Research and Development in Education (RaDEn)

Research Article

Analysis the quality of critical thinking and creativity questions in high school biology subjects with the Rasch model

Tasya Novian Indah Sari [a,1,*], **Anna Rakhmawati** [a,2]

[a] Biology Education Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Special Region of Yogyakarta 55281, Indonesia
[1] tasyanovian.2022@student.uny.ac.id *; [2] anna_rakhmawati@uny.ac.id
* Corresponding author

Abstract: A good instrument is an instrument that can measure students' abilities accurately. The research aims to analyze the quality of critical thinking and creativity items on water pollution material in terms of validity, reliability, and level of difficulty using the Rasch model. The research used a quantitative descriptive method involving biology learning experts and 173 class X high school students. The research was carried out in December 2023-January 2024. The data collection method used expert assessment sheets and critical thinking and creativity essay questions. Data were analyzed using the Rasch model assisted by Quest. The results showed that of the 10 critical thinking essay questions, 3 questions were invalid, while of the 8 creativity essay questions, 1 question was invalid based on the suitability of the INFIT MNSQ and OUTFIT T scores. The results of the reliability test with Cronbach's alpha showed that the reliability value of the critical thinking instrument was 0.74 and creativity 0.79. Average difficulty level of questions with threshold values in the medium and difficult categories. Overall, the instrument was feasible and question items that did not fit the Rasch model were removed.

Keywords: assessment instruments; item response theory; level of difficulty; reliability; validity

1. Introduction

The 21st century is marked by the rapid development of science, globalization, economics, and technology. Education today needs to prepare students who can face developments in the 21st century (Thornhill-Miller et al., 2023). One of the important 21st century skills that students have is the ability to think critically and creatively (Siahaan et al., 2023; Zainil et al., 2023). High-level thinking skills, especially critical and creative thinking skills, are very necessary for utilizing students' knowledge for problem-solving and decision making in various areas of life (Amin et al., 2020; Kardoyo et al., 2020; Simanjuntak et al., 2021; Sumarni & Kadarwati, 2020). Creative thinking skills relate to students' ability to understand problems and find solutions to various existing strategies or methods, while critical thinking skills help students to overcome information that does not have a strong basis and make decisions that are meaningful and based on valid information (Kardoyo et al., 2020).

The results of a survey conducted by researchers regarding critical thinking skills and creativity in biology learning show that the critical thinking skills of high school students in Magetan Regency are 50% in the medium category and 50% in the low category, while creativity is 64% in the low category and 36% in the medium category. The results of this survey are supported by several previous studies that the critical thinking skills of students in Indonesia are still relatively low (Meryastiti et al., 2023; Pantiwati et al., 2022; Satria et al., 2023) and creativity skills are also still relatively low (Leasa et al., 2021; Priyanto & Dharin, 2021). The results of the survey and several previous studies are also supported by interviews with biology teachers and deputy principals for curriculum at one of the public high schools in Magetan Regency to encourage students to think critically and creatively, the school has provided outreach to teachers to present questions that not just a short answer. The short answer questions presented make students less accustomed to thinking. At a minimum, the questions presented by the teacher contain discourse and familiarize students with reading this discourse. Despite this, teachers still do

Citation: Sari, T. N. I & Rakhmawati, A. (2024). Analysis the quality of critical thinking and creativity questions in high school biology subjects with the Rasch model. *Research and Development in Education (RaDEn)*, *4*(1), 602-616. https://doi.org/10.22219/raden.v4i1.32758

Received: 18 March 2024
Revised: 22 April 2024
Accepted: 3 May 2024
Published: 7 June 2024

not implement this. Interview results show that there are still many teachers who do not use critical or creative thinking questions in learning and tend to use questions that have been provided by publishers. Time limitations and school administrative burdens make teachers choose to use existing questions rather than create questions that support students' higher-level thinking abilities. Therefore, one of the hopes of teachers and deputy principals in the curriculum field is that they can present an instrument that familiarizes students with questions that are not just answered briefly.

Test instruments are part of learning outcomes that are important in determining the success of learning (Astuti et al., 2023). To obtain the right data or information, an instrument is needed that can provide the information you want to know. A good instrument is an instrument whose results are valid and reliable so that it can measure students' abilities accurately (Darmana et al., 2021; Sudihartinih & Prabawanto, 2020). The validity and reliability of the test instrument is important. Instrument validity is an absolute requirement to produce a valid instrument. Valid means being able to measure what should be measured (Sudihartinih & Prabawanto, 2020), while reliability describes that a test must be able to measure something consistently that can be relied upon or trusted (Fietri et al., 2021). Apart from validity and reliability, other aspects that support the quality of test instruments are differentiation and level of difficulty. According to Tarmizi et al. (2021) the different power and level of difficulty of the questions is something that must be included in the process of analyzing the questions so as to obtain a quality test instrument.

Analysis of test instruments can be carried out using various approaches, one of which is classical test theory. In classical test theory, the aspects that really determine the quality of the questions are the level of difficulty and the distinguishing power of the questions. However, the characteristics of the questions produced by classical test theory change according to the student's abilities (Susdelina et al., 2018). To overcome the weaknesses of classical theory, an alternative approach emerged, namely the Rasch model (Darmana et al., 2021).

Rasch model analysis is a statistical analysis of suitability that provides information that the data obtained is ideal or illustrates that people who have high abilities provide answer patterns to items according to their level of difficulty (Rizbudiani et al., 2021). Analysis of instrument quality with the Rasch model has several advantages compared to analysis with classical test theory, namely the Rasch model can identify wrong answers, inaccurate assessments and predict missing data (Hamdu et al., 2020). Besides that, Lidinillah et al. (2020) said that instrument quality analysis using the Rasch model is able to measure the number of items and respondents independently, meaning that the data produced does not depend on the number of items or the number of respondents.

Apart from the importance of instrument quality analysis, the trend of research on instrument quality analysis using the Rasch model in Indonesia is still minimal. For example, research results Prasetya and Pratama (2023) analysis of the quality of critical thinking essay questions using the Rasch model on the digestive system material obtained good results for measuring the critical thinking skills of high school students. Research by Sa'diyah et al. (2020) that the results of the analysis of the quality of critical thinking multiple choice questions using the Rasch model are valid and reliable for measuring junior high school students' critical thinking skills on physics concepts. The study results were similar Tania et al. (2021) it was reported that the analysis of the validity and reliability of the critical thinking essay question instrument on ecological material obtained good results. In addition to the Rasch model analysis on critical thinking instruments, studies conducted by R. R. Sari et al. (2022) that the Rasch model analysis on creativity instruments for teachers also provides valid and reliable results. Research by Eliaumra et al. (2022) analysis of the quality of diagnostic instruments with the Rasch model to measure the creativity of prospective biology teachers provides good results. Finally, the use of item quality analysis using the Rasch model on biology concept creativity question instruments for biology teachers also provides valid and reliable results (Bui et al., 2020).

Based on the results of previous research, the use of the Rasch model for analyzing critical and creative thinking items shows positive results. The use of the Rasch model for item analysis in biology learning assessment has been used in middle school, high school, and university students and teachers. The instruments developed generally take the form of essay questions and multiple-choice questions. For junior and senior high school levels, the use of the Rasch model for analyzing creativity instruments is still minimal, this is shown by the findings of previous research results in the last 5 years, while critical thinking regarding biological material is still limited. Apart from that, existing instruments for critical thinking and creativity in biology learning refer to the 2013 curriculum, whereas currently, they have begun to adapt to the independent curriculum. So, we need an instrument that is valid and reliable and can help teachers in implementing the independent curriculum and familiarizing students with critical and creative thinking. Furthermore, analysis of the quality of instruments is quite crucial because it can determine whether the instrument is effective or not in measuring student abilities. To be able to describe students' critical thinking abilities and creativity optimally, a truly appropriate instrument is needed. Therefore, based on the results of the analysis of field needs in schools and previous research studies, this research aims to analyze the quality of critical thinking and creativity items in terms of validity, reliability, level of difficulty, and differences in each instrument on water pollution material.

2. Materials and Methods

2.1 Types of research

The research was carried out as a unit of Analyze, Design, Develop, Implementation and Evaluation (ADDIE) research stages in developing student worksheets Project Based Learning (PjBL) based on local potential using quantitative descriptive methods. The research was carried out from December 2023 to January 2024.

Table 1. Distribution of critical, creative thinking indicators and water pollution materials

Indicators of critical thinking skills	Water pollution materials	Item number
Critical thinking skills		
Build basic skills	Characteristics of water pollution	6
	Impact of water pollution	7
Making further explanations	Characteristics of water pollution	4
	Impact of water pollution	2
Making conclusions	Characteristics of water pollution	1 & 5
Giving simple explanations	Causes of water pollution	3 & 10
Making estimates along with integration	Efforts to overcome water pollution	8 & 9
Creativity		
Fluency	Causes of water pollution	1 & 2
Flexibility	Impact of water pollution	3 & 4
Originilaty	Efforts to overcome water pollution	5 & 6
Elaboration	Efforts to overcome water pollution	7 & 8

2.2 Research subject

The subjects of this research were two biology learning experts from the Faculty of Mathematics and Natural Sciences, Yogyakarta State University and 173 students of SMAN 1 Karas, Magetan Regency. The qualifications of the biology learning experts

involved in this research are 1) minimum Doctoral or Ph.D., 2) scientific concentration in the environmental field and 3) scientific concentration in education or assessment.

2.3 Research Methods and Instruments

Data collection methods were carried out using non-test and test methods. The non-test instrument used in this research was a learning expert assessment sheet regarding critical thinking skills and creativity instruments (Table 1). The learning expert's assessment of critical thinking and creativity instruments was carried out by placing a checklist on each question item based on the assessment indicators in Table 1. Question items that did not meet the indicators in Table 1 were corrected by the researcher before conducting the field trial. Furthermore, Table 2 present instrument assessment grid used in this study.

The test instruments in this research were 10 essay questions on critical thinking skills and 8 essay questions on creativity skills on water pollution material presented in Indonesian. Indicators of critical thinking in this research include 1) building basic skills, 2) making further explanations, 3) making conclusions, 4) giving simple explanations and 5) making estimates along with integration (Ennis, 2011), while creativity indicators include 1) fluency, 2) flexibility, 3) originality and 4) elaboration (Torrance, 1969). Testing of the question instruments was carried out online with the help of the Google Form platform.

Table 2. Instrument assessment sheet grid by biology learning experts

No	Aspect	Assessment Indicators
1	Construction aspects	The questions are formulated logically
		Question items are formulated according to the question indicators
		The question items do not give rise to multiple interpretations
		The questions do not provide clues to the answer
		The answer to a question item does not depend on the previous answer
		Correspondence between question items and levels of Bloom's taxonomy
		Correspondence between question items and indicators of critical thinking/creativity
		The images, data, or graphs presented in the questions function well
2	Language aspects	The questions use language that is by Indonesian language rules
		The questions use language that is easy to understand and communicative
3	Material aspects	The material presented is appropriate for the level of education
		The material presented is clear and by standard concepts
		Accuracy in using biological terms
		The questions in the questions are formulated according to the question indicators in the grid
		The assessment aspects are by the assessment rubric on the grid

2.4 Data analysis technique

Data from trials of critical thinking skills and creativity were analyzed using the Rasch model. The validity of critical thinking and creativity instruments is based on the INFIT MNSQ and OUTFIT T fit value. According to Setyowarno (2017) INFIT MNSQ and OUTFIT T fit values can be used to compare the determination of each item or item with the criteria model shown in Table 3 and Table 4. In addition, reliability analysis uses Cronbach's alpha value (Table 5). Finally, the Rasch model analysis also provides a ranking of the difficulty level of the questions with threshold values (Table 6). Rasch model analysis assisted by QUEST software.

Table 1. INFIT MNSQ conditions Rasch model

Value INFIT MNSQ	Information
> 1.33	Does not fit the Rasch model
0,77 – 1.33	Fits the Rasch model
< 0.77	Does not fit the Rasch model

Source: Setyowarno (2017)

Table 2. OUTFIT T conditions Rasch model

Value OUTFIT T	Information
OUTFIT T ≤ 2.00	Matches the Rasch model / pass questions
OUTFIT ≥ 2.00	Not suitable for the Rasch model / fallout questions

Source: Setyowarno (2017)

Table 3. Instrument reliability index criteria

Reliability index	Reliability criteria
0.00 – 0.19	The degree of reliability is very low
0.20 – 0.39	Low degree of reliability
0.40 – 0.59	Medium degree of reliability
0.60 – 0.79	High degree of reliability
0.80 – 1.00	The degree of reliability is very high

Source: Sumardi (2020)

Table 4. Threshold value (difficulty level of question items)

Value Threshold	Criteria
b > 2	Very difficult
1 < b ≤ 2	Difficult
-1 ≤ b ≤ 1	Medium
-1 > b ≥ -2	Easy
b < -2	Very easy

Source: E. D. K. Sari and Mahmudi (2024)

3. Results

3.1 Empirical Validity (Expert Judgment)

The instrument prepared in this research is an instrument for critical thinking and creativity on water pollution material for class X or phase E students in Indonesian. The critical thinking and creativity question instruments developed were in the form of essay questions and each consisted of 10 questions for the critical thinking question instrument

(Table 1) and 8 questions for the creativity question instrument (Table 1). The question instrument is equipped with a critical thinking and creativity instrument grid and an essay question assessment rubric with a maximum score of 4 for each number, starting from the range 0 to 4. Before testing and analyzing the Rasch model, the critical thinking and creativity question instrument is carried out empirically with expert opinion on each instrument.

The biology learning experts involved in this research are two lecturers from the Faculty of Mathematics and Natural Sciences, Yogyakarta State University who have doctoral qualifications in the fields of environmental science and education. Experts assess critical thinking skills and creativity instruments from two aspects, namely material and assessment (Table 2). The expert provides an assessment of each question item based on its suitability to the expert assessment indicators in Table 2. The suggestions and input provided by the expert in this research are used as material for improvement to present appropriate critical thinking and creativity instruments from the assessment and material aspects (Table 7).

Table 5. Expert advice and input on critical thinking and creativity instruments

Instrument	Expert advice and input
Critical thinking	Correct grammar according to Enhanced Indonesian spelling
	Relate the material to other life in the water, not just humans
	Add quality standards according to the type of pollution as a reference for students
	The water pollution presented is more varied, not just water pollution due to leather industry waste
Creativity	Correct grammar according to Enhanced Indonesian spelling
	Some items about creativity do not reflect indicators of creativity

Pay attention to the results of the following research!

Based on the results of Arifah Dwi Astuti research regarding water quality in the Gandong River area, Magetan Regency, the following data were obtained.

No	Parameters	Result
1.	pH	7.5
2.	BOD	115 mg/l
3.	COD	274 mg/l
4.	TSS	314 mg/l
6.	Oil and fat	5 mg/l
5.	Ammonia	872 mg/l
6.	Chrome	314 mg/l

Based on the data obtained above, please make conclusions regarding the water quality in the area based on the facts presented.

Figure 1. Example of question items before expert revision (critical thinking question instrument)

Examples of improvements made are in Figures 1 and Figure 2. Figure 1 presents critical thinking questions before they were corrected based on expert input. Figure 1 presents environmental pollution questions related to several water pollution parameters. In this question, students are asked to conclude research results based on environmental quality standards. The expert provided input that to discuss environmental quality standards, the standards used must be determined based on the aquatic environment. Each body of water has different quality standards, so in Figure 2 the questions that have been corrected by the researcher present the environmental quality standards for the textile

industry. With the help of environmental quality standards for the textile industry, students are expected to be able to provide further explanations based on research results on questions with environmental quality standards.

Critical and creative thinking question instruments that have been declared feasible by experts will then be subjected to limited trials. A limited trial was carried out at SMAN 1 Karas, Magetan Regency, involving 173 students. The trial was carried out with the help of the Google Forms platform. After the trial was carried out, the trial results were analyzed using the Rasch model assisted by QUEST software.

Pay attention to the results of the following research, to answer questions number 4 and 5!

Based on the results of research conducted by Arifah Dwi Astuti regarding the influence of the disposal of leather processing industry waste in the waters of the Gandong River, Regency

Magetan obtained the following data. Research conducted by Arifah Dwi Astuti used 6 water chemical parameters (Table 1). To find out more about the water quality in the Gandong River, the results of this research must be compared with quality standards, one of which is the quality standard for textile industry waste which has been set by the Ministry of Environment and Forestry of the Republic of Indonesia (Table 2).

Table 1. Research Results on Gandong River Water Quality

No	Parameters	Result
1.	pH	7.5
2.	BOD	115 mg/l
3.	COD	274 mg/l
4.	TSS	314 mg/l
6.	Oil and fat	5 mg/l
5.	Ammonia	872 mg/l
6.	Chrome	314 mg/l

Table 2. Standard Quality Standards for Textile Industry Wastewater

No	Parameters	Water discharge (m3/day)		
		≤100	100 < x < 1000	≥ 1000
1.	BOD (mg/L)	60	45	35
2.	COD (mg/L)	150	125	115
3.	TSS (mg/L)	50	40	30
4.	Total phenols (mg/L)	0.5	0.5	0.5
5.	Total chromium (mg/L)	1	1	1
6.	Total ammonia (mg/L)	8	8	8
7.	Sulfide (mg/L)	0.3	0.3	0.3
8.	Fat oil	3	3	3
9.	pH	6-9	6-9	6-9
10.	Color (pt-co)	200	200	200
11.	Maximum discharge (m3/tonne product)	100	100	100

Based on the research data obtained in Table 1, please provide further explanation regarding the water quality of the Gandong River based on the research results in Table 1 with the textile industry wastewater quality standards in Table 2.

Figure 2. Example of question items after expert revision (critical thinking question instrument)

3.2 Validity of Question Items

The results of the suitability analysis of the item fit for the critical thinking instrument with INFIT MNSQ and OUTFIT T showed that of the 10 questions prepared (Table 8), 3 questions did not fit the Rasch model and these questions were discarded. The items that do not fit the Rasch model are item number 4, item number 5, and item number 6. Item number 4 received an INFIT MNSQ value of 0.72 < 0.77 (not suitable for the Rasch model) and an OUTFIT T value of -2.0 < 2.00 (fits the Rasch model). Item number 5 obtained an INFIT MNSQ value of 0.72 < 0.77 (does not fit the Rasch model) and an OUTFIT T value of -2.3 (fits the Rasch model). Item number 6 obtained an INFIT MNSQ value of 1.41 < 1.33 (fits the Rasch model) and an OUTFIT T value of 4.4 > 2.00 (does not fit the Rasch model). Based on the suitability of the INFIT MNSQ and OUTFIT T values, items number 4 and 5 do not meet the minimum limit value of INFIT MNSQ value < 0.77, while item number 6 does not meet the maximum limit value of OUTFIT T value > 2.00. So, the number of

critical thinking questions that were declared valid was 7 questions. Even though several question items were dropped, there were still indicators of critical thinking and water pollution sub-materials in the questions used.

The results of the suitability analysis of the item fit for the creativity instrument with INFIT MNSQ and OUTFIT T showed that of the 8 items prepared (Table 9), there was 1 item that did not fit the Rasch model and this item was discarded. The items that do not match the Rasch model are item number 1. Item number 1 has an INFIT MNSQ value of 1.46 > 1.33 (not suitable for the Rasch model) and an OUTFIT T value of 2.2 > 2.00 (not suitable with Rasch model). Based on the suitability of the INFIT MNSQ and OUTFIT T values, item number 1 does not meet the maximum limit value of INFIT MNSQ value > 1.33 and does not meet the maximum limit value of OUTFIT T value > 2.00. So, the number of creativity questions that were declared valid was 7 questions. Even though several question items were dropped, there were still indicators of creativity and water pollution sub-materials in the questions used.

Table 6. INFIT MNSQ and OUTFIT T values for critical thinking items

Item	INFIT MNSQ	OUTFIT T	Information
1	0.82	-1.8	Matches the Rasch model / pass questions
2	1.03	0.1	Matches the Rasch model / pass questions
3	1.13	1.0	Matches the Rasch model / pass questions
4	0.72	-2.0	Not suitable for the Rasch model / fallout questions
5	0.72	-2.3	Not suitable for the Rasch model / fallout questions
6	1.41	4.4	Not suitable for the Rasch model / fallout questions
7	0.98	-0.7	Matches the Rasch model / pass questions
8	1.11	0.7	Matches the Rasch model / pass questions
9	0.99	-0.1	Matches the Rasch model / pass questions
10	1.08	0.5	Matches the Rasch model / pass questions

Table 7. INFIT MNSQ and OUTFIT T values for creativity questions

Item	INFIT MNSQ	OUTFIT T	Information
1	1.46	2,2	Not suitable for the Rasch model / fallout questions
2	0.92	-1.1	Matches the Rasch model / pass questions
3	0.92	0.00	Matches the Rasch model / pass questions
4	0.85	-1.5	Matches the Rasch model / pass questions
5	1.07	1.0	Matches the Rasch model / pass questions
6	0.92	-0.9	Matches the Rasch model / pass questions
7	0.98	1.4	Matches the Rasch model / pass questions
8	0.86	-1.7	Matches the Rasch model / pass questions

3.3 Instrument Reliability

Apart from the suitability of item fit, the next item analysis is the reliability test. The results of the reliability analysis of the critical thinking instrument using the Rasch model show an instrument reliability value of 0.74 or the reliability of the critical thinking instrument is in the good category and the reliability value of the creativity instrument is 0.79 or the reliability of the creativity instrument is in the good category (Table 10).

Table 8. Reliability value of critical thinking and creativity instruments

Reliability Value of Critical Thinking Instruments	Reliability Value of Creativity Instruments
0.74	0.79
High degree of reliability	High degree of reliability

3. 4 Difficulty Level of Question Items

The level of difficulty of the critical thinking and creativity instrument items in the Rasch model analysis is seen based on the threshold value. The difficulty level of critical thinking questions in Table 11 shows that 5 questions are in the medium category, namely numbers 2, 3, 8, 9, and 10. Meanwhile, 2 questions are in the difficult category, namely number 1 and number 7. Based on Table 11, the difficulty level of critical thinking questions starts from numbers 1, 7, 8, 2, 10, 3, and 9. Critical thinking indicators with the highest level of difficulty are building basic skills item number 1 and making conclusions item number 7.

Table 9. Level of Difficulty of Critical Thinking Instrument Items

Item	Threshold Value	Interpretation
1	0.50	Difficult
2	-0.52	Medium
3	-0.56	Medium
7	0.45	Difficult
8	-0.35	Medium
9	-0.59	Medium
10	-0.52	Medium

Next is the level of difficulty of creativity questions in Table 12 showing that 3 questions are in the medium category, namely items number 2, 3, and 5. Meanwhile, 4 questions are in the difficult category, namely numbers 4, 6, 7 and 8. The difficulty level of the creativity question instrument starts from numbers 4, 8, 7, 6, 5, 3 and 2. The creativity indicators with the highest level of difficulty in the questions are the indicators of flexibility (item number 4), originality (item number 6), and elaboration (item number 7 and 8).

Table 10. Creativity question instrument difficulty level

Item	Threshold Value	Interpretation
2	-0.26	Medium
3	-0.23	Medium
4	0.32	Difficult
5	-0.19	Medium
6	0.06	Difficult
7	0.09	Difficult
8	0.24	Difficult

4. Discussion

The question instrument developed in this research was used to measure high school students' critical thinking skills and creativity on water pollution material. In detail, the question instrument developed consisted of 10 questions to measure critical thinking skills and 8 essay questions to measure creativity skills. First, before testing and analysis of the Rasch model, the question instrument is validated by a biology learning expert. Biology learning experts assessed the instrument on two aspects, namely assessing the assessment and material aspects (Table 2). The results of the assessment by experts are used as a consideration for improving the instrument before testing (Figure 1 and Figure 2). According to Eliaumra et al. (2022) content validity is a drinking requirement that an instrument must have. Even though it is only a value, content validity is an important quality indicator of the validity of the instrument and provides an overview of the feasibility and practicality of the instrument. A similar opinion was expressed by Safitri et al. (2024) that the content validity of an instrument describes the extent to which the question items on the instrument represent the content and objects to be measured or the extent to which they reflect the behavioral characteristics that will be measured, in this research there are critical thinking skills and creativity. If the instrument items meet content validity, then the instrument items can be said to measure content aspects.

The results of the validity test of critical thinking and creativity items based on the INFIT MNSQ and OUTFIT T values, Table 8 shows that 3 items on critical thinking are not misfits and in Table 9 1 item on creativity is a misfit. According to Muntazhimah et al. (2020) that question items that do not meet the validity criteria can be categorized as invalid items (misfit) and cannot be maintained, but must be discarded or replaced. Erfan et al. (2020) and Petra and Aziz (2020) convey that an instrument must be able to measure what it should measure. Apart from that, the question instruments used as student evaluation materials must be valid and reliable (Lia et al., 2020).

The next analysis is the reliability of the instrument. A measurement result can be trusted if several times carrying out measurements on the same group of subjects, relatively similar measurement results are obtained, as long as the aspect being measured in the subject has not changed (Farida & Musyarofah, 2021). The results of the reliability analysis of the critical thinking skills instrument obtained a value of 0.74 and the reliability of the creativity skills instrument obtained a value of 0.79. The results of this study are similar to opinions Ubadillah et al. (2022) that good instrument reliability is reliability above 0.60. Research results by Ndiung and Jediut (2020) that the high reliability of the items indicates that this instrument is quite adequate and can actually be used. The validity and reliability of each item in the questionnaire is important, because it relates to the accuracy and reliability of the instrument. Viewed from a reliability perspective, the two critical thinking and creativity instruments prepared are categorized as good or reliable.

The results of measuring the level of difficulty of the items on the critical thinking and creativity questions instrument show that most of the questions have a medium level of difficulty and some questions have a high level of difficulty (Table 11 and Table 12).

According to Tauhidah and Rofi'ah (2023) The level of difficulty of the questions must vary, items that do not vary will make it difficult to accurately classify students' abilities. The higher the difficulty level value, the more difficult the question item is, conversely, if the difficulty level value is lower, the question item is judged to be easier (Saputri et al., 2023). The level of difficulty of a question item can be influenced by the student's ability to take the test. Therefore, educators or question makers must consider the level of difficulty of the questions so that students' abilities are better trained and not stuck with questions that are low-level or easy so that with poor-quality questions, students will find it difficult to develop their ability to think.

The results of this research produced an instrument for critical and creative thinking in biology subjects, especially in water pollution. The research results show that the instrument is valid and reliable. Critical and creative thinking items that did not fit the Rasch model were removed in this study. Therefore, teachers can use this instrument as an assessment tool as well as familiarize students with questions oriented toward critical and creative thinking skills.

5. Conclusions

Instrument validity and reliability are absolute requirements for a suitable instrument. A suitable instrument is an instrument that is capable of measuring the object being measured and is consistent. Apart from the validity and reliability of the instruments used as evaluation tools by educators, they must also have varying levels of difficulty. The results of the Rasch model analysis of the critical thinking instrument from the 10 essay questions contained 3 misfit questions and 7 fit questions with a degree of reliability of 0.74. Meanwhile, the Rasch model analysis of the creativity instrument from the 8 essay questions contained 1 misfit question and 7 fit questions with a degree of reliability of 0.79. Question items that are not suitable for this research are not used for the next stage, however, there are still indicators and sub-materials that represent these question items. Overall, the difficulty level of the questions is in the medium category and some questions are in the high category. Empowering students' critical thinking skills and creativity is a big task for all stakeholders in the education sector, one of which is teachers. Efforts that can be made are to familiarize students with questions oriented towards critical thinking skills and creativity so that the research results can be used by future researchers to measure students' critical thinking skills and creativity or can be used by teachers.

Author Contributions: S.T.N.I. : writing original draft, writing original instruments, methodology, collecting and analysis data; and R.A : reviewing instruments, reviewing draft and finalizing of publication manuscript

Acknowledgments: Thank you to SMAN 1 Karas, Magetan Regency for giving permission to carry out research and the Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta

Conflicts of Interest: Authors declare there are no conflicts of interest.

6. References

Amin, S., Utaya, S., Bachri, S., Sumarmi, & Susilo, S. (2020). Effect of problem-based learning on critical thinking skills and environmental attitude. *Journal for the Education of Gifted Young Scientists*, *8*(2), 743–755. https://doi.org/10.17478/jegys.650344

Astuti, R. T., Dewi, M., & Pratiwi, R. Y. (2023). Pedagogik online test instrument to measure students' critical thinking ability. *Jurnal Riset Pedagogik*, *7*(1), 276–286. https://doi.org/10.20961/jdc.v7i1.71328

Blegur, J., Rajagukguk, C. P. M., Sjioen, A. E., & Souisa, M. (2023). Innovation of analytical thinking skills instrument for throwing and catching game activities for

elementary school students. *International Journal of Instruction*, *16*(1), 723–740. https://doi.org/10.29333/iji.2023.16140a

Bui, T. L. T., Kazarenkov, V. I., & de Tran, V. (2020). Application of Rasch model to develop a questionnaire for evaluating the quality of teaching for students' creativity development. *International Journal of Learning, Teaching and Educational Research*, *19*(8), 278–296. https://doi.org/10.26803/ijlter.19.8.15

Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of Rasch model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 329–345. https://doi.org/10.24815/jpsi.v9i3.19618

Eliaumra, E., Samaela, D. P., & Muhdin, N. K. (2022). Developing diagnostic test assessment to measure creative thinking skills of Biology preservice teacher students. *Research and Evaluation in Education*, *8*(2), 152–168. https://doi.org/10.21831/reid.v8i2.50885

Ennis, R. H. (2011). *The nature of critical thinking: an outline of critical thinking dispositions and abilities*. https://www.yumpu.com/en/document/view/50722770/the-nature-of-critical-thinking-an-outline-of-critical-our-faculty

Erfan, M., Maulyda, M. A., Ermiana, I., Hidayati, V. R., & Widodo, A. (2020). Validity and reliability of cognitive tests study and development of elementary curriculum using Rasch model. *Psychology, Evaluation, and Technology in Educational Research*, *3*(1), 26–33. https://doi.org/10.33292/petier.v3i1.51

Farida, F. & Musyarofah, A. (2021). Validitas dan reliabilitas dalam analisis butir soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab* , *I*(1), 34–44. https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB

Fietri, W. A., Zulyusri, & Violita. (2021). Analisis butir soal biologi kelas xi madrasah aliyah sakinah kerinci menggunakan program komputer anates 4.0 for windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, *7*(1), 28–35. https://doi.org/10.15548/nsc.v7i1.2329

Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items quality analysis using Rasch model to measure elementary school students' critical thinking skill on stem learning. *Jurnal Pendidikan Indonesia*, *9*(1), 61. https://doi.org/10.23887/jpi-undiksha.v9i1.20884

Kardoyo, K., Nurkhin, A., Muhsin, M., & Pramusinto, H. (2020). Problem-based learning strategy: Its impact on students' critical and creative thinking skills. *European Journal of Educational Research*, *9*(3), 1141–1150. https://doi.org/10.12973/EU-JER.9.3.1141

Leasa, M., Batlolona, J. R., & Talakua, M. (2021). Elementary students' creative thinking skills in science in the Maluku islands, Indonesia. *Creativity Studies*, *14*(1), 74–89. https://doi.org/10.3846/cs.2021.11244

Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model. *Research and Evaluation in Education*, *6*(1), 41–50. https://doi.org/10.21831/reid.v6i1.30112

Lidinillah, D. A. M., Aprilia, M., Suryana, D., & Ahmad, A. B. (2020). Development of creativity instrument through Rasch model analysis. *Universal Journal of Educational Research*, *8*(4), 1620–1627. https://doi.org/10.13189/ujer.2020.080455

Meryastiti, V., Supeno, S., Ridlo, Z. R., & Rahayuningsih, R. (2023). Improving critical thinking skills of junior high school students in science learning using the development of interactive e-module based macromedia flash. *Journal of Innovative Science Education*, *12*(2), 163–172. https://doi.org/10.15294/JISE.V12I2.55080

Muntazhimah, M., Putri, S., & Khusna, H. (2020). Rasch model untuk memvalidasi instrumen resiliensi matematis mahasiswa calon guru matematika. *Jurnal Kajian Pendidikan Matematika*, *6*(1), 65–74. https://doi.org/10.30998/jkpm.v6i1.8144

Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, *10*(1), 94. https://doi.org/10.25273/pe.v10i1.6274

Pantiwati, Y., Kusniarti, T., Permana, F. H., Nurrohman, E., & Sari, T. N. I. (2022). The effects of the blended project-based literacy that integrates school literacy movement strengthening character education learning model on metacognitive skills, critical thinking, and opinion expression. *European Journal of Educational Research*, *12*(1), 145–158. https://doi.org/10.12973/eu-jer.12.1.145

Petra, T. Z. H. T., & Aziz, M. J. A. (2020). Investigating reliability and validity of student performance assessment in Higher Education using Rasch Model. *Journal of Physics: Conference Series*, *1529*(4). https://doi.org/10.1088/1742-6596/1529/4/042088

Prasetya, W. A., & Pratama, A. T. (2023). Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(1), 76–91. https://doi.org/10.21831/pep.v27i1.58873

Priyanto, D., & Dharin, A. (2021). Students creativity development model and its implementation in Indonesian islamic elementary school. *Pegem Journal of Education and Instruction, 11*(3), 81–87. https://doi.org/10.14527/pegegog.2021.00

Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar : Jurnal Pendidikan Matematika*, *12*(2), 399–412. https://doi.org/10.24042/ajpm.v12i2.9939

Sa'diyah, L. H., Siahaan, P., Suhendi, E., Samsudin, A., Aminudin, A. H., Rais, A., Sari, I., & Rachmadtullah, R. (2020). Critical thinking instrument test (ctit): Developing and analyzing sundanese students' critical thinking skills on physics concepts using Rasch analysis. *International Journal of Psychosocial Rehabilitation*, *24*, 2020. https://doi.org/10.37200/IJPR/V24I8/PR281423

Safitri, I., Lestarani, D., Imtikhanah, R. D. N. W., Akbarini, N. R., Sari, M. W., Fitrah, M., Ilyas, Sista, R. T., Setyono, I. D., Setyaningtyas, R. F., Ndayizeye, O., & Hapsan, A. (2024). *Teori pengukuran dan evaluasi* (A. Hapsan, Ed.; 1st ed.). CV. Ruang Tentor. https://books.google.co.id/books?id=NWf_EAAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis instrumen assesmen : validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, *09*(05), 2986–2995. https://doi.org/10.36989/didaktik.v9i5.2268

Sari, E. D. K., & Mahmudi, I. (2024). *Analisis pemodelan Rasch pada assessment pendidikan (analisis dengan menggunakan aplikasi winstep)* (1st ed.). PT. Pena Persada Kerta Utama. https://www.researchgate.net/publication/378210331_BUKU_ANALISIS_PEMODELAN_RASCH_PADA_ASESMEN_PENDIDIKAN

Sari, R. R., Sugiharto, B., & Widoretno, S. (2022). Instrument analysis of biology teachers' needs to assess students' creative thinking skills using Rasch model. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *7*(2), 241–258. https://doi.org/10.24042/tadris.v7i2.13265

Satria, E., Sarumaha, Y. A., Satria, E., Hs, N., & Anas. (2023). Building students' critical thinking skill through problem-based learning model. *Jurnal Kajian Pendidikan FKIP Universitas Dwijendra*, *14*(1). https://doi.org/10.46650/wa.14.1.1409.92-98

Setyowarno, D. (2017). *Upaya meningkatkan kualitas butir soal dengan analisis aplikasi quest*. https://staffnew.uny.ac.id/upload/198810132015041004/lainlain/PPM%20Panduan%20Quest.pdf

Siahaan, E. Y. S., Muhammad, I., Dasari, D., & Maharani, S. (2023). Research on critical thinking of pre-service mathematics education teachers in Indonesia (2015-2023): A bibliometric review. *Jurnal Math Educator Nusantara: Wahana Publikasi Karya Tulis Ilmiah Di Bidang Pendidikan Matematika*, *9*(1), 34–50. https://doi.org/10.29407/jmen.v9i1.19734

Simanjuntak, M. P., Hutahaean, J., Marpaung, N., & Ramadhani, D. (2021). Effectiveness of problem-based learning combined with computer simulation on students' problem-solving and creative thinking skills. *International Journal of Instruction*, *14*(3), 519–534. https://doi.org/10.29333/iji.2021.14330a

Sudihartinih, E., & Prabawanto, S. (2020). Test instrument validation in plane geometry using Rasch model. *Mathematics Education Journals*, *4*(2), 102–115. http://ejournal.umm.ac.id/index.php/MEJ

Sumardi, S. (2020). *Teknik pengukuran dan penilaian hasil belajar [Measurement techniques and assessment of learning outcomes]*. Deepublish.

Sumarni, W., & Kadarwati, S. (2020). Ethno-stem project-based learning: Its impact to critical and creative thinking skills. *Jurnal Pendidikan IPA Indonesia*, *9*(1), 11–21. https://doi.org/10.15294/jpii.v9i1.21754

Susdelina, S., Perdana, S. A., & Febrian. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan Rasch model. *Jurnal Kiprah*, *VI*(1), 41–48. https://doi.org/10.31629/kiprah.v6i1.574

Tania, T., Sajidan, & Harlita. (2021). Analyzing the quality of instrument for critical thinking skill and assessing students' critical thinking skill in ecology using the Rasch model. *Proceeding Biology Education Conference*, *17*(1), 120–126.

Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Analisis butir soal pilihan ganda tema sehat itu penting kelas v sd negeri 04 kota bengkulu. *Elementary School Education Journal*, *4*(2), 124. https://doi.org/10.30651/else.v4i2.7090

Tauhidah, D., & Rofi'ah, N. L. (2023). Validation of undergraduate science process skills tests: Rasch model analysis. *Jurnal Research and Development in Education*, *3*(1), 51–57. https://doi.org/10.22219/raden.v3i1

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, *11*(3). https://doi.org/10.3390/jintelligence11030054

Torrance, E. P. (1969). *Creativity*. Dimensions Publishing Company.

Ubadillah, M., Marwoto, P., Wiyanto, Rusilowati, A., Subali, B., Mindyarto, B. N., & Isnaeni, W. (2022). Development of habits of mind instruments in the context 23 of basic physics practicum: Efa and Rasch mode. *Journal of Educational Cultural and Psychological Studies*, *26*(12), 23–49. https://www.ledonline.it/ECPS-Journal/

Yasin, S. N. T. M., Yunus, M. F. M., & Ismail, I. (2018). The use of Rasch measurement model for the validity and reliability. *Journal of Counseling and Educational Technology*, *1*(2), 22. https://doi.org/10.32698/0111

Zainil, M., Kenedi, A. K., Rahmatina, Indrawati, T., & Handrianto, C. (2023). The influence of a STEM-based digital classroom learning model and high-order thinking skills on the 21st-century skills of elementary school students in Indonesia. *Journal of Education and E-Learning Research*, *10*(1), 29–35. https://doi.org/10.20448/jeelr.v10i1.4336