of mathematical CTS of students is an instrument in the form of a description test. According to Sahira & Penggabean (2022), test questions play a role in reviewing the ability of students. Items must go through item analysis to be feasible and qualified to measure students' mathematical CTS. This statement is supported by the opinion of Parisu et al. (2024); Fauziana & Wulansari (2021); the use of item analysis can help educators to improve the quality of questions by revising or removing less effective questions, and it can help educators to diagnose how well students understand the material that has been taught. One of the statistical methods that can be applied to analyse items is the Rasch model approach.

The Rasch approach is a statistical approach that can be used to evaluate test data and each item's reliability and validity (Parisu et al., 2024). According to Wulandari et al. (2025), the Rasch model has the advantage that missing data is predicted and analysed using individual responses. In addition, the Rasch model provides an alternative approach to using raw test scores or data. Therefore, if the Rasch model measure is applied to raw data from test results, it tends to produce a similar interval measurement scale, thus providing precise information regarding learners' capacity and the quality of questions presented by educators. The opinion expressed by Erfan et al. (2020) using the Rasch model analysis will produce item characteristic information. The Rasch model was first developed by Georg Rasch in the 60s and popularised using raw data created by Ben Wright, consisting of dichotomous data in the form of true and false (Chan et al., 2014). According to Erfan et al. (2020), this model has become popular as a tool often used in educational research.

Research conducted by Wulandari et al. (2025) titled "Analysis of Mathematics Knowledge Items for Grade V Elementary School Using the Rasch Model" discusses the results of the analysis of mathematical knowledge items using the Rasch model. The questions analysed were multiple-choice questions of 10 items. The research results obtained were eight questions declared valid; besides that, the reliability obtained was quite good, at 0.65. Furthermore, this study was able to show the level of difficulty of the question.

Furthermore, Parisu et al. (2024) conducted other research titled "Analysis of Mathematics Basic Knowledge Items Using the Rasch Approach". The research was conducted to improve the quality of tests and provide diagnostic information about students' understanding by analysing the items. In addition, the purpose of the research was to produce a quality test device to measure students' basic knowledge of mathematics. The test instrument analysed was 25 multiple-choice questions. The results obtained were twenty-four items declared valid, and the instrument's reliability was in a good category, namely 0.94. In addition, this study produces a classification of the items' difficulty level.

Previous studies conducted by Wulandari et al., 2025; Parisu et al., 2024 have tested items testing the quality of items to measure students' ability. In the study, the questions analysed were multiple-choice and did not explain specifically what abilities would be measured. Furthermore, no one has validated description questions based on indicators of critical thinking skills that Karim & Normaya (2015 have put forward, namely Interpretation, analysis, evaluation, inference for SMA / SMK students using the RASCH approach.

The Rasch model was chosen because it is a statistical method for assessing item reliability and validity and classifying item difficulty. This research focuses on the analysis of mathematical knowledge description items distributed in class X at SMK YPE Nusantara Slawi and SMA Ihsaniyah Kota Tegal to obtain test instruments to measure quality mathematical CTS so that later it can be used to measure the mathematical CTS of SMK / SMA level students in class X.

Based on the problem description above, the researcher wants to know whether the description question instrument, based on the indicators of mathematical critical thinking skills that have been made, meets the criteria of validity and reliability based on the Rasch model.

RESEARCH METHOD

Type of Research

This type of research is Research and Development (R&D) with the Define, Design, Development, and Disseminate model, with restrictions only until the Development stage. The flow of R&D in this study can be seen through Figure 1, as follows:

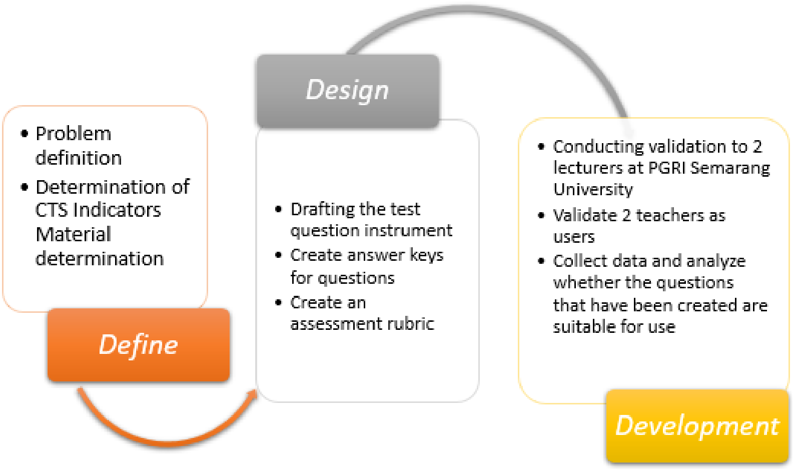

Problem definition
Determination of CTS Indicators Material determination
Define
Design
Drafting the test question instrument
Create answer keys for questions
Create an assessment rubric
Conducting validation to 2 lecturers at PGRI Semarang University
Validate 2 teachers as users
Collect data and analyze whether the questions that have been created are suitable for use
Development

Figure 1. Research Flow

Participants

This research was conducted at SMK YPE Nusantara Slawi and SMA Ihsaniyah Tegal City. The research subjects comprised 51 students from SMK YPE Nusantara Slawi and 29 from SMA Ihsaniyah Tegal City. The sampling technique used was saturated sampling; the research sample used was all members of the population in the observation class. The reason for choosing the saturated sampling technique is that researchers want to include the population that is the object of research, so that it can produce accurate research results. The tool used in this research is the mathematical CTS test instrument on arithmetic rows and series material. This research was conducted from November 15, 2024, to December 15, 2024.

Data collection method

This data collection method includes interviews with mathematics teachers at the research site, library research, questionnaires, and assessment of student answers in working on CTS indicator-based questions that have been developed. Library research was conducted to obtain information regarding CTS indicators that many experts and researchers have proposed and to consider which indicators to use. Furthermore, a questionnaire was conducted to conduct expert validation. Assessment of student answers based on the assessment rubric has been made and has passed through expert and user validation.

Data Analysis Technique

Data analysis in this study used a qualitative approach to describe the feasibility of the instruments that had been made. Furthermore, a quantitative approach is used to determine the validity and reliability of the questions after collecting data on the research sample—quantitative analysis with the Rasch approach assisted by WinStep 3.73 software.

Explanation of Research Flow

At the define stage, researchers identified problems that took place in the learning process so that the content of the test instrument was adjusted and developed. This research was carried out for 4 weeks, starting from the design stage, namely, the researcher designed a draft instrument of mathematical CTS test questions on the material of arithmetic rows and series. Furthermore, the test question instrument was validated by two experts, namely lecturers from PGRI Semarang University, one teacher from SMK YPE Nusantara Slawi, and one teacher from SMA Ihsaniyah Tegal City. Furthermore, the development stage is that the researcher adjusts the mathematical CTS test instrument according to the direction on the validation sheet that has been given to experts and teachers, after revision, the test question instrument is given to students who have been selected as research subjects and analyzes the results of student artistry tests using Rasch model analysis with the Winstep version 3.73 application.

The instrument used is a mathematical CTS question that refers to the CTS indicators put forward by Karim & Normaya (2015), namely Interpretation, analysis, evaluation, and inference. Researchers refer to these indicators because they are considered easy; besides that, these indicators can quantify the mathematical CTS of students. Each indicator is made up of one question. Then the instrument was given to 2 experts and 2 math subject teachers. The validator profile can be seen through Table 1 as follows:

Table 1: Validator Profile

No	Name	Academic Degree	Institution	Field of Expertise	Academic Position
1.	Dr. FX. Didik Purwosetiyono, M.Pd	Doctor of Education	UPGRIS	Critical Thinking and Educational Assessment	Lector
2.	Dr. Ida Dwijayanti, M.Pd	Doctor of Education	UPGRIS	Curriculum and TPACK	Head Lector
3.	Hanifathul Hidayat, M.Pd	Master of Education	SMA Ihsaniyah Kota Tegal	Mathematics Education	Teacher
4.	Resmiyati, S.Pd	Bachelor of Education	SMK YPE Nusantara Slawi	Mathematics Education	Teacher

The data from the assessment by the validator is then analysed using qualitative descriptive techniques, using percentages by calculating the score achieved from all indicators with the following formula:

$$N = \frac{\text{Assessment Result Score}}{Maximum\ Score} \times 100\%$$

By using the formula above, the category of expert validation results can be seen through Table 2, as follows:

Table 2: Criteria for Expert Validation Results

Problem Item	Category
$86\% \leq N < 100\%$	Very good
$72\% \leq N < 85\%$	Good
$58\% \leq N < 71\%$	Fair
$44\% \leq N < 57\%$	Less
$N \leq 44\%$	Very Poor

In the instrument validation process, researchers received criticism and suggestions from validators. The suggestions on the questions that have been made are from the language writing system, which is considered not in accordance with the EYD writing rules, and some questions are still considered unclear. The researcher makes revisions to adjust the questions according to the validator's suggestions. Then, on the assessment rubric that researchers have made, there are many adjustments to the assessment points, so that researchers can change many assessment points so that it is feasible to use. Furthermore, after being assessed as feasible by the validator, the researcher conducted an initial trial at SMK YPE Nusantara Slawi, precisely in class X MPLB 1, consisting of 18 students. From the results of the instrument field test, almost all students had difficulty working on the questions that were asked. These results were considered inaccurate by the researcher, so the researcher took the whole dataset to get accurate data and analysed it using the Rasch model, assisted by Winstep software.

By using the Rasch model, the test results obtained are then analysed using the help of Winstep software version 3.73 to determine the interaction between respondents and statement items and detect by producing a logit value, which is used to calculate the value because it can reflect the probability of items from a group of respondents. According to Wibisono, 2016 in Muntazhimah et. al., (2020) states that in the Rasch model, raw data cannot be directly analyzed, but must first be converted into the form of an 'odds ratio' to then carry out a logarithmic transformation into logit units as a manifestation of the respondent's probability of responding to an item. The resulting data is then analysed based on aspects of item suitability (content validity), instrument reliability, person and item separation, unidimensionality (construct validity), and item difficulty analysis.

In Rasch model analysis, item fit is evaluated to assess how well the items contribute to the measurement of the underlying construct. Through the output of the Winstep software, we can obtain a number of item parameters that fit the Rasch model, as well as Cronbach's Alpha values that show the overall reliability test results. According to Sumintono & Widhiarso (2015), the criteria for checking item suitability (content validity) and unsuitable individuals are evaluated through several indicators, namely Outfit MNSQ (Mean-Square), Outfit ZSTD (Z-Standard), and item correlation values or Pt Measure Corr (Point Measure Correlation). The criteria are as follows:

a. MNSQ Outfit value is accepted if it lies in the range $0.5 < \text{MNSQ} < 1.5$

b. The ZSTD Outfit value is accepted if it lies in the range of $2.0 < \text{ZSTD} < 2.0$.

c. Pt. Measure Corr is accepted if it lies in the range $0.4 < \text{Pt. Measure Corr} < 0.85$

Instrument items are considered suitable if they have at least two of these requirements; otherwise, if they are below two of the requirements, then they are considered unsuitable (Sumintono, 2018). As for the value of seeing the value of item dimensionality (construct validity), some criteria have been determined according to Sumintono & Widhiarso (2015), namely, as Raw variance value explained by measures & unexplained variance in contrasts $\mathbf{1^{St} - 5^{th}}$, the criteria used can be seen through table 3 as follows:

Table 3. Construct Validity Categories

No	*Raw Variance explained by measures*	Category	*Unexplained variance in contrasts* $\mathbf{1^{St} - 5^{th}}$	Category
1	$< 20\%$	Bad	$> 15\%$	Bad
2	$\geq 20\%$	Fair	$10 - 15\%$	Fair
3	$\geq 40\%$	Good	$5 - 10\%$	Good
	60%	Very Good	$< 5\%$	Very Good

Source : Sumintono & Widhiarso, 2015 in Ngadi, (2023)

The unidimensionality test is used to evaluate the construct validity of the Rasch model. In this test, there is only one hidden attribute that underlies the respondent's answer to the item (Lord, 1980, in (Ridho, 2011)) and evaluates the interaction between the respondent's answer to the item. In the unidimensionality test, the value of the raw variance measures is at least 20%. However, for the case of unexplained variance, the quality of the instrument is categorised as good if it has a value of less than 15% (Sumintono & Widhiarso, 2014, in (Ngadi, 2023)).

Respondents provide answers regarding the concepts expressed through statements or answers in the description items. These items form the dimensions of a variable and are arranged in the form of a questionnaire (Sebariani et. al., 2023). In this study, the questionnaire was replaced with a rubric to assess the results of students' answers when working on the instrument of students' mathematical CTS test questions. The following is Table 4, person and item reliability, which serves as the basis for determining the reliability of both people and items:

Table 4. Person and Item Reliability

Value	Category
> 0,94	Excellent
0,91 – 0,94	Very good
0,81 – 0,90	Good
0,67 – 0,80	Fair
< 0,67	Poor

Source : Sumintono & Widhiarso, 2015 in (Ngadi, 2023)

Table 5. Cronbach's Alpha Categories

Value	Category
> 0,8	Very good
0,7 – 0,8	Good
0,6 – 0,7	Fair
0,5 – 0,6	Poor
< 0,5	Very Bad

Source : Sumintono & Widhiarso, 2015 in (Ngadi, 2023)

Person and Item Separation Index serves to estimate the tools obtained that differentiate the skills of learners. Item quality is also tested through the value of person and item separation (Sumintono & Widhiarso, 2014; Wati & Mahtari, 2017) with the aim of determining the grouping map (strata) of respondents and items. According to Akthar (2017a), in (Ngadi, 2023), the greater the value of person-item separation, the better because person and item separation have a long range or are more varied. There are values for the separation index that range from 0 to infinity; the higher the separation, the better. The criteria for the person index and item separation are shown in Table 6 as follows:

Table 6. Separation Index

Person and Item Strata Index	Criteria
>5	Special
3 – 4	Very good
2 – 3	Good
>= 1,5	Acceptable
< 1,5	Unacceptable

Rasch modelling divides item difficulty into four groups based on the size value (logit) and the standard deviation (SD) value of the logit item. The following are the criteria for item difficulty shown in Table 7, as follows:

Table 7. Criteria for Level of Difficulty

Measure (logit)	Interpretation of item difficulty level
$Measure\ (logit) < -SD$	Easy
$-SD \leq Measure\ (logit) < 0{,}00$	Medium
$0{,}00 \leq Measure\ (logit) < SD$	Difficult
$Measure\ (logit) > SD$	Very Difficult

RESULTS AND DISCUSSION

Define Stage:

In the define stage, the first step taken is to involve and clearly define the learning objectives measured by the items developed, including identifying the mathematical CTS of students. Furthermore, the preparation of mathematical questions using arithmetic rows and series aims to measure the level of mathematical CTS in accordance with the indicators that have been formulated. The following are research indicators for the preparation of mathematical CTS test instruments, which refer to the CTS indicators Karim & Normaya (2015) interpretation, analysis, evaluation, and inference. The explanation of these indicators that researchers have adapted is as follows:

Table 8. Indicators of CTS

No	Indicator	Achievement indicator
1	Interpretation Provide opinions/ideas, according to the theory/concept, and clear answers to the problems given.	Learners are able to provide correct answers to the problems given based on views/opinions/ideas according to theory and concepts.
2	Analyze Identifying statements, the relationship questions, and the concepts contained in the problem can be done by making the right	Answers correctly based on the analysis of information and problems that have been presented in the

No	Indicator	Achievement indicator
	mathematical model, accompanied by a clear explanation. Furthermore, the results of the analysis of the information presented and the problem need to be conveyed or presented to provide a better understanding.	problem by connecting information and problems, so that they are able to write formulas and make mathematical models to solve problems in the problem.
3	Evaluation Using the right strategy in solving the problem, complete and correct in performing calculations.	Learners are able to provide correct answers through the strategies made to perform the right calculations to solve the problems presented.
4	Inference Making conclusions based on strong evidence involves the process of identifying various arguments or assumptions, as well as looking for alternative solutions. In addition, it is important to keep the situation and evidence in mind.	Learners are able to provide answers to the problems presented in the problem through conclusions that have been made based on appropriate evidence.

Adaptation of mathematical CTS indicators described by Karim & Normaya (2015)

Design Stage:

In the design stage, researchers continued the preparation of question items, scoring guidelines, and assessment rubrics for CTS based on the cognitive domain in the form of four essay questions on arithmetic rows and series material as a measuring tool to describe the mathematical CTS of students. 2 lecturers and 2 teachers validated the four questions as validators of question instruments. The following are the results of the validation of the four validators. All aspects declared feasible for use can be seen in the following table:

Table 9 Expert Validation Results

Aspect	Percentage %	Category	Feasibility
Material	84,37	Good	Feasible
Suitability of material with CTS indicators	79,68	Good	Feasible
construction	93,75	Very Good	Feasible
Language	85,93	Good	Feasible

The validation results show that the instrument is feasible to use, and there are no aspects that are categorised as inappropriate. The assessment rubric, after going through the revision stage, is as follows:

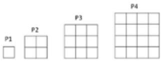
Table 10. CTS Assessment Rubric

Indicator	Score Description	Score
Interpretation	Did not write the answer	0
	Gave the wrong answer	1
	Gives an incorrect answer to the given problem based on almost correct views/opinions/ideas.	2
	It gives the correct answer to the given problem based on views/opinions/ideas according to the theory and concept, but it is not precise and unclear.	3
	Gives the correct answer to the given problem based on views/opinions/ideas according to the theory and concepts, but not clearly.	4
	Gives the correct answer to the given problem based on the views/opinions/ideas according to the theory and concept, precisely and clearly.	5
Analysis	Did not write the answer	0
	Gave the wrong answer	1
	Gives an incorrect answer based on the analysis of information and problems that have been presented, which are not precise and not clear.	2
	Gives a correct answer based on the analysis of information and problems that have been presented in the problem, but not yet precise and unclear.	3
	It gives a correct answer based on analysing the information and problems that have been presented, but are not yet clear.	4
	Provide answers correctly based on the analysis of information and problems that have been presented in the problem, precisely and clearly	5
Evaluation	Did not write the answer.	0
	Gave the wrong answer.	1
	Gave a correct answer through the strategy made and performed calculations that were not yet precise, and yet unclear.	2
	Gave a correct answer through the strategy made and performed calculations that were not yet precise, and yet unclear.	3
	Gave the correct answer through the strategy made, and did the right calculation, yet it was not clear.	4
	Gave the correct answer through the strategy created and performed calculations that were precise and clear.	5

Indicator	Score Description	Score
Inference	Did not write the answer	0
	Gave the wrong answer	1
	Hanya memberikan satu jawaban dari beberapa jawaban yang mungkin dan tidak memberikan kesimpulan	2
	Gives the correct answer to the problem presented in the problem through a conclusion that has been made based on evidence that is not yet precise and unclear.	3
	Gives the correct answer to the problem presented in the problem through the conclusion that has been made based on precise and unclear evidence.	4
	Gives the correct answer to the problem presented in the problem through the conclusion that has been made based on precise and clear evidence.	5

Furthermore, the test instrument was made with the selected CTS indicators. Researchers made one question from each indicator after going through the revision stage as follows:

Table 11: CTS Test Instrument

No	Indicator	Problem
1	Interpretation	Some squares are grouped and arranged in such a way that each group produces a new square, as shown below:  Can you determine the next arrangement of the squares?
2	Analysis	Mrs. Ariyanti received a poetry competition prize of Rp3,000,000. She wants to distribute some of the money to her seven nieces. The youngest niece gets the smallest share, and each niece gets more money from the younger niece according to the arithmetic sequence pattern. If the first niece receives Rp350,000 and the third niece receives Rp250,000, how much money does Mrs. Ariyanti have left after distributing the money to her seven nieces and nephews?
3	Evaluation	Sari added up all the pages of her book from page 1 to page 70 and got a total of 2445. However, it turns out that there is one missed page that is not counted. Is it true that there is one page that Sari has not counted? Prove it
4	Inference	Rosa is saving for the next 5 years. In the first month, Rosa saves Rp. 50,000, and the next month Rp. 75,000, - and so on. There are several methods/ways to find out the total savings Rosa has in year 5. What method do you think is the most appropriate, and explain your reasons?

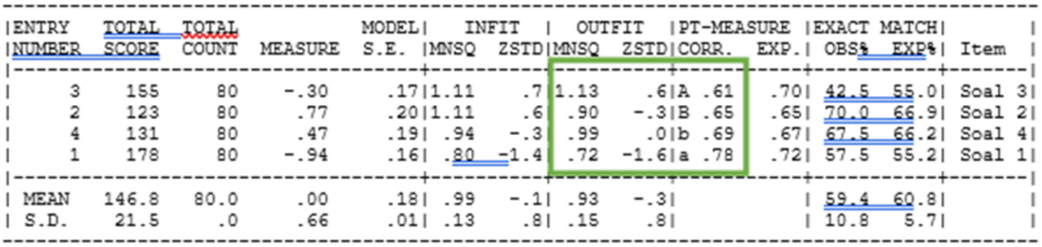
Development Stage:

At this stage, the researcher tested the mathematical CTS question instrument on a small scale, where the test question instrument was given to class X students of SMK YPE Nusantara Slawi and SMA Ihsaniyah Kota Tegal. The test results were then recorded using Microsoft Excel and analysed using the Winstep program. Information on the results of the analysis includes the following:

Analysis of the suitability of the test instrument items (content validity of the instrument)

The results of the analysis of item validity using the Rasch model show high effectiveness, because it can produce a reliable analysis (Sari et. al., 2016). The following are the results of the analysis of the validity of the question items carried out using the Winstep application. The results obtained in the form of Misfit order, which can be seen in Figure 2:

Figure 2. Misfit Order Analysis Results



```
---------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL                 MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT   MEASURE   S.E. |MNSQ   ZSTD|MNSQ   ZSTD|CORR.   EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|     3    155     80      -.30     .17|1.11     .7|1.13     .6|A .61    .70| 42.5  55.0| Soal 3|
|     2    123     80       .77     .20|1.11     .6| .90    -.3|B .65    .65| 70.0  66.9| Soal 2|
|     4    131     80       .47     .19| .94    -.3| .99     .0|b .69    .67| 67.5  66.2| Soal 4|
|     1    178     80      -.94     .16| .80   -1.4| .72   -1.6|a .78    .72| 57.5  55.2| Soal 1|
|------------------------------------+----------+----------+-----------+-----------+------|
| MEAN   146.8   80.0       .00     .18| .99    -.1| .93    -.3|           | 59.4  60.8|      |
| S.D.    21.5     .0       .66     .01| .13     .8| .15     .8|           | 10.8   5.7|      |
---------------------------------------------------------------------------------------------
```

The explanation of the validity test using the Rasch model will be explained in the table below:

Table 12. Results of Problem Item Suitability Analysis

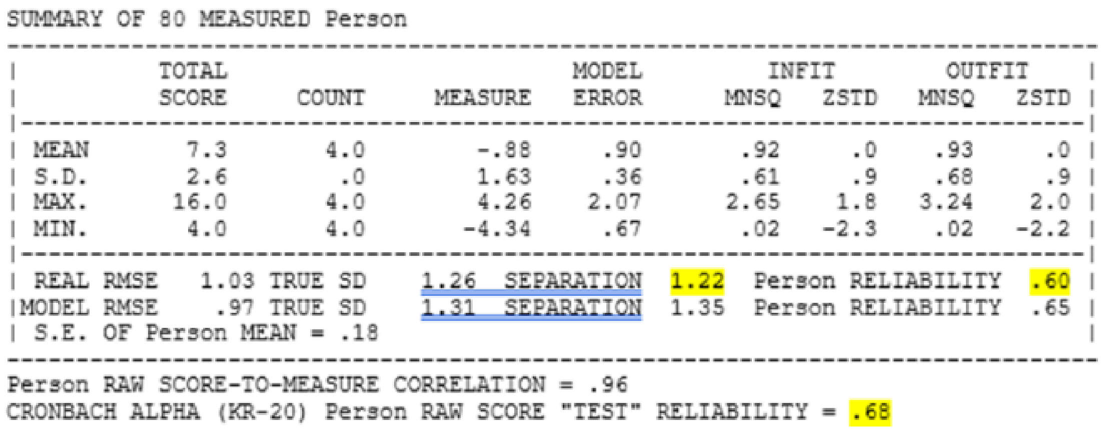
No	Item Number	Measurement Accuracy Criteria			Decision
		Outfit MNSQ	Outfit ZSTD	PT-Meancorr	
1	Soal 3	1,13	0,6	0,61	Valid
2	soal 2	0,90	-0,3	0,65	Valid
3	soal 4	0,99	0,0	0,69	Valid
4	soal 1	0,72	-1,6	0,78	Valid

Based on the results of the analysis of the table above, it can be seen that the items in Rasch modelling obtained information that 4 items are in the valid category because they meet all categories of Outfit MNSQ, Outfit ZSTD, and Pt Measure Colleration. The results of this study indicate that the existing questions have met the standards set and can ensure that the students' mathematical critical thinking skills have been tested through relevant and quality questions. This finding is reinforced by other research conducted by Parisu et al. (2024) and Wulandari et al. (2025) that the items are declared valid if they meet at least 2 item fit requirements. From the results of the item fit analysis, it can be seen that the Outfit MNSQ value of the four items lies between 0.5 and 1.5, which means that the four items meet this criterion. Furthermore, it can be seen that the Outfit ZSTD value of the four items lies between -2.0 to 2, which means that the four items meet this criterion. In the Point Measure Colleration value, the four items lie between 0.4 and 0.85. The conclusion that can be drawn is that the four questions can assess students' mathematical CTS. This is in accordance with the purpose of validity testing, which serves to assess how effective a measurement tool, such as a questionnaire or test, is in measuring what should be measured. In this study, what we want to measure is learners' mathematical CTS.

Analisis Reliabilitas Instrumen

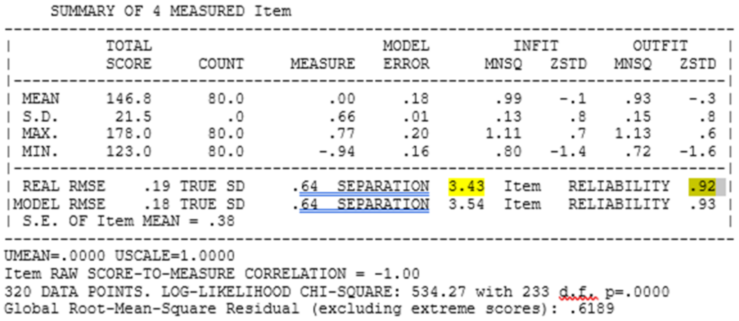
Reliability, which comes from the word "reliability", refers to the degree to which a measurement result can be trusted. Measurement results are considered reliable if, in several instances, carrying out measurements on the same group of subjects, consistent results are obtained, provided that the aspects measured in the subject remain unchanged (Ramadhan et. al.,2024). A high reliability value indicates that the items in the test provide consistent and reliable results when repeated under the same conditions. The results of the overall analysis of the instruments used in this study are detailed in the following figure:

Figure 3. Summary Statistic (Summary of 80 Measured Persons



```
SUMMARY OF 80 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD    MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       7.3       4.0        -.88     .90        .92     .0     .93     .0 |
| S.D.       2.6        .0        1.63     .36        .61     .9     .68     .9 |
| MAX.      16.0       4.0        4.26    2.07       2.65    1.8    3.24    2.0 |
| MIN.       4.0       4.0       -4.34     .67        .02   -2.3     .02   -2.2 |
|-----------------------------------------------------------------------------|
| REAL RMSE    1.03 TRUE SD     1.26  SEPARATION  1.22  Person RELIABILITY  .60 |
|MODEL RMSE     .97 TRUE SD     1.31  SEPARATION  1.35  Person RELIABILITY  .65 |
| S.E. OF Person MEAN = .18                                                    |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .68
```

Figure 4. Summary Statistics (Summary of 4 measured Items)



```
     SUMMARY OF 4 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     146.8      80.0         .00     .18       .99    -.1    .93    -.3 |
| S.D.      21.5        .0         .66     .01       .13     .8    .15     .8 |
| MAX.     178.0      80.0         .77     .20      1.11     .7   1.13     .6 |
| MIN.     123.0      80.0        -.94     .16       .80   -1.4    .72   -1.6 |
|-----------------------------------------------------------------------------|
| REAL RMSE     .19 TRUE SD      .64  SEPARATION  3.43  Item    RELIABILITY  .92 |
|MODEL RMSE     .18 TRUE SD      .64  SEPARATION  3.54  Item    RELIABILITY  .93 |
| S.E. OF Item MEAN = .38                                                     |
-------------------------------------------------------------------------------
UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
320 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 534.27 with 233 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): .6189
```

From the summary statistical output analysis conducted using Winstep, results were obtained in the form of tables, which were divided into two categories. The first table presents a summary of the measurements for individuals, i.e., learners. In contrast, the second table presents a summary of the items based on the suitability of the four items analysed. The reliability value for learners was recorded at 0.60, which indicates that the reliability is weak, as it falls below 0.67. In contrast, the reliability of the questions scored 0.92, indicating that the reliability was very high, falling within the range of 0.91 to 0.94. The results of this analysis are in line with Wulandari et al. (2025), which shows that the person reliability value reaches 0.86, which is classified as good, while item reliability is only 0.65, which is in the weak category. The results of this analysis are in line with research conducted by D. Pratama (2020), which found that the difference between the value of person reliability is good while item reliability is classified as weak. The significant difference between these two values shows that the consistency of students' answers in answering the mathematics test questions is still relatively weak, even though the quality of the items can be categorised as good. This can happen because the number of respondents is less than 100. This opinion is in line with the opinion of D. Pratama (2020) because the number of respondents is less than 100 people, which causes item reliability to be classified as weak, causing differences with person reliability, which is found to be very good. This opinion has been proven by several previous studies that have been conducted by Adi (2025) by analysing 25 items using 421 (more than 100) respondents to get a consistency of responses from positive respondents. The quality of questionnaire items in measuring instruments has a very good or extraordinary level of reliability, namely, obtained person reliability of 0.89 and item reliability of 0.99. Research conducted by Nuryanti et al., (2018) with a total of 20 items and 90 respondents, getting results in the form of item reliability of 0.95 classified as excellent while person reliability of 0.70 is classified as sufficient, there is a significant difference between item reliability and person reliability, this is because the number of respondents is only 90 (less than 100). Based on several previous studies, it can be concluded that to get consistent results between item reliability and person reliability, it is necessary to use as many respondents as possible in order to get maximum results.

Although there is a significant difference between item reliability and person reliability, the reliability of the interaction between the person or student and the overall question item from the Alpha Cronbach value of 0.68 shows that the reliability of the interaction of students with the question items is categorised as sufficient because the value is more than 0.67 and less than 0.80. This assessment is based on Cronbach's Alpha value, which can assess the relationship between individuals and all question items as a whole. Thus, it can be concluded that the level of consistency of students in taking the test is sufficient, or in other words, the overall reliability is in the sufficient category.

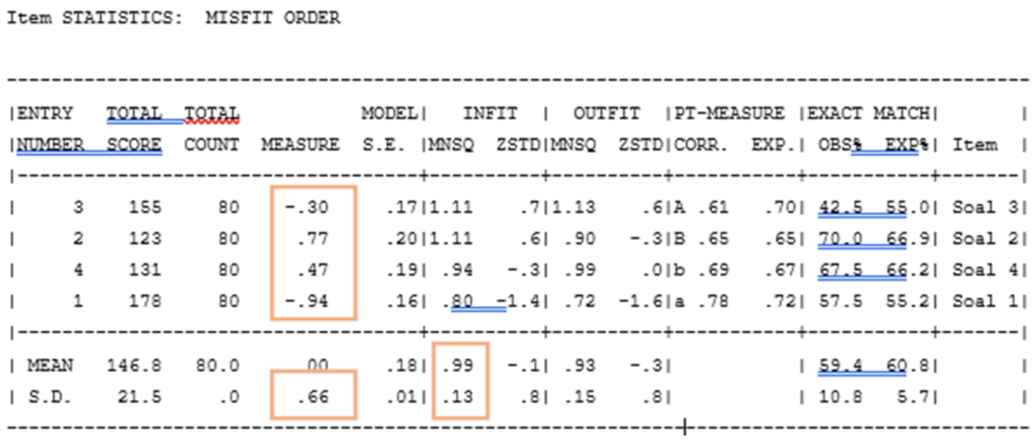
Item dan Item Separation Index

In grouping people and items, it can be seen through the separation value, where the greater the separation value, the better the overall quality of the instrument in identifying groups of respondents and items. The value of person separation can be seen through Figure 2. Summary Statistic (Summary of 80 Measured Persons) is 1.22. If the person's strata H formula is used, namely = $\frac{[(4 \times Separation)+1]}{3}$ (Ngadi, 2023). Then $H = \frac{[(4 \times 1{,}22)+1]}{3} = 1{,}96$ rounded to 2, this states that there are 3 2 groups of students. As for the value of item separation known through Figure 3. Sumary Statistiic (Sumary of 4 measured Items) is 3.43, then $H = \frac{[(4 \times 3{,}43)+1]}{3} = 4{,}906$ rounded to 5, this states that the question has good value because it divides the question in 5 groups. The results of this study indicate that person separation is acceptable because it is categorized as person separation > 5 and for item separation is categorized as excellent. This statement is in line with the view expressed by Sumintono & Widhiarso (2015), which states that the more the level of separation, the better the quality of the instrument developed. In other words, the higher the item separation value, the more accurate the resulting measurement. In addition, according to Ngadi (2023), the index value of the separation of respondents and items can be used to determine the consistency of the ability being measured and the consistency of the level of difficulty of the items; this means that the index value can indicate the level of reliability.

Analisis Tingkat Kesukaran

The difficulty level of an item affects the number of respondents who tend to give the correct answer. In Rasch modelling, to evaluate the difficulty level of an item, we can refer to Table 6, which shows the item measure values used in item categorisation. This criterion gives an idea of the difficulty level of the item. The logit value for each item is presented in the item measure, organised from highest to lowest. The logit value is an indicator that shows the difficulty level of an item. The more difficult an item is, the higher the logit value it has. An item is considered good if the difficulty level is balanced and proportional. The Misfit Order is shown in Figure 5. Item Statistic - Misfit order as follows:

Figure 5. Item Statistic: Misfit Order



```
Item STATISTICS:  MISFIT ORDER

------------------------------------------------------------------------------------------
|ENTRY    TOTAL  TOTAL                 MODEL|   INFIT    |  OUTFIT   |PT-MEASURE |EXACT MATCH|       |
|NUMBER   SCORE  COUNT  MEASURE   S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item  |
|------------------------------------+----------+----------+-----------+-----------+-------|
|     3     155     80    -.30     .17|1.11    .7|1.13    .6|A .61   .70| 42.5  55.0| Soal 3|
|     2     123     80     .77     .20|1.11    .6| .90   -.3|B .65   .65| 70.0  66.9| Soal 2|
|     4     131     80     .47     .19| .94   -.3| .99    .0|b .69   .67| 67.5  66.2| Soal 4|
|     1     178     80    -.94     .16| .80  -1.4| .72  -1.6|a .78   .72| 57.5  55.2| Soal 1|
|------------------------------------+----------+----------+-----------+-----------+-------|
| MEAN    146.8   80.0     .00     .18| .99   -.1| .93   -.3|           | 59.4  60.8|       |
| S.D.     21.5     .0     .66     .01| .13    .8| .15    .8|           | 10.8   5.7|       |
------------------------------------------------------------+-----------------------------
```

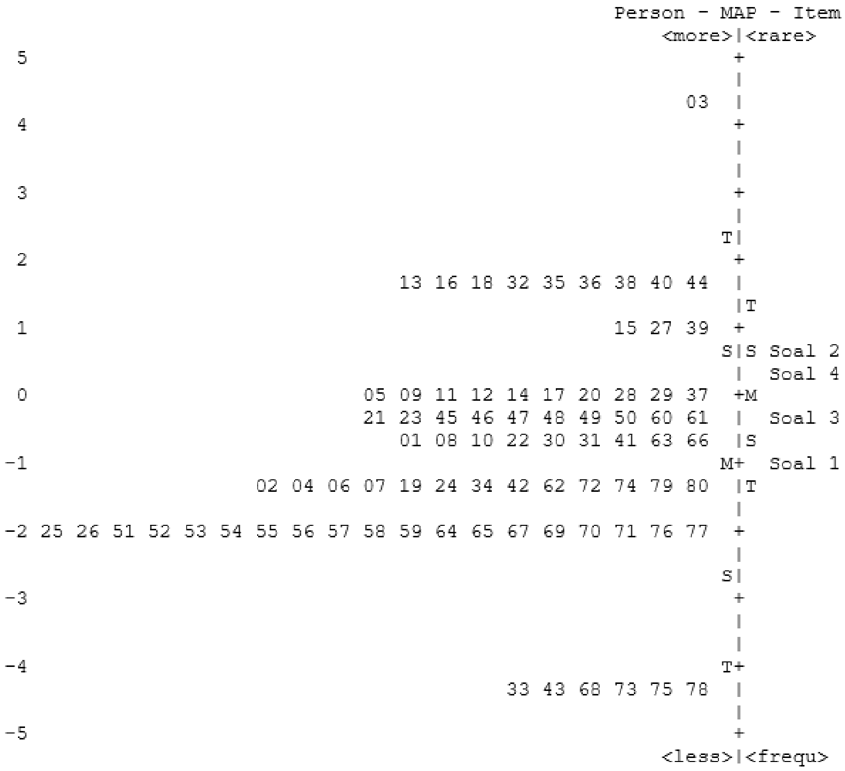
The results of the analysis of the difficulty level of the CTS test items can be interpreted as the difficulty level of each item based on the measure's logit value, which describes the relative difficulty level of each item in the test. From the figure above, the Standard Deviation (SD) value is 0.66. The results of the analysis for the CTS test instrument can be seen in the following table:

Table 12. Results of Logit (Measure) Value Analysis

Question No.	Logit (Measure) Value	Measure Value Analysis	Category
1	-0,94	*Measure* (Logit) < -0,66	Easy
2	0,77	*Measure* Logit > 0,66	Very Difficult
3	-0,30	-0,66 < *Measure* (Logit) < 0,00	Medium
4	0,47	0,00 ≤ *Measure* (Logit) < 0,66	Difficult

From the table above, it can be explained that Problem number 1 with the Interpretation indicator is categorised as easy, the analysis of the measure value shows -0.94 < -0.66. Furthermore, in question number 2, the analysis indicator is categorised as very difficult, because 0.77> 0.66. In question number 3, the evaluation indicator is categorised as moderate because the logit value is in the range -0.66 to 0.00. Furthermore, question number 4, which shows the inference indicator, is categorised as difficult because the logit value is in the range 0.00 to 0,66. From the analysis of the level of difficulty, it can be concluded that two CTS test questions are still classified as difficult. This analysis is in line with research conducted by Wulandari et al. (2025), which uses the Rasch analysis model. The results showed that the level of difficulty of the items consisted of two questions classified as very easy, seven questions classified as easy, and one question classified as difficult. In addition, research conducted by Fauziana & Wulansari (2021) shows the results of analysing the level of difficulty of items that are divided into several categories. From these results, there are four questions categorised as easy, four questions categorised as medium, and two questions categorised as difficult. Furthermore, the analysis of the level of difficulty of items can also be seen through variable maps, which are displayed in Figure 6:

Figure 6.Variable Map



```
                                                   Person - MAP - Item
                                                       <more>|<rare>
 5                                                           +
                                                             |
                                                         03  |
 4                                                           +
                                                             |
                                                             |
 3                                                           +
                                                             |
                                                            T|
 2                                                           +
                                 13 16 18 32 35 36 38 40 44  |
                                                             |T
 1                                                 15 27 39  +
                                                            S|S Soal 2
                                                             |  Soal 4
 0                            05 09 11 12 14 17 20 28 29 37  +M
                              21 23 45 46 47 48 49 50 60 61  |  Soal 3
                                 01 08 10 22 30 31 41 63 66  |S
-1                                                          M+  Soal 1
                     02 04 06 07 19 24 34 42 62 72 74 79 80  |T
                                                             |
-2 25 26 51 52 53 54 55 56 57 58 59 64 65 67 69 70 71 76 77  +
                                                             |
                                                            S|
-3                                                           +
                                                             |
                                                             |
-4                                                          T+
                                          33 43 68 73 75 78  |
                                                             |
-5                                                           +
                                                       <less>|<frequ>
```

From the results of testing the map variable, it is in accordance with the analysis of the level of difficulty using the logit value that the easiest question is question number 1 of the interpretation indicator, question number 2 of the analysis indicator is a question categorized as very difficult, question number 3 of the evaluation indicator is a question categorized as moderate and question number 4 of the inference indicator is a question categorized as difficult. Although question number 2 of the analysis indicator is classified as very difficult, it can be used in tests with the aim of measuring students' mathematical CTS because the results of validity testing support it and meet 3 testing criteria, namely the MNSQ Outfit value of 0.90, the ZSTD Outfit value of -0.3, and Pt. Measure Corr 0.65 so that this question is considered capable of measuring what is to be measured. If examined in terms of the ability of students to answer test instruments, it can be seen in the variable map image that students with codes 03, 13, 16, 18, 32, 35, 36, 38, 40, 44 can answer test instruments better than other students. However, 3 learners have a logit value of +1, namely learners with codes 15, 27, and 39. In comparison, the average ability of students in answering test instruments is at a value of $0 logit,$

which is a total of 29. Furthermore, students who have a low ability to answer the CTS test instruments are 32 students.

The results of this study analysing the level of difficulty lead to the preparation of mathematical CTS test instruments. In this situation, educators can understand the extent of students' mathematical CTS abilities and analyse what types of questions students categorise as very difficult questions. Based on the analysis that has been done, educators should be able to provide innovations in the learning process with the aim of increasing students' CTS.

CONCLUSION

Mathematical CTS test question instruments with indicators of Interpretation, analysis, evaluation, and inference, where each indicator is given one question, have good content validity test results, or in other words, all items are said to be valid. Furthermore, the Unidimensionality Test obtained raw variance explained by measures is 57.2% and overall, this construct validation is categorised as good. Based on the reliability test results, the person reliability value is 0.60 and the item reliability value is 0.92. In contrast, the Cronbach alpha value, which is a measure of reliability, reaches 0.68, which indicates that the level of reliability is categorised as sufficient. Based on the results of the analysis of the level of difficulty, it can be concluded that the questions for the interpretation indicator are categorised as easy, the analysis indicator is categorised as very difficult, the evaluation indicator is categorised as moderate, and the inference indicator is categorised as difficult. The results of this analysis show that the test instrument is feasible to use because it has met the validity and reliability tests. For the quality of the items categorised as excellent, however, there are still questions that are categorised as difficult by students. The results of the analysis conducted using the Rasch model can be a guide to evaluate the level of mathematical CTS of students. In addition, these results are also useful for assessing the learning process that has been implemented, so that educators can design more effective learning methods with the aim of increasing students' mathematical CTS.

REFERENCES

Adi, N. R. M. (2025). Analisis Kualitas Instrumen Literasi Media: Validitas dan Reliabilitas Menggunakan Model Rasch. *Briliant: Jurnal Riset Dan Konseptual*, *10*(2), 323–336. https://doi.org/10.28926/briliant.v10i2.2048

Ajizah, E., & Putu Artayasa, I. (2022). Validitas Bahan Ajar IPA Berbasis Problem Based Learning Untuk Meningkatkan Keterampilan Berpikir Kritis dan Sikap Ilmiah Peserta Didik. *Journal of Classroom Action Research*, *4*(2), 147–153. https://doi.org/10.29303/jcar.v4i1.1855

Anderha, R. R., & Maskar, S. (2021). Pengaruh Kemampuan Numerasi Dalam Menyelesaikan Masalah Matematika Terhadap Prestasi Belajar Mahasiswa Pendidikan Matematika. *Jurnal Ilmiah Matematika Realistik*, *2*(1), 1–10. https://doi.org/10.33365/ji-mr.v2i1.774

Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioural Sciences*, *129*, 133–139. https://doi.org/10.1016/j.sbspro.2014.03.658

Dini Riyantini Sari, Nanan Sekarwana, Zahrotur Rusyda Hinduan, & Bambang Sumintono. (2016). Analisis Tingkat Kepuasan Masyarakat terhadap Dimensi Kualitas Pelayanan Tenaga Pelaksana Eliminasi Menggunakan Pemodelan Rasch. *JSK: Jurnal Sistem Kesehatan*, *2*(1), 47–55. https://jurnal.unpad.ac.id/jsk_ikm

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (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*(1), 11–19. https://doi.org/https://doi.org/10.23887/ijerr.v3i1.24080

Fauziana, A., & Wulansari, A. D. (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

Hadiyanti, H., Susongko, P., & Munadi. (2024). Pengembangan Instrumen Higher Order Thinking Skill Mata Pelajaran Matematika Dengan Rasch Model. *Journal of Education Research*, *5*, 399–407. https://doi.org/https://doi.org/10.37985/jer.v5i1.765

Irfiani, V., Junaedi, I., & Waluya, S. B. (2023). Systematic Literature Review: Kemampuan Berpikir Kritis Matematis Siswa Ditinjau dari Adversity Quotient. *Jurnal Pendidikan Matematika*, *1*(2), 11. https://doi.org/10.47134/ppm.v1i2.157

JANNAH, U., Rosi, M., & Hafsi, A. R. (2023). Profil Kemampuan Numerasi Siswa Terhadap Kecerdasan Emosional. *As-Salam: Jurnal Studi Hukum Islam & Pendidikan*, *12*(1), 98–112. https://doi.org/10.51226/assalam.v12i1.513

Jayadi, A., Putri, D. H., & Johan, H. (2020). Identifikasi Pembekalan Ketrampilan

Abad 21 Pada Aspek Keterampilan Pemecahan Masalah Siswa SMA Kota Bengkulu dalam Mata Pelajaran Fisika. *Jurnal Kumpara Fisika*, *3*(1), 25–32. https://doi.org/https://doi.org/10.33369/jkf.3.1.25-32

Karim, K., & Normaya, N. (2015). Kemampuan Berpikir Kritis Siswa dalam Pembelajaran dalam Pembelajaran Matematika dengan Menggunakan Model Jucama di Sekolah Menengah Pertama. *EDU-MAT: Jurnal Pendidikan Matematika*, *3*(1). https://doi.org/10.20527/edumat.v3i1.634

Lutfiah, F. C., Juniati, D., Khabibah, S., Surabaya, U. N., & Wetan, J. L. (2023). Pengaruh Mathematical Literacy Terhadap Peningkatan Critical Thinking: Literature Review. *SUPERMAT Journal Pendidikan Matematika*, *7*(2), 208–218. https://doi.org/https://doi.org/10.33627/sm.v7i2.1588

Muliana, G. (2021). Analisis Kemampuan Berpikir Kritis Matematis Siswa Kelas X pada Materi Persamaan Logaritma Ditinjau dari Kemandirian Belajar. *MATH LOCUS: Jurnal Riset Dan Inovasi Pendidikan Matematika*, *2*(1), 15–22. https://doi.org/10.31002/mathlocus.v2i1.1475

Muntazhimah, Syifani Putri, & Hikmatul Khusna. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 65. https://doi.org/10.30998/jkpm.v6i1.8144

Mustika Wati, & Saiyidah Mahtari. (2017). *Artikel Penelitian / Article Reviu Pengembangan Instrumen Kognitif Fisika Siswa SMP*. *1*(1), 45–56.

Ngadi, N. (2023). Analisis Model Rasch Untuk Mengukur Kompetensi Pengetahuan Siswa Smkn 1 Kalianget Pada Mata Pelajaran Perawatan Sistem Kelistrikan Sepeda Motor. *Jurnal Pendidikan Vokasi Otomotif*, *6*(1), 1–20. https://doi.org/10.21831/jpvo.v6i1.63479

Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, *4*(2), 224–233. https://doi.org/10.21831/jipi.v4i2.21442

Parisu, C. Z. L., Ekadayanti, W., Sisi, L., Juwairiyah, A., & Kasmawati. (2024). Analisis Butir Soal Pengetahuan Dasar Matematika Menggunakan Pendekatan Rasch. *SCIENCE TECH : Jurnal Ilmu Pengetahuan Dan Teknologi*, *10*(1), 36–45.

Pratama, B. A., & Mardiani, D. (2022). Kemampuan berpikir kritis matematis antara siswa yang mendapat model problem-based learning dan discovery learning. *Jurnal Inovasi Pembelajaran Matematika: PowerMathEdu*, *1*(1), 83–92. https://doi.org/10.31980/pme.v1i1.1368

Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy : Jurnal Pendidikan Islam*, *7*(1), 61–70. https://doi.org/10.32923/tarbawy.v7i1.1187

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education*, *6*(2), 10967–10975.

https://doi.org/10.31004/joe.v6i2.4885

Ridho, A. (2011). *MULTIDIMENSIONALITAS PADA TES POTENSI AKADEMIK*. 1–16. https://repository.uin-malang.ac.id/1855/

Rizti, T. M., & Prihatnani, E. (2021). Efektivitas Model Pembelajaran 3CM (Cool-Critical-Creative-Meaningfull) terhadap Kemampuan Berpikir Kritis Siswa SMP. *Mosharafa: Jurnal Pendidikan Matematika*, *10*(2), 213–224. https://doi.org/10.31980/mosharafa.v10i2.945

Rosita Dewi Nur, I., Herman, T., & Mariyana, R. (2019). Logical-Mathematics Intelligence in Early Childhood. *International Journal of Social Science and Humanity*, *May*, 105–109. https://doi.org/10.18178/ijssh.2018.v8.944

Sahira, P. E., & penggabean, E. M. (2022). Development of Test Instruments to Measure Middle School Students' Creative Thinking Ability. *EDUMATIKA JURNAL MIPA*, *2*(1), 167–171. https://doi.org/https://doi.org/10.56495/emju.v2i4.300

Salsabila, U. H., Ilmi, M. U., Aisyah, S., Nurfadila, N., & Saputra, R. (2021). Peran Teknologi Pendidikan dalam Meningkatkan Kualitas Pendidikan di Era Disrupsi. *Journal on Education*, *3*(01), 104–112. https://doi.org/10.31004/joe.v3i01.348

Sebariani, N., Ningsih, I., & Khoiruddin, A. Y. (2023). Faktor-Faktor Yang Mempengaruhi Loyalitas Nasabah Bank Syariah Indonesia. *Jurnal Syarikah*, *9*(2), 197–208. https://doi.org/https://doi.org/10.30997/jsei.v9i2.9884

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

Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. *173*(Icei 2017), 38–42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., Islam, U., Indonesia, I., Widhiarso, W., & Mada, U. G. (2014). *untuk Penelitian Ilmu-Ilmu Sosial*. *November*.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch Pada Assessment Pendidikan. *AplikAsi RascH PemodelAn Pada Assessment Pendidikan*, *September*, 1–24.

Wulandari, O., Muhtarom, M., & Sumarno, S. (2025). *Analisis Butir Soal Pengetahuan Matematika Kelas V Sekolah Dasar Menggunakan Model Rasch*. *07*(1), 293–302.

Zain, N. K., & Marhayati. (2024). Level of Critical Thinking Abilities of Madrasah Aliyah Negeri 1 Jombang Students in Solving Geometry Problems: A Review Based on Cognitive Style. *Kontinu: Jurnal Penelitian Dikdaktik Matematika*, *08*(2), 120–134. https://doi.org/http://dx.doi.org/10.30659/kontinu.8.2