Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

October 2022. Vol. 10, No. 4
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 842-855

Development of Critical Thinking Ability Test using the Rasch Model on Substance Pressure Materials

***Eliza, Tomo Djudin, Erwina Oktavianty**
Department of Physics Education, FKIP, Tanjungpura University.
Jl Prof. Dr. H. Hadari Nawawi, Bansir Laut, Southeast Pontianak, Pontianak City,
West Kalimantan, Indonesia. Postal Code: 78115
* Correspondence Author e-mail: eliza@student.untan.ac.id

Received: August 2022; Revised: September 2022; Published: October 2022

Abstract

This study aims to produce an instrument to measure students' critical thinking skills on pressure equipment in class VIII SMP. The type of research used is Research and Development (R&D). The type of research used is research and development (R&D). The test instrument developed was in the form of multiple choice based on eight development steps, namely potential and problems, information gathering, product design, product validation, product revision, product testing, product revision, and final product. Validation of the instrument was developed by 3 validators, namely 2 lecturers of Physics Education at FKIP Tanjungpura University and 1 science teacher at the school. The validation given to the validator is in the form of a material validation questionnaire and question validation. The results of material expert validation data obtained 84.72%, validation obtained from content validation with an Aiken index of 0.78, included in the high category and suitable for use. The results of the data analysis of teacher and student response questionnaires obtained 92.60% and 86.26% categorized as very feasible. The results of construct validity using the Rasch model with the application of Winstep 3.73 obtained result 3, which was declared inappropriate because it did not meet the MNSQ, ZSTD, and Pt Measure Corr criteria. The reliability of the Cronbach alpha test value obtained a value of 0.73 with a fairly high category. Regarding the level of difficulty of the item, 4 question items must be discarded because they are included in the very easy and very easy categories. The overall discriminatory power is in the good category with a discriminatory index of 0.40 DP < 70. The results of the study show that out of 30 things, 23 things are appropriate to be used to measure students' critical thinking skills with good-quality items.

Keywords: Test development, Critical Thinking, Rasch Model

How to Cite: Eliza, E., Djudin, T., & Oktavianty, E. (2022). Development of Critical Thinking Ability Test using the Rasch Model on Substance Pressure Materials. *Prisma Sains : Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 10*(4), 842-855. doi:https://doi.org/10.33394/j-ps.v10i4.5987

https://doi.org/10.33394/j-ps.v10i4.59873

Copyright© 2022, Eliza et al.
This is an open-access article under the CC-BY License.

INTRODUCTION

The industrial revolution 4.0 requires educators to explore ways to utilize human potential to stay ahead so that they can be competent in the 21st century. One of the challenges for educators is how to measure the skills and characteristics of society in the industrial revolution era, namely a knowledge-based society (Ramadhan & Iriani, 2013). According to Widana et al. (2019), 21st-century skills in education are a set of 4C skills consisting of critical thinking, creativity, collaboration, and communication.

Critical thinking is a decision-making process based on the existing evidence, the context of the situation, and the ideas being considered (Facione, 2011). When studying at school (Daniati et al., 2018), critical thinking is a skill that needs to be developed, taught, trained, and accustomed to. There is no way to estimate students' critical thinking skills without using test instruments because critical thinking can be developed by teaching and can

also be improved through appropriate evaluation steps. Therefore, knowledge of critical thinking aspects and indicators is needed in making critical thinking ability test instruments. According to Facione (2015), an indicator of a person's critical thinking ability is whether they can interpretation, analysis, inference, evaluation, explanation, and self-regulate.

Indonesia is ranked 69th out of 76 countries studied in PISA 2015, according to the results of the PISA and TIMSS surveys. In line with the results of the TIMSS survey of junior high school (SMP) students with the characteristics of cognitive level questions, it shows that Indonesian students consistently scored lower than students in other countries on questions measuring critical thinking skills. Indonesia was ranked 50th out of 53 countries that contributed to the event (Karim & Normaya, 2015). Based on these results, shows that the ability of Indonesian students when dealing with high-level skills is still low, including critical thinking skills. According to Yulian (2016), the scores obtained by Indonesian students on PISA and TIMSS are still low due to the nature of the test questions that tend to focus on problem-solving, while students in Indonesia are not accustomed to working on arguing, problem-solving, and communicating questions.

According to the results of research conducted by (Jamaluddin et al., 2020), the development of higher order thinking skills, such as critical thinking skills, has not been planned intentionally or systematically by teachers in most of the lesson plans. Many find the habits of teachers who design and implement learning oriented to mastery of concepts with low cognitive levels. Though the ability to think critically belongs to a high level of cognitive level. Students are not accustomed to or trained in developing their critical thinking potential in dealing with problems, thereby reducing their competence and independence and reducing their critical thinking skills (Arifin, 2017). In line with the results of the pre-research conducted in several junior high schools in the Tebas sub-district (Junior High School 1 Tebas, Junior High School 3 Tebas, Junior High School 6 Tebas, and Junior High School 7 Tebas), one of the science teachers said that the material on substance pressure for class VIII was interesting to study because of the following: this is common in everyday life. However, when questions are asked about problems that occur in the environment, students cannot describe problem solving and critical thinking skills and have not been able to acquire new knowledge for students.

Based on these problems, it shows that in the matter of substance abuse, students' critical abilities are still lacking. Many students do not understand the concept of substance pressure, so the average score obtained is only 65 for those who get low scores on daily tests. The results of interviews with science teachers show that when applying assessments, teachers do not have knowledge of critical thinking skills, nor do they have special instruments to assess these skills. This is because the teacher is not experienced in compiling questions that assess students' critical thinking skills. In addition, students have not been trained in taking tests oriented to critical thinking skills, so teachers plan and carry out learning activities oriented to mastery of concepts. This makes students unable to answer and examine questions regarding problem solving well. Research conducted by (Rahaju et al., 2020), found that the problems found in school learning are questions that tend to test many aspects of memory that do not train students' higher-order thinking skills. One of the reasons is the lack of development of thinking skills. Students are critical because they are not accustomed to critical thinking when learning takes place by the teacher, especially when assessment is used. Based on the results of pre-research that has been carried out, this research focuses on developing a measurement instrument in the form of a critical thinking ability test on substance abuse material in junior high school to train students to use their thinking methods more effectively, especially in critical thinking.

In general, the implementation of education in applying critical thinking skills assessment is still low, at around 45% (Satria Mukti & Istiyono, 2018). Another finding is that critical thinking skills focus on measuring how well students perform against competency standards and take tests. In addition, not many researchers have looked at the

items or tools used in depth, and researchers have not explored the abilities of students based on responses. Based on the results of observations made by (Febriano et al., 2021)stated that the facts in the field show that the teachers have not paid attention to indicators of critical thinking skills, especially interpreting, analyzing, and evaluating. From the study of questions contained in student learning textbooks, daily assessments, and interviews, it has been found that the questions given are not following the typical problems and do not help students develop critical thinking skills. To overcome these kinds of problems, it is necessary to have a test instrument that is used to measure students' critical thinking skills with appropriate and quality data analysis tools. From the results of research conducted by (Sabekti & Khoirunnisa, 2018) regarding the development of critical thinking skills tests using the Rasch model, the data analysis obtained shows that 12 of the 13 question items meet the criteria for good questions according to the Rasch model, with instrument reliability of 0.72 (enough). This study provides evidence that the Rasch model can be used to develop instruments to measure students' critical thinking skills. In contrast to previous studies to measure critical thinking skills, many use this type of essay test. In this study, we used multiple-choice tests because the multiple-choice format is still little applied in assessing critical thinking skills. After all, students easily guess the answer from several available answer choices. However, not all indicators of critical thinking ability can be measured by multiple-choice tests. According to (Facione, 2013), a person's critical thinking does not have to fulfill all aspects of critical thinking, but one or several aspects can be chosen depending on the focus of the discipline being studied. Therefore, the researcher uses four critical thinking indicators, which refer to the indicators used by (Karim & Normaya, 2015), namely, interpreting, analyzing, evaluating, and inference.

In determining the results of the test assessment, the teacher does not see the analysis process, even though the test is part of the assessment that must be valid and reliable (Solichin, 2017). The analytical tool that is often used by teachers and used as a reference in analyzing the instrument is the classical test theory (Classical Test Theory). However, classical measurement theory has several limitations, where the level of difficulty of the items depends on the ability of the respondents, so it is not appropriate to use classical measurement theory to assess students' abilities and analyze items (Novinda et al., 2019). Therefore, to overcome the shortcomings of classical measurement theory, modern measurement theories are needed, such as Item Response Theory (IRT), one of which is the most famous Rasch model. The Rasch model is advantageous in terms of predicting missing data due to its ability to analyze individual response patterns. According to Sumintono & Widhiarso (2013), the Rasch model is used to analyze test instruments that can describe the relationship between items and respondents, which brings measurements to more objective and precise results. The Rasch model is more accurate than classical theory analysis in terms of determining the accuracy of an item. From the explanation that has been explained, this research examines the development of a critical thinking ability test using Rasch model analysis on substance pressure material in Class VIII Junior High School.

Apart from the description that has been explained, the selection of the Rasch Model to analyze data is still rarely used because there are still many studies that use classical test theory, which still has many weaknesses and shortcomings. The data analysis technique and development model used is the differences between this research and previous research. In previous studies, the analysis of items was still classically used with a limited-scale trial in only one school, while in this study, the item analysis tool used the Rasch model with the Winsteps program with large-scale trials in several schools in the Tebas District. In general, the purpose of this study is to produce a test that can measure students' critical thinking skills with good quality items with four critical thinking indicators, which refer to the indicators used by Karim & Normaya (2015) as follows (Table 1).

Table 1. Aspects and Indicators of critical thinking in research

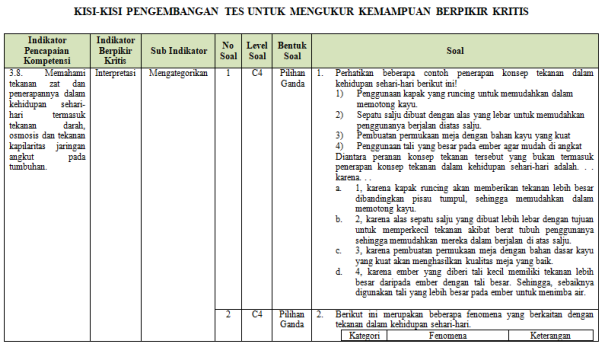
No	Critical thinking aspect	Indicator
1	Interpretation	a) Categorize b) Coding c) Clear meaning
2	Analysis	a) Checking ideas b) Detect arguments c) Analyze arguments
3	Inference	a) Questioning the evidence b) Estimating alternatives c) Describe conditions using deductive and inductive reasoning
4	Evaluation	a) Assess the credibility of the claim b) Assessing the quality of the arguments made using inductive and deductive reasoning

METHOD

The research and development method (research and development) is the method used in this study, with reference to the development model from Borg & Gall adapted by (Sugiyono, 2016) which consists of 8 development steps, namely: 1) potential and problems; 2) information gathering; 3) initial product design; 4) design validation; 5) design revision; 6) product trial; 7) product revision; and 8) final product result. The development stage of the critical thinking ability test is based on the steps by Mardapi (2012).

The population in this study were students in class VIII in the District of Tebas. In the sample based on Rasch modeling with a value of ±0.5 logit 99% confidence level, which means the sample size range is 108–243 samples (Sumintono, B, & Widhiarso, 2015)

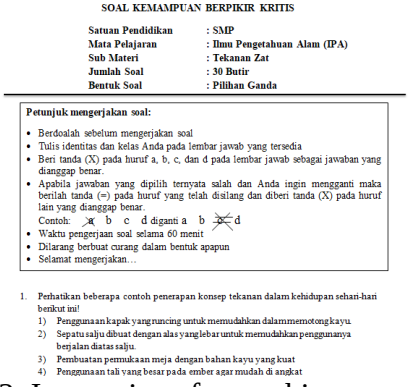
The population in this study were students in class VIII in the District of Tebas. In the sample based on Rasch modeling with a value of ±0.5 logit 99% confidence level, which means the sample size range is 108–243 samples (Sumintono, B, & Widhiarso, 2015). The research sample was taken from the average low, medium, and high UN IPA scores in 2019 in Tebas District, which amounted to 120 students. Technical data analysis in the form of practicality questionnaires from teachers and students, as well as a test totaling 30 items in the form of multiple choice. The instrument's validity is in the form of content validation through the assessment of 3 material and question expert validators, namely 2 Physics Education lecturers, FKIP Tanjungpura University, and 1 science teacher, while construct validation is analyzed using the Rasch model on Winstep software version 3.73. The results of the content validation are processed through the V Aiken formula. While the construct validation was processed and analyzed using the Rasch model through Winstep software version 3.73. The following are the instruments developed at the product design stage, namely as follows (Figure 1-Figure 4).



KISI-KISI PENGEMBANGAN TES UNTUK MENGUKUR KEMAMPUAN BERPIKIR KRITIS

Indikator Pencapaian Kompetensi	Indikator Berpikir Kritis	Sub Indikator	No Soal	Level Soal	Bentuk Soal	Soal
3.8. Memahami tekanan zat dan penerapannya dalam kehidupan sehari-hari termasuk tekanan darah, osmosis dan tekanan kapilaritas jaringan angkut pada tumbuhan.	Interpretasi	Mengategorikan	1	C4	Pilihan Ganda	1. Perhatikan beberapa contoh penerapan konsep tekanan dalam kehidupan sehari-hari berikut ini! 1) Penggunaan kapak yang runcing untuk memudahkan dalam memotong kayu. 2) Sepatu salju dibuat dengan alas yang lebar untuk memudahkan penggunanya berjalan diatas salju. 3) Pembuatan permukaan meja dengan bahan kayu yang kuat 4) Penggunaan tali yang besar pada ember agar mudah di angkat Diantara peranan konsep tekanan tersebut yang bukan termasuk penerapan konsep tekanan dalam kehidupan sehari-hari adalah. . . karena. . a. 1, karena kapak runcing akan memberikan tekanan lebih besar dibandingkan pisau tumpul, sehingga memudahkan dalam memotong kayu. b. 2, karena alas sepatu salju yang dibuat lebih lebar dengan tujuan untuk memperkecil tekanan akibat berat tubuh penggunanya sehingga memudahkan mereka dalam berjalan di atas salju. c. 3, karena pembuatan permukaan meja dengan bahan dasar kayu yang kuat akan menghasilkan kualitas meja yang baik. d. 4, karena ember yang diberi tali kecil memiliki tekanan lebih besar daripada ember dengan tali besar. Sehingga, sebaiknya digunakan tali yang lebih besar pada ember untuk menimba air.
			2	C4	Pilihan Ganda	2. Berikut ini merupakan beberapa fenomena yang berkaitan dengan tekanan dalam kehidupan sehari-hari. Kategori | Fenomena | Keterangan

Figure 1. The Grid of Critical Thinking



SOAL KEMAMPUAN BERPIKIR KRITIS

Satuan Pendidikan : SMP
Mata Pelajaran : Ilmu Pengetahuan Alam (IPA)
Sub Materi : Tekanan Zat
Jumlah Soal : 30 Butir
Bentuk Soal : Pilihan Ganda

Petunjuk mengerjakan soal:

- Berdoalah sebelum mengerjakan soal
- Tulis identitas dan kelas Anda pada lembar jawab yang tersedia
- Beri tanda (X) pada huruf a, b, c, dan d pada lembar jawab sebagai jawaban yang dianggap benar.
- Apabila jawaban yang dipilih ternyata salah dan Anda ingin mengganti maka berilah tanda (=) pada huruf yang telah disilang dan diberi tanda (X) pada huruf lain yang dianggap benar.
  Contoh: a b c d diganti a b ~~c~~ d
- Waktu pengerjaan soal selama 60 menit
- Dilarang berbuat curang dalam bentuk apapun
- Selamat mengerjakan…

1. Perhatikan beberapa contoh penerapan konsep tekanan dalam kehidupan sehari-hari berikut ini!
   1) Penggunaan kapak yang runcing untuk memudahkan dalam memotong kayu.
   2) Sepatu salju dibuat dengan alas yang lebar untuk memudahkan penggunanya berjalan diatas salju.
   3) Pembuatan permukaan meja dengan bahan kayu yang kuat
   4) Penggunaan tali yang besar pada ember agar mudah di angkat

Figure 2. Instructions for working on questions

1. Perhatikan beberapa contoh penerapan konsep tekanan dalam kehidupan sehari-hari berikut ini!
   1) Penggunaan kapak yang runcing untuk memudahkan dalam memotong kayu.
   2) Sepatu salju dibuat dengan alas yang lebar untuk memudahkan penggunanya berjalan diatas salju.
   3) Pembuatan permukaan meja dengan bahan kayu yang kuat
   4) Penggunaan tali yang besar pada ember agar mudah di angkat

   Diantara peranan konsep tekanan tersebut yang bukan termasuk penerapan konsep tekanan dalam kehidupan sehari-hari adalah. . . karena. . .
   a. 1, karena kapak runcing akan memberikan tekanan lebih besar dibandingkan pisau tumpul, sehingga memudahkan dalam memotong kayu.
   b. 2, karena alas sepatu salju yang dibuat lebih lebar dengan tujuan untuk memperkecil tekanan akibat berat tubuh penggunanya sehingga memudahkan mereka dalam berjalan di atas salju.
   c. 3, karena pembuatan permukaan meja dengan bahan dasar kayu yang kuat akan menghasilkan kualitas meja yang baik.
   d. 4, karena ember yang diberi tali kecil memiliki tekanan lebih besar daripada ember dengan tali besar. Sehingga, sebaiknya digunakan tali yang lebih besar pada ember untuk menimba air.

2. Berikut ini merupakan beberapa fenomena yang berkaitan dengan tekanan dalam kehidupan sehari-hari.

Kategori	Fenomena	Keterangan
I	 Dongkrak Hidrolik	Tukang cuci mobil sedang menaikkan mobil agar memudahkan proses pencucian mobil
II		Sepatu salju dipakan agar

Figure 3. Display item questions

KUNCI JAWABAN DAN RUBRIK SKORING INSTRUMEN PENILAIAN KEMAMPUAN BERPIKIR KRITIS MATERI TEKANAN ZAT

1. Interpretation				
Indikator	**No Item**	**Pertanyaan**	**Kunci Jawaban**	**Rubrik Penskoran**
1.1 Mengkategorikan	1	Perhatikan beberapa contoh penerapan konsep tekanan dalam kehidupan sehari-hari berikut ini ! 1) Penggunaan kapak yang runcing untuk memudahkan dalam memotong kayu. 2) Sepatu salju dibuat dengan alas yang lebar untuk memudahkan penggunanya berjalan diatas salju. 3) Pembuatan permukaan meja dengan bahan kayu yang kuat. 4) Penggunaan tali yang besar pada ember agar mudah di angkat. Diantara peranan konsep tekanan tersebut yang bukan termasuk penerapan konsep tekanan dalam kehidupan sehari-hari adalah…. karena . . . a. 1 karena kapak runcing akan memberikan tekanan lebih besar dibandingkan pisau tumpul, sehingga memudahkan dalam memotong kayu. b. 2 karena alas sepatu salju yang dibuat lebih lebar dengan tujuan untuk memperkecil tekanan akibat berat tubuh penggunanya sehingga memudahkan mereka dalam berjalan di atas salju. c. 3 karena pembuatan permukaan meja dengan bahan dasar kayu yang kuat akan menghasilkan kualitas meja yang baik. d. 4 karena ember yang diberi tali kecil memiliki tekanan lebih besar daripada ember dengan tali besar. Sehingga, sebaiknya digunakan tali yang lebih besar pada ember untuk menimba air.	C	• Skor 1, Jika jawaban benar • Skor 0, jika jawaban salah

Figure 4. Answer key

RESULTS AND DISCUSSION

The product developed is a test of critical thinking skills on substance abuse material in class VIII SMP in the form of 30 multiple-choice questions with a total of 120 students in SMP in Tebas District. The purpose of this research and development is to develop a test that is appropriate to be used in measuring or assessing students' critical thinking skills related to the concept of substance pressure in junior high school. At the potential and problem stages, it begins with preliminary research by analyzing needs through interviews and observations in each school. Based on the results of interviews and observations that have been carried out with each class VIII teacher at Junior High School 1 Tebas, Junior High School 3 Tebas, Junior High School 6 Tebas, and Junior High School 7 Tebas, it was found that all schools used the 2013 curriculum with methods and learning models in schools. diverse. Each of these schools has the potential to develop test tools that can assess cognitive domains to support the various methods and models taught by teachers.

There are several problems with the school assessment process. For example, teachers do not have special instruments to assess students' critical thinking skills, which leads to the teacher's lack of attention to students' critical thinking skills. In addition, students critical thinking skills cannot be measured because teachers only design and carry out learning activities that focus on mastery of concepts, meaning that many students are unable to answer and analyze problems related to problem-solving properly. And the last problem is in determining the results of the teacher's assessment by not paying attention to the analysis process so that an inaccurate test analysis is obtained in the assessment process. At the information collection stage, it is done by collecting interview data or collecting information through relevant reference sources such as books, the internet, or journals critical thinking skills using the Rasch model analysis or other supporting journals.

At the product design stage, there are steps to compile test specifications and write test questions. At the test specification stage, to assess students' critical thinking skills on substance pressure in class VIII Junior High School, the first step is to determine the purpose of the test. The second step is to compile a test grid by identifying KI and KD as well as the material to be developed first. Then, describe KI into learning indicators and make indicators according to critical thinking indicators modified from the development carried out by (Facione, 2015) namely interpretation, analysis, evaluation, and inference. Finally, make a grid of questions based on critical thinking indicators.

In the process of working on critical thinking skills, the specified duration is 60 minutes with a total of 30 multiple-choice questions. At the product validation stage, there is a test review stage where the product validation results are obtained from expert validator data, namely from 2 physics education lecturers at FKIP Tanjungpura University and 1 science teacher from the designated school. Material validation resulted in a feasibility percentage of 86.67% based on the presentation aspect, 90% based on the content quality

aspect, 77.78% based on the construction aspect, and 84.44% based on the language aspect. The average of all aspects was obtained by 84.72% in the very feasible category. For the validation of the 30 questions that have been made, it is stated that 30 items are valid with V > 0.70, with 17 items having an Aiken index of > 0.70 and 13 items having an Aiken index of 0.80, which means they have a validity coefficient. high content that supports the overall test item content (Azwar, 2015).

Based on the results of the analysis, it was concluded that all items were declared valid. This was supported by the average Aiken index obtained, which was 0.78 in the high category. Furthermore, the assessment was also carried out by subject teachers in four schools using a response questionnaire. This response questionnaire was created to correct the shortcomings of the developed test and to evaluate the test as learning material. The diagram below shows the results of the teacher's response as follows.

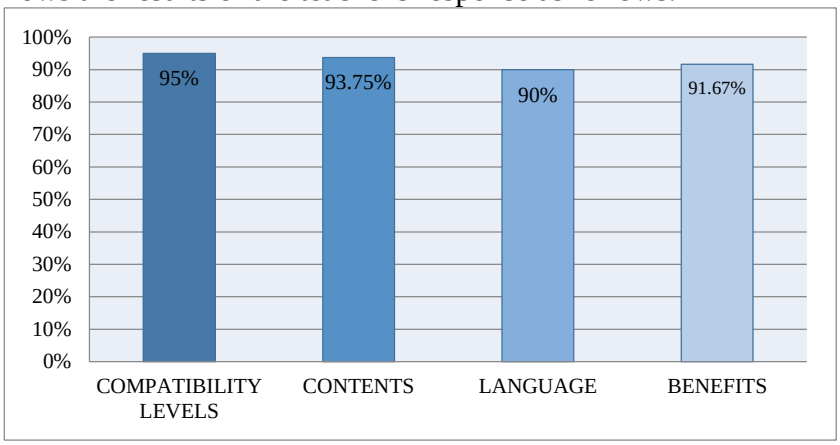


Figure 5. The Teacher Response Analysis Results Diagram

From the diagram of the results of the teacher's response analysis, it can be seen that the percentage of conformity obtained is 95%, the content aspect is 93.75% in the very appropriate category, the language aspect is 90% in the very appropriate category, and the usefulness aspect is 91.76% in the very feasible category.

Based on the percentage score of each aspect, an average of 92.60% is obtained in the very feasible category. The validation results show that there are several suggestions and improvements to be made to the tests developed by the validator, so that product revisions and improvements can be made. Several questions have been revised, namely questions numbers 3, 6, 7, 8, 9, 12, 13, 15, 17, 18, 22, 23, and 29. Sentence editing and image revision are carried out to make the formulation of the questions clear and easy to understand by learners. According to Arikunto (2016), a clear formulation of questions can increase the quality of good items. After the product is repaired according to the validator's suggestions, the next step is a small product trial. Small-scale trials were carried out on class VIII D students, totaling 32 students at Junior High School 1 Tebas, by providing test products and student response questionnaires. Small-scale trials were conducted to find out the responses of students, how long it took them to answer questions, and the obstacles faced by students. The results of the student response questionnaire analysis can be seen in diagram 6.

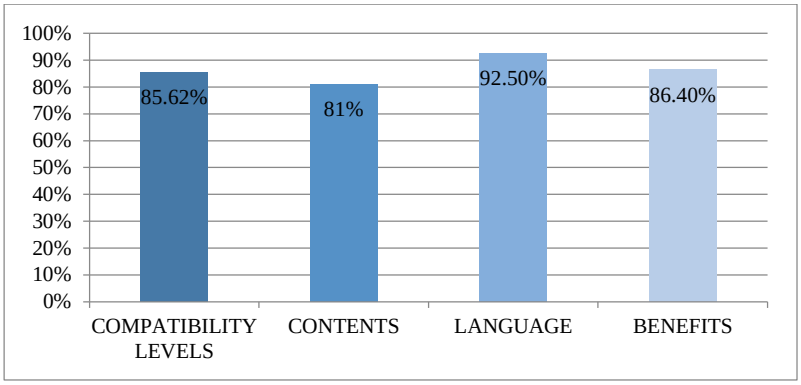


Figure 6. Diagram of student response questionnaire results

The results of small-scale trials obtained through response questionnaires got a very decent category from all aspects, with a percentage level of suitability of 85.62% in the very feasible category, 81% content aspects in the appropriate category, language aspects of 92.50%, and usefulness aspects of 86.26%. While reviews and suggestions by students regarding critical thinking ability test questions can be found, it can be concluded that the questions distributed are very challenging and require high analysis in order to do them. Some students think the questions are easy to understand, but some students think the questions are still difficult to understand because the answer choices are too long. In addition, regarding the time provided to work on the questions, based on the responses of students, it is known that students feel that the duration of time is in accordance with the number of questions that must be answered.

Responses and suggestions given by students are used as material to revise the developed test. After a small-scale trial, a large-scale trial was conducted on 120 students in 4 schools, namely at Junior High School 1 Tebas in class VIII F with as many as 32 students; Junior High School 3 Tebas in class VIII C with as many as 28 students, Junior High School 6 Tebas in class VIII C with as many as 30 students; and Junior High School 7 Tebas in class VIII B with as many as 30 students. In large-scale trials, students are asked to work on questions and the results will be analyzed for the quality of the items in the form of validity, reliability, level of difficulty, and discriminating power of questions using the Rasch model with Winstep software version 3.73, which will automatically know each value. The results of the construct validation analysis can be carried out by analyzing the MNSQ, ZSTD, and PT Measure-ALL Cor values received according to the quality criteria of the items. Based on data analysis, it was found that 27 item items were accepted because they met the requirements of the predetermined criteria, while 3 item items were rejected because they did not meet the ZSTD and PT Measure Cor outfit criteria, so they did not fit and the questions had to be discarded. The MNSQ outfit value is 0.99 in the person column and the items included in the fit criteria are in the range of 0.5 <MNSQ <1.5, meaning that the test can be used to measure students' critical thinking skills on substance pressure material in class VIII Junior High School. For the outfit, the ZSTD value obtained is 0 for a person, and 0.1 item is in the range -2.0<ZSTD<2.0 so it is possible that the data has a rational value. In Point Measure Correlation (PTMEA COR), the values obtained are all positive, so the items can be used to measure a test.

The reliability obtained through the Winstep 3.73 application shows the results of the person reliability obtained by 0.70. This shows the results of students' answers when taking tests are quite good. The results of Item reliability were obtained at 0.89, indicating that the quality of the test items is classified as good, while Cronbach's Alpha was obtained at 0.73, indicating the interaction or level of consistency between respondents and items included in the fairly high category. The reliability value taken for the test characteristics developed in this study is only the Cronbach alpha reliability value, with a minimum value of 0.60.

The level of difficulty of the items in the Rasch model is reviewed according to (Sumintono & Widhiarso, 2015), which can be classified into 4 categories based on the value of the measure, namely very easy items (<-1), easy items (-1 to 0), items are difficult (0 to 1), and items are very difficult (>1). The level of difficulty of the items was analyzed using Winstep version 3.73. The results of the analysis of the level of difficulty based on students' answers to the large-scale trial test that has been carried out can be seen in Table 2.

Table 2. The Results of the Analysis of the Difficulty of the Items

Question number	Difficulty Level	Category	Question number	Difficulty level	Category
1	0.55	Difficult	16	-0.49	Easy
2	-0.35	Easy	17	-0.26	Easy
3	-0.63	Easy	18	-0.44	Easy
4	0.12	Difficult	19	0.16	Difficult
5	-0.21	Easy	20	-0.17	Easy
6	1.05	Very difficult	21	-0.44	Easy
7	0.82	Difficult	22	0.82	Difficult
8	-0.84	Easy	23	0.00	Difficult
9	-0.17	Easy	24	0.78	Difficult
10	-1.02	Very easy	25	0.55	Difficult
11	-1.35	Very easy	26	0.85	Difficult
12	0.97	Difficult	27	0.82	Difficult
13	0.04	Difficult	28	-0.68	Easy
14	-0.84	Easy	29	-0.30	Easy
15	1.09	Very difficult	30	-0.04	Easy

Distinguishing power in this study can be seen from the value of the separation. The higher the separation value, the better the overall instrument quality between respondents and items, because it can recognize groups of respondents and items (Sumintono, B, & Widhiarso, 2015). Based on the Person Srata formula, we obtained H = [(4x2.87) + 1]/3 = 4.16, then rounded to 4. The results obtained indicate that there are four groups of items according to the level of difficulty received by respondents, categorized into very difficult, difficult, easy, and very easy. The next step is to look at the mean logit item obtained at 0.00 with a standard deviation of 0.66, which means that the discriminatory index is in a good category. For the characteristics of the test developed in this study, the discriminatory power used was at least 0.40 with sufficient category.

The next step is the revision of the final product after the critical thinking ability test is considered feasible to use because it has met the characteristics and good quality of the test. The final step is the final product, where at this stage a test development product has been obtained that can be used by teachers in measuring students' critical thinking skills on substance pressure material in class VIII Junior High School. The results of the recapitulation of the analysis of the quality of the tests developed so that they are feasible to use can be seen in Table 3.

Table 3. Recapitulation of the Results of the Analysis of the Quality of the Tests Developed

Question Number	Developed Test Quality				Information
	1	2	3	4	
1	Invalid	0.73	0.55	0,66	Thrown away
2	Valid	0.73	-0.35	0,66	Worn
3	Valid	0.73	-.0.63	0,66	Worn
4	Valid	0.73	0.12	0,66	Worn
5	Valid	0.73	-0.21	0,66	Worn
6	Valid	0.73	1.05	0,66	Thrown away
7	Invalid	0.73	0.82	0,66	Thrown away

Question Number	Developed Test Quality				Information
	1	2	3	4	
8	Valid	0.73	-0.84	0,66	Worn
9	Invalid	0.73	-0.17	0,66	Thrown away
10	Valid	0.73	-1.02	0,66	Thrown away
11	Valid	0.73	-1.35	0,66	Thrown away
12	Valid	0.73	0.97	0,66	Worn
13	Valid	0.73	0.04	0,66	Worn
14	Valid	0.73	-0.84	0,66	Worn
15	Valid	0.73	1.09	0,66	Thrown away
16	Valid	0.73	-0.49	0,66	Worn
17	Valid	0.73	-0.26	0,66	Worn
18	Valid	0.73	-0.44	0,66	Worn
19	Valid	0.73	0.16	0,66	Worn
20	Valid	0.73	-0.17	0,66	Worn
21	Valid	0.73	-0.44	0,66	Worn
22	Valid	0.73	0.82	0,66	Worn
23	Valid	0.73	0.00	0,66	Worn
24	Valid	0.73	0.78	0,66	Worn
25	Valid	0.73	0.55	0,66	Worn
27	Valid	0.73	0.85	0,66	Worn
28	Valid	0.73	0.82	0,66	Worn
29	Valid	0.73	-0.68	0,66	Worn
30	Valid	0.73	-0.30	0,66	Worn

Information:
1 = Validity (Valid, Invalid)
2 = Reliability (minimum Cronbach Alpha 0.60)
3 = Difficulty level of the item (Range -1 to 1)
4 = Distinguishing Power (Minimum 0.40)

Based on the recapitulation of the results of the developed test analysis, the number of questions that can be used is 23 questions out of 30 questions tested in the field. There are 7 item questions that are not suitable for use because 3 questions do not meet the criteria for fit validity and 4 questions have a very easy and very difficult level of difficulty so that these questions must be eliminated because they cannot be used to measure students' critical thinking skills.

Discussion

Critical thinking skills are very important in the world of education, and to help students develop these skills, it is necessary to create a learning environment that encourages critical thinking. This can be started by teaching students how to be active learners (Hartanto, 2020). In this case, the teacher plays an important role in improving critical thinking skills. As stated by Arifin (Marfhadella, 2021), students will not develop higher-order thinking skills if the teacher does not provide opportunities to develop and direct them at school. Based on the results of the needs analysis conducted through interviews, information was obtained regarding the need for high-level abilities, one of which is critical thinking skills, namely: 1) test instruments that can accommodate the 2013 curriculum and help students develop critical thinking skills; 2) instruments that can be implemented by teachers to students according to their characteristics; and 3) test instruments that meet measurement standards, are valid, and reliable. From the needs in the field, an urgently needed thinking ability test can be used in measuring students' critical thinking skills to help

students understand and improve critical thinking skills in learning physics with good quality items. This is based on the opinion (Zhou et al., 2013), that the instrument is said to be good if the results are reliable and able to assess what should be assessed. To make a good research instrument, it must first undergo several tests to ensure it meets the criteria for a good test.

The results of the validity analysis obtained indicate that the product is very feasible to use in terms of material validation, with an average overall aspect of 84.72%. While the validation of the questions calculated by content validation uses the Aiken formula. According to Retnawati (2016), the use of Aiken V Index to assess the entirety of each item with indicators that will be assessed using these items. Therefore, all items are declared valid according to the average Aiken index score of 0.78, which is included in the high category. Test assessments were also carried out by subject teachers in four schools by providing teacher response questionnaires. Aspects assessed in the teacher's response questionnaire include aspects of the level of suitability, content, language, and usefulness. Based on the teacher's response questionnaire, shows that the test is very feasible to use, with a feasibility percentage of 92.60%. Based on these results, it can be concluded that the test to measure students' critical thinking skills is very feasible to use so that it can be tested in the field.

The analysis of the instrument results provides empirical evidence that the test items are of good quality. This can be seen from the results of product trials. The results of the small-scale trial found that the tests that had been developed were included in the criteria for proper and good use in the evaluation of learning. The Rasch model can be used to determine the quality of a test instrument that can be measured (construct validity) (Baghaei, 2014). This is because the Rasch model can show items that are difficult for respondents to do (Abdul Aziz et al., 2014), and items that are associated with the ability of the respondents (Sumintono, B, & Widhiarso, 2015).

The results of large-scale trials with 120 students. In analyzing the validity data obtained using the Rasch model on the Winstep program version 3.73. The Rasch model can produce a wider variety of results when used to test the validity of an instrument because the validity of an instrument can be tested using various criteria (Othman et al., 2014). The results of the validity showed that 27 of the 30 items were accepted because they met the requirements of the predetermined criteria. However, 3 items were rejected, namely 1, 7, and 9, because the ZSTD and PT Measure Cor outfit values did not meet the predetermined criteria. The MNSQ outfit value obtained from 30 questions is between 0.79 to 1.21 with a mean of 0.99. This value is included in the fit criteria because it lies within the interval $0.5 < MNSQ < 1.5$. The ZSTD outfit value obtained by the items is between -1.9 to 4.0. Within that range, several questions are included in the criteria that do not fit because they do not fit the range of $-2.0 < ZSTD < +2.0$, namely the number of questions 1.7 and 9. While for Pt Measure, some questions do not meet the criteria or do not fit, specifically questions 1, 2, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 16, 23, 24, 25, 26, 27 and 30 that score less than 0.4 and thus do not match the value range of 0.4 Pt Measure Corr 0.85. However, item numbers 2, 3, 4, 5, 10, 11, 12, 13, 14, 16, 23, 24, 25, 26, 27 and 30 can still be accepted and maintained due to the MNSQ outfit value and the ZSTD outfit value being in the range that matches the criteria. While questions 1.7 and 9 were declared unfit because they did not meet the requirements for the ZSTD outfit and PT Measure-ALL Cor. Based on available data and information, it appears that respondents who are not suitable (misfit) cannot be included in the study or must be eliminated. According to Sumintono & Widhiarso (2015), an item is said to be valid or fit into the model if it meets at least two of the three criteria that must be met. Thus, based on the results of the validity of the items, 3 items are said to be invalid because they do not meet the item quality criteria, so they must be discarded. In line with the opinion of Tadeko (2013), suggests that theoretically, the item is not appropriate for measuring students' critical thinking skills.

The pattern of respondents' answers can be measured if the MNSQ infit value is 1.00 and the MNSQ outfit is 0.99. This indicates that the respondent's answer pattern on the

overall respondent's answer pattern instrument is good. According to Ardiyanti (2016), the MNSQ infit and outfit MNSQ values, which are close to 1.00 are better. For a person, the ZSTD infit value is 0.1 and the ZSTD outfit is 0.1, meaning that the overall pattern of respondents' answers is in line with the model because the infit ZSTD and ZSTD outfit values are said to be good if they are closer to 0.0 (Ardiyanti, 2016).

In addition to meeting the validity requirements, the test instrument must also be reliable to produce accurate results. According to the superiority of reliability in the Rasch model, the reliability of the instrument can be ascertained from three variables namely, person reliability, item reliability, and Cronbach's alpha. Based on the value of Cronbach's alpha on the reliability of the items obtained, the instrument is said to have a fairly high consistency level, therefore be used for research because it has met the minimum requirement of > 0.7 (Bond & Fox, 2007); Sumintono & Widhiarso, 2013). A significant difference in value indicates that the reliability of the learner's person is low, so it is categorized as sufficient. This shows that the level of consistency of students' answers to questions about critical thinking skills on substance abuse material is included in the weak or less consistent category. And for the reliability of the test items that have been developed to have a high consistent level, it means that the quality of the items in the test has a good reliability aspect.

According to Hambleton & Swaminathan (Istiyono et al., 2014), if the difficulty level of a test varies, the test is considered good in its difficulty level. At the item difficulty level of 30 items of critical thinking skills of students on substance pressure material in class VIII Junior High School, there are 2 items (7%) in the very difficult category, namely questions number 6 and 15. In the difficult category, there are 12 items (40%) namely numbers 1, 4, 7, 12, 13, 19, 22, 23, 24, 25, 26 and 27. In the easy category, there are 14 questions (46%) namely numbers 2, 3, 5, 8, 9, 14, 16, 17, 18, 20, 21, 28, 29, and 30, while in the very easy category, there are 2 items (7%) namely numbers 10 and 11. Item number 15 is one of the most difficult questions because 49 out of 120 students can answer correctly. While question number 11 is included in the category of very easy questions because 103 of the 120 students were able to correctly answer it, According to Bistok Sirait (Hanifah, 2017), the level of difficulty of the questions occurs because of the complexity of the topic and the conditions of the student's answers. This is also supported by the opinion (Azis, 2016) which states that the level of difficulty of the questions can be seen from the number of test respondents who answered the question correctly, compared to the number of students who took the test. The more questions are answered, the easier it is to classify the questions. Conversely, the fewer questions answered, the more difficult the questions become. Rasch observed that the ability of students to answer correctly on dichotomous questions can be compared with the level of difficulty of the questions, according to Wahyuni (2015). The characteristics of the tests that have been developed in this study are based on the level of difficulty of the items used in the range of -1 to 1 with the criteria of easy and difficult items. Very easy and very difficult questions cannot be used as a good measure of the ability of students. As supported by the opinion of Daryanto (Yani et al., 2014), a good question is a question that is not too difficult or too easy to answer. Because very easy questions do not develop students in increase efforts to solve them, while very difficult questions will make students become desperate and not have the enthusiasm and effort to try again because it is beyond their capabilities.

The discriminating power of items in Rasch modeling is used to analyze the level of individual abilities of students who can answer high questions or low questions. Identifying the respondent group can also be done with the respondent separation index. The discriminatory index is in a good category. The value of the separation item is 4.16. According to Linacre (Napitupulu, 2017), a separation index that exceeds 2 has a good value. In line with the opinion of Sumintono & Widhiarso (Erfan et al., 2020), said that the higher the item separation value, the better the quality or quality of the instrument for the respondents and the overall items.

CONCLUSION

In general, the critical thinking ability test using the Rasch model analysis in this study is appropriate to be used in measuring students' critical thinking skills in the substance pressure material for class VIII SMP in Tebas District. The results of the feasibility test from expert validation showed that from the material expert validation it obtained feasibility with a score of 84.72% in the very feasible category, while for question validation it obtained eligibility with an Aiken index of 0.78 which was included in the high category. For the practicality test aspect, which is based on the teacher and student response questionnaire, it is stated that it is very suitable to be used for testing in the field. Meanwhile, for the quality of the test items based on the validity based on the MNSQ, ZSTD, and PT Measure Corr scores, 27 items were declared valid because they had met the criteria. If the level of difficulty of the items used is in the range of -1 to 1 with the criteria of difficult and easy items, so that 4 questions must be eliminated because they are included in very difficult and very easy questions, then Rasch modeling has distinguishing power with categories both with the separation index of items H = 4.16 which indicates that this item has a good response distribution pattern so that the developed test is feasible or valid to be used in measuring students' critical thinking skills.

RECOMMENDATION

Students' critical thinking skills need to be developed and improved in schools. Therefore, in evaluating learning, teachers are expected to use critical thinking skills tests as an alternative tool for practice questions for students. On the other hand, students must be actively involved in learning and not only focus on the material being studied but also develop concepts and practice questions to improve critical thinking skills. And lastly, for related researchers, it is recommended to use multiple choice tests by adding open reasons to get the latest forms and reasons why students answer the answer choices.

ACKNOWLEDGMENTS

The completion of this research cannot be separated from the help of various parties. On this occasion, the authors would like to thank all those who have provided support, criticism, and suggestions in carrying out the research; in particular, the principal and the physics teacher at the researcher's place, who have been willing to give permission and help research so that this research can be completed.

REFERENCES

Abdul Aziz, A., Jusoh, M. S., Amlus, M. H., Omar, A. R., & Awang Salleh, T. S. (2014). Construct Validity: A Rasch Measurement Model Approaches. *Journal of Applied Science and Agriculture*, *9*(12), 7–12. https://www.researchgate.net/publication/266676182

Ardiyanti, D. (2016). Aplikasi Model Rasch pada Pengembangan Skala Efikasi Diri dalam Pengambilan Keputusan Karir Siswa. *Jurnal Psikologi*, *43*(3), 248. https://doi.org/10.22146/jpsi.17801

Arifin, Z. (2017). Mengembangkan Instrumen Pengukur Critical Thinking Skills Siswa pada Pembelajaran Matematika Abad 21. *Jurnal THEOREMS (The Original Research of Mathematics)*, *1*(2), 92–100.

Arikunto, S. (2016). *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Bumi Aksara.

Azis, A. (2016). Analisis Tes Buatan Guru Bidang Studi Matematika Kelas V SD 1 Katobengke. *Edumatica : Jurnal Pendidikan Matematika*, *6*(1), 15–24. https://doi.org/10.22437/edumatica.v6i01.2998

Azwar. (2015). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Baghaei, P. (2014). The Rasch Model as a Construct Validation. *Rasch Measurement Transaction*, *22*(1).

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model : Fundamental Measurement in the HumBond, T. G., & Fox, C. M. (2007). Applying the Rasch Model : Fundamental Measurement in the Human Sciences Second Edition University of Toledo.an Sciences Second Edition University of Toledo*.

Daniati, N., Handayani, D., Yogica, R., & Alberida, H. (2018). Analisis Tingkat Kemampuan Berpikir Kritis Peserta Didik Kelas VII SMP Negeri 2 Padang tentang Materi Pencemaran Lingkungan. *Atrium Pendidikan Biologi*, *1*(2), 1–10.

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19.

Facione, P. a. (2011). Critical Thinking : What It Is and Why It Counts. In *Insight assessment* (Issue ISBN 13: 978-1-891557-07-1.). https://www.insightassessment.com/CT-Resources/Teaching-For-and-About-Critical-Thinking/Critical-Thinking-What-It-Is-and-Why-It-Counts/Critical-Thinking-What-It-Is-and-Why-It-Counts-PDF

Facione, P. a. (2013). *Criticatal Thinking: What It Is and why It Counts*. The California Academic Press.

Facione, P. a. (2015). *Critical Thinking: What It Is and Why It Counts*. The California Academic Press.

Febriano, R., Tandililing, E., & Enawaty, E. (2021). Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Matematis Dengan Menggunakan Analisis Model Rasch Pada Siswa SMP. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, *10*(9), 1–12. https://jurnal.untan.ac.id/index.php/jpdpb/article/viewFile/49587/75676590680

Hanifah, N. (2017). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Sosio E-KONS*, *6*(1).

Hartanto. (2020). Pengembangan Instrumen Assesmen Kemampuan Berpikir Kritis Siswa Kelas V SeKolah Dasar Pada Pelajaran Tematik. *Akrab Juara*, *5*(1), 43–54.

Istiyono, E., Mardapi, D., & Suparno, S. (2014). PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA (PysTHOTS) PESERTA DIDIK SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Jamaluddin, J., Jufri, A. W., Muhlis, M., & Bachtiar, I. (2020). Pengembangan Instrumen Keterampilan Berpikir Kritis Pada Pembelajaran IPA di SMP. *Jurnal Pijar Mipa*, *15*(1), 13–19. https://doi.org/10.29303/jpm.v15i1.1296

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

Mardapi, D. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Nuha Medika.

Marfhadella, P. (2021). *Pengembangan Assesment Untuk Mengukur Kemampuan Berpikir Kritis Materi Interaksi Makhluk Hidup Dengan Lingkungan Di SMP Se-Kota Bengkulu*. Institut Agama Islam Negeri (IAIN) Bengkulu.

Napitupulu, D. (2017). Rasch Model Analysis on Acceptance Instrument of Optical Marker for Safe Vehicle Marking. *Jurnal Penelitian Transportasi Darat*, *19*, 37–48.

Novinda, M. R. R., Silitonga, H. T. M., & Hamdani. (2019). Pengembangan tes pilihan ganda menggunakan model Rasch materi gerak lurus kelas X Pontianak. *Jurnal Pendidikan Dan Pembelajaran*, *8*(6), 1–11. https://jurnal.untan.ac.id/index.php/jpdpb/article/view/33452

Othman, N., Salleh, S. M., Hussin, H., & Wahid, H. A. (2014). Assessing Construct Validity and Reliability of Competitiveness Scale Using Rasch Model Approach. *The 2014 WEI International Academic Conference Proceedings*, 113–120.

Rahaju, E. B., Fardah, D. K., Wijayanti, P., & Ismail. (2020). Kemampuan Guru-guru Matematika SMP Kabupaten Ponorogo dalam mengembangkan soal Berpikir Tingkat

Tinggi. *Jurnal Pendidikan Matematika Raflesia*, *05*(01), 75–81.

Ramadhan, M.A., & Iriani, T. (2013). Pembelajaran Berbasis Softkills, Sudahkah Dilakukan Guru? *Paper Presented at Prosiding Semnas Politik Pendidikan Nasional: Program Pascasarjana UNY*.

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Parama Publishing.

S Wahyuni, K. M. (2015). Analisis Kemampuan Peserta didik dengan Model Rasch. *In Seminar Nasional Evaluasi Pendidikan*.

Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model Untuk Mengembangkan Instrumen Pengukuran Kemampuan Berikir Kritis Siswa Pada Topik Ikatan Kimia. *Jurnal Zarah*, *6*(2), 68–75. https://doi.org/10.31629/zarah.v6i2.724

Satria Mukti, T., & Istiyono, E. (2018). Instrument for Assessing the Critical Thinking Ability of X Grade High School Students on Biology Learning. *BIOEDUKASI: Jurrnal Pendidikan Biologi*, *11*, 105–110. http://dx.doi.org/10.20961/bioedukasi-uns.v11i2.21624

Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Jurnal Manajemen & Pendidikan Islam*, *2*(2), 192–213. www.depdiknas.go.id/evaluasi-proses-

Sugiyono. (2016). *Metode Penelitian Pendidikan Pendidikan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Sumintono, B, & Widhiarso, W. (2013). *Aplikasi Model Rasch Pada Assesment Pendidikan*. Tim Komunikata Publishing House.

Sumintono, B, & Widhiarso, W. (2015). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Tim Komunikata Publishing House.

Tadeko, N., & S. (2013). Analisis Pemahaman Soal Grafik Mahasiswa Pendidikan MIPA Menggunakan Pemodelan Rasch. *Jurnal Inovasi Dan Pembelajaran Fisika*, *1*(1).

Widana, I. W., Adi, S., Herdiyanto, H., Abdi, J,. Marsito, M., & Istiqomah, I. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Matematika*. Direktorat Pembinaan Sekolah Menengah Atas.

Yani, A., Asri, A. F., & Burhan, A. (2014). Analisis Tingkat Kesukaran, Daya Pembeda Dan Fungsi Distraktor Soal Ujian Semester Ganjil Mata Pelajaran Produktif Di Smk Negeri 1 Indralaya Utara Tahun Pelajaran 2012/2013. *Jurnal Pendidikan Teknik Mesin*, *1*(2), 98–115.

Yulian, V. N. (2016). Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Melalui Metode Pembelajaran Metode Inkuiri Berbantuan Software Algebrator. *Jppm*, *9*(1), 20–24.

Zhou, Q., Huang, Q., & Tian, H. (2013). Developing Students' Critical Thinking Skills by Task-Based Learning in Chemistry Experiment Teaching. *Creative Education*, *04*(12), 40–45. https://doi.org/10.4236/ce.2013.412a1006.