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

July 2023. Vol. 11, No. 3
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 706-718

The Development of Critical Thinking Skills Assessment Instrument Based on Nearpod in Junior High School Science Learning

Rohmatul Maghfiroh, *Sri Wahyuni, Zainur Rasyid Ridlo
Science Education Study Program, Department of Mathematics and Natural Sciences Education, Faculty of Teacher Training and Education, Jember University. Jl. Kalimantan No.37, Jember 68121, Indonesia.

*Corresponding Author e-mail: sriwahyuni.fkip@unej.ac.id

Received: May 2023; Revised: June 2023; Published: July 2023

Abstract

Critical thinking skills are one of the skills needed in the 21st century, where to be able to assess critical thinking skills, assessment instruments are required. The aim of the research is development critical thinking skills instrument based on nearpod in junior high school science learning that meets valid, reliable, difficult, and practical criteria as well as critical thinking skills analysis. This research is a Research and Development research (R&D) with the GLAI model. The validity of the product is assessed by three validators consisting of one lecturer and two science teachers. Testing of assessment instruments was carried out on 125 grade VII students. The average product validity assessment of the three validators was 81.70% with valid criteria. Based on the field tests conducted, it is known that 10 items are valid with a reliability score of 0.602. At the difficulty level, the product has 1 easy item, 8 sufficient items, and 1 difficult item. In practicality, 97.66% was obtained with very practical criteria. In the analysis of critical thinking skills, the results of interpretation were obtained by 28.68% (very low), analysis by 31.20% (very low), inference by 43.44% (very low), and explanation by 47.80 % (low). So it is known that the critical thinking skills instrument based on nearpod in junior high school science learning was developed was valid, reliable, difficult levels, practical, and effective in measuring critical thinking skills.

Keywords: Development, assessment instrument, critical thinking skills, nearpod, junior high school, science learning

How to Cite: Maghfiroh, R., Wahyuni, S., & Ridlo, Z. (2023). The Development of Critical Thinking Skills Assessment Instrument Based on Nearpod in Junior High School Science Learning. *Prisma Sains : Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 11*(3), 706-718. doi:https://doi.org/10.33394/j-ps.v11i3.7876

https://doi.org/10.33394/j-ps.v11i3.7876

INTRODUCTION

Critical thinking skills come from three words, namely skill, thinking, and critical. Skill is the ability to do a job easily and meticulously (Ridlo et al., 2020). In the 21st century, students are not only expected to have a master of theoretical knowledge but students are also expected to master specific skills in the 21st century (Ekayanti et al., 2022; Zubaidah Siti, 2019). 21st-century skills include *Communication, Critical Thinking, Collaboration, and Creativity* (4C skills) (Redhana, 2019). One of the necessary 21st-century skills is critical thinking skills (Bilad et al., 2022). Thinking according to Lismaya (2019) is a cognitive process activity to absorb various forms of information obtained so that it can make a decision related to the problem at hand. Critical thinking according to (Facione, 2011) is self-regulation in making decisions that produce interpretation, analysis, evaluation, and conclusions as well as explaining a fact, concept, methodology, parameter, or contextual consideration that is used as a reference in decision making (Ridlo et al., 2020). Critical thinking skills are cognitive processes carried out by students, where students analyze the problems faced in sequence and specifically to find solutions to solve problems (Sulianto et al., 2018). Critical thinking skills

are fundamental to problem-solving. This skill is important for students to get to the root of the problem and how to get to the root of the problem so that the right solution to the problem at hand can be found. The application of critical thinking skills can be done in a variety of disciplines. Teachers play an important role in developing and developing learning programs that focus more on strengthening these skills (Ridlo et al., 2020). Therefore, teachers who act as facilitators must be good at making learning more meaningful for students (Ridlo et al., 2020).

Research conducted by (Rosmalinda et al., 2021) shows that the critical thinking skills of SMP Negeri 1 Belitang III students are still low with a percentage of 58.1%. This is seen from the ability of students who have not met the criteria for critical thinking in solving PISA-type questions. The results of research conducted by Rosmalinda et al. (2021) are supported by research conducted by Hartini and Sukardjo (2015) which states that instruments in the form of critical thinking skills are still not widely used. The results of an *interview* with one of the science teachers of SMPN 4 Jember obtained information that also stated that SMPN 4 Jember had never been given questions to measure critical thinking skills. The low implementation of critical thinking skills assessment will have an impact on graduates having low critical thinking skills both at the elementary school to tertiary levels (Reta, 2012). According to Sani (2016), assessment is an effort carried out systematically to collect valid and reliable data or information, where assessment is used as a benchmark for making policies related to an education program. The assessment in the independent curriculum on learning outcomes consists of three domains, namely cognitive, psychomotor, and affective. The assessment standards contained in Permendikbud no. 21 of 2022 article 9 state that the assessment of student learning outcomes consists of two forms, namely formative and summative. Formative assessment aims to monitor and improve the learning process and evaluate the achievement of learning objectives. With formative assessments, information will also be obtained about learning difficulties and student development. Meanwhile, summative assessment is carried out to assess students as a reference to determine class advancement and graduation from educational units.

Critical thinking skills include cognitive domains that can be assessed using formative tests. Formative tests to measure critical thinking skills can be multiple-choice objective tests. In addition to multiple-choice questions, the cognitive realm can be assessed using essay questions (Martanti et al., 2021) However, the fact is that the assessments that have been carried out by teachers are still mostly oriented toward cognitive assessments in the form of multiple choice objective tests that are oriented to knowledge only and have not referred to critical thinking skills (Nyoman Putriadi & dkk, 2020). This is reinforced by (Putra et al., 2021) who stated that the question items given by teachers are still not classified as question items that can develop critical thinking skills. According to (Jember et al., 2018), in addition to using methods and models that can familiarize critical thinking skills, teachers must also provide questions that familiarize students to think critically. The solution offered to overcome the problems mentioned is to develop an assessment instrument for critical thinking skills. The development of critical thinking skills assessment instruments is needed because, with critical thinking skills assessment instruments, teachers can find out the level of critical thinking of students which can be used as a reference for the preparation of learning plans to be implemented (Setiana, 2018) Assessment instruments that are well arranged and based on the level of student thinking ability will improve students' thinking power, especially critical thinking skills (Fika Amalia & Endang Susilaningsih, 2014).

The assessment using paper that has been carried out so far is less effective because it requires a lot of costs and also a long time (Nugroho & Airlanda, 2020). In addition to requiring a lot of money and also a long time, paper assessments are also not environmentally friendly. Test sheets and answer sheets will accumulate more and more and pollute the environment. The solution to replace paper-based assessments is to use computerized *assessments online.* One platform that can be used to conduct computerized assessments is nearpod. The difference

between this study and previous research is that this study uses the nearpod platform and the questions developed contain subskills, indicators of critical thinking skills in materials substances, and changes. Substance material and its changes are one of the class VII science materials in the independent curriculum. In previous studies, research has been carried out related to the development of science literacy-based assessment instruments using Quizizz to measure hots by Azizah et al. (2023) on circulatory system material in humans. Then the research that has been carried out by Hamidah and Wulandari (2021) is the development of hots-based assessment instruments using the Quizizz application on fluid materials and harmonic vibrations.

Nearpod is a website-based application, and to access it students do not need to install applications on their devices. This Nearpod has several advantages including that it can be operated both on laptops and devices (Nurhamidah, 2021). Besides being easy to operate both on laptops and devices, student answers can also be written in pdf form which will make it easier for teachers to correct essay answers. *Nearpod* is one such software that supports learning especially assessment. In the nearpod, there are activity features that can be used for assessment, such as interactive quizzes, multiple-choice tests, memory tests, tests with long answers, tests with short answers, and students can answer on the nearpod with pictures (Minalti & Erita, 2021). This study aims to develop critical thinking skills instruments based on nearpod in junior high school science learning that meet valid, reliable, difficult, and practical criteria as well as critical thinking skills analysis.

METHOD

This research is included in the type of *Research and Development* (R&D) development research that uses the GLAI development model. GLAI is a design used to develop assessment instruments proposed by Bowling et al. (2008) which consists of six stages, namely defining content, developing and selecting test items, review by experts, focus group interviews, pilot study data collection, and evaluation. The more detailed steps of the GLAI development model, it is presented in the following Figure 1.

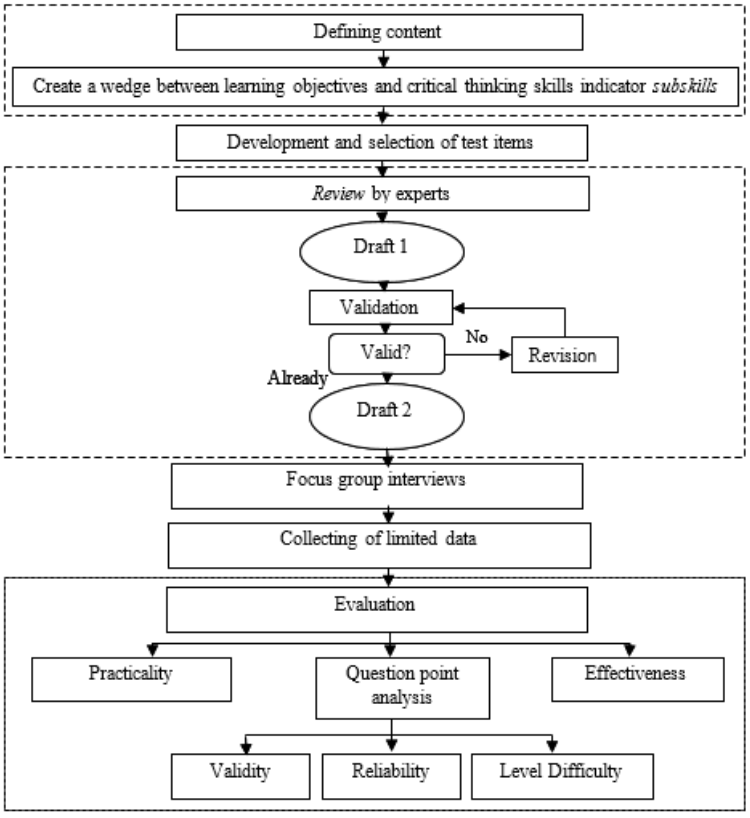

Figure 1. Research Design

The research subjects in this study were grade VII students of SMPN 4 Jember with details on a limited trial consisting of 30 students and a field test consisting of 125 students. Data obtained in this study were obtained from interviews, observations, and expert validation as well as analysis from limited trials and field tests. Meanwhile, data analysis techniques use expert validation analysis, question item analysis, assessment implementation analysis, and critical thinking skills analysis. The analysis of question items consists of analyzing the validity of the question items, reliability, and level of difficulty.

In this research, the data analysis techniques used include expert validity, empirical validity, practicality tests, and effectiveness tests of critical thinking skills assessment instruments based on nearpod. The data analysis techniques used are described as follows.

Expert validity

To find out the magnitude of the value of validity, experts can use the following formula. Information:

$$V_a = \frac{T_{SE}}{T_{SH}}$$

V_a = Expert validation
T_{SE} = Total score achieved
T_{SM} = Total expected score

Furthermore, the percentage results of each validator are averaged and criterion based on the following Table 1.

Table 1. Criteria validity instrument

Validity Criteria (%)	Validity Categories
$90.01 < x \leq 100.00$	Very valid
$70.01 < x \leq 90.00$	Valid
$50.01 < x \leq 70.00$	Less valid
$25.00 < x \leq 50.00$	Not valid

(Wahyuni* et al., 2022)

Empirical validity

Empirical validity is known when the developed question items have been used to make measurements. Empirical validity in research includes analysis of question items ranging from question item validity tests to difficulty tests.

a. Validity of question items

The validity of the problem in the form of a description is calculated using the *product moment correlation* formula proposed by Karl Pearson as follows.

$$r_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum y)^2\}\{n\sum y^2 - \sum y^2\}}}$$

where:

r_{xy} = correlation coefficient of question item score and total score
x = grain score
y = Total score
n = number of samples

Hadil calculations using the above formula are then criterion based on the following criteria in Table 2.

Table 2. Criteria for the validity of question items

r_{xy}	Question Item validity criteria
$r_{calculate} > r_{table}$	Valid
$r_{calculate} \leq r_{table}$	Not Valid

(Hamidah & Wulandari, 2021)

b. Reliability

In this study, researchers calculated the reliability of the assessment instrument developed to determine its efficacy. The assessment instrument developed is in the form of *essay* questions so that to calculate its reliability can be calculated using the *Alpha* formula as follows.

$$r_i = (\frac{k}{k-1})(1 - \frac{\sum \sigma_b^2}{\sigma_t^2})$$

Remarks :

r_i = Instrument Reliability

k = number of question items

$\sum \sigma_b^2$ = Number of grain variances

σ_t^2 = total variance

Based on Guiltford's reference, reliability criteria are categorized as follows in Table 3.

Table 3. Question item reliability criteria

Reliability Coefficient	Category
0.80-1.00	Very High
0.60-0.79	High
0.40-0.59	Sufficient
0.20-0.39	Low
0.00-0.19	Very low

(Hamidah & Wulandari, 2021)

c. Difficulty level

The difficulty level of the question item is shown in the form of a number called the question item difficulty index. To determine the level of difficulty, subjective test items in the form of descriptions or *essays* are as follows.

$$P = \frac{\bar{X}}{X_{max}}$$

where:

P = The difficulty level of the question item

$\bar{X}$ = Average score

X_{max} = Maximum score on a specific question item

The value obtained from the calculation can be interpreted with the following criteria in Table 4.

Table 4. Criteria for the difficulty level of the question item

Difficulty Index	Criterion
0.00-0.29	Difficult
0.30-0.70	Sufficient
0.71-1.00	Easy

(Wahyuni* et al., 2022)

Practicality Test

The research conducted in its implementation was assessed by three observers to know the practicality of the implementation of the assessment. To find out the percentage of implementation of the assessment is calculated using the following formula.

$$(p) = \frac{\Sigma x}{n} \times 100\%$$

Information:
p = implementation of assessment
Σx = Total number of scores obtained
n = Total overall score

The score results obtained from the calculation are then criterion based on the following criteria.

Table 5. Practicality criteria

Implement ability (%)	Category
$k \geq 90$	Very practical
$80 \leq k < 90$	Practical
$70 \leq k < 80$	Quite practical
$60 \leq k < 70$	Less practical
$k < 60$	Very impractical

(Wahyuni* et al., 2022)

Effectiveness Test

The effectiveness of the critical thinking skills assessment instrument based on nearpod can be seen from the result of the analysis of critical thinking skills. The percentage of students' critical thinking skills can be known from the acquisition of calculations using the following formula.

$$Percentage = \frac{Total\ score}{Maximum\ Score} \times 100\%$$

The percentage acquisition of critical thinking skills is further categorized based on the following criteria in Table 6.

Table 6. Criteria for Critical Thinking Skills

Interpretation (%)	Category
$81.25 < X \leq 100$	Very high
$71.5 < X \leq 81.25$	High
$62.5 < X \leq 71.5$	Sufficient
$43.75 < X \leq 62.5$	Low
$0 < X \leq 43.75$	Very low

(Wahyuni* et al., 2022)

RESULTS AND DISCUSSION

Development research that has been carried out produces a product in the form of a critical thinking skills assessment instrument based on nearpod. The resulting critical thinking skills assessment instrument is in the form of 10 essays with indicators of critical thinking skills. The development stage of this assessment instrument uses the GLAI development model which consists of 6 stages, namely defining content, developing and selecting test items, review by experts, focus group interviews, pilot study data collection, and evaluation.

Defining content

The initial stage carried out before the content definition stage is to analyze problems and needs to be related to assessment. Problem analysis is carried out by conducting literature study activities related to assessment instruments in previous studies. Then the needs analysis was carried out by observation and also an interview with one of the science teachers at SMPN 4 Jember. From the interviews that have been conducted, information was obtained that SMPN 4 Jember has never been given questions to find out students' critical thinking skills, and also the assessment carried out previously only used Google form when online and also used paper when offline.

At the stage of defining content, researchers identify concepts to be tested using instruments. Researchers at this stage create a wedge between learning objectives and *subskills* indicators of critical thinking skills. After obtaining the slice, researchers use it as a reference in developing critical thinking skills assessment instruments.

Development and selection of test items

At this stage, researchers create questions based on slices of learning objectives and *subskills* indicators of critical thinking skills that have been created in the previous stage. The development and selection of test items consist of determining the form of questions adjusted to the slices that have been made, determining the number of questions, and also scoring questions, making question cards as well as inputting questions on the nearpod. The results of the development that have been made can be seen in Figure 2.

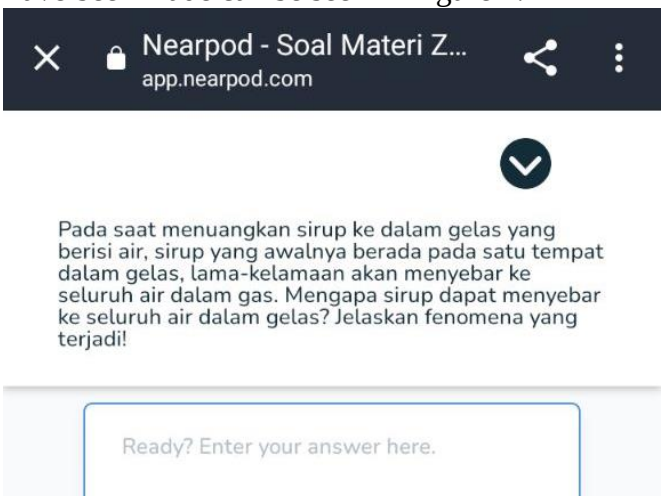

Figure 2. Result of the development assessment instrument

Expert reviews

At this stage, researchers validate experts to find out the validity of the assessment instruments that have been made. In this validation, experts provide assessments, comments, and also suggestions on assessment instruments that have been developed. In the validation sheet, there is also a decision from experts regarding the validity of the instrument whether it needs to be revised, or whether the assessment instrument can be directly tested on a limited basis. The results of expert validation are presented in the Table 7.

Table 7. Expert validation

No	Assessed aspects	Validator 1	Validator 2	Validator 3	Average
1	Material	84.50%	81%	80%	81.83%
2	Construction	84.50%	79%	82%	81.83%
3	Language	84%	80.40%	80%	81.46%
Average percentage					81.70%
Score criteria					Valid

Based on the analysis that has been carried out on the validation results, it can be seen that the critical thinking skills assessment instrument based on nearpod in junior high school science learning is included in the valid score category. Even though it has been classified as a valid score, there still needs to be an improvement from comments and suggestions from validators. Comments and suggestions from validators in detail can be seen in Table 8.

Table 8. Validator comments and suggestions

Validator	Validator comments and suggestions
Validator 1	The questions should represent indicators of critical thinking skills
Validator 2	More specific question description
Validator 3	-

In the *review* stage of experts, each validator provides a decision regarding the feasibility of the instrument developed by the researcher. The validator's decision regarding the feasibility of the instrument is that the three validators stated that the instrument developed by the researcher was suitable for research with minor revisions.

Focus group interviews

The focus group interview phase was conducted after conducting a limited trial. For the limited trial, researchers used one class and at this stage, researchers conducted interviews with 5 students. From the interviews that have been conducted, information was obtained that the sentences used in the questions can be understood by students.

Pilot study data collection

In the collection phase of the pilot study, researchers conducted a limited trial on 30 students of grade VII D SMPN 4 Jember. Data obtained from limited trials are then collected for analysis ranging from validity, reliability, level of difficulty, practicality of instruments, and analysis of critical thinking skills. The acquisition of validity tests in limited trials after being processed using the SPSS application can be seen in Table 9.

Table 9. Limited trial validity results

Question Number	r calculate	r table	Information
1	0.464	0.361	Valid
2	0.548	0.361	Valid
3	0.622	0.361	Valid
4	0.576	0.361	Valid
5	0.434	0.361	Valid
6	0.465	0.361	Valid
7	0.457	0.361	Valid
8	0.534	0.361	Valid
9	0.427	0.361	Valid
10	0.461	0.361	Valid

Determination of the validity of question items can be seen from the correlation between the r table and also r count. If the r table is greater than the r table then the question item is declared invalid and if the r table is smaller than the r count then the question item is declared valid. In a limited test conducted by researchers with n = 30, where n is obtained from the number of students who take part in the limited trial. From the research that has been done by researchers, it is known that questions number 1, 2, 3, 4, 5, 6,7, 8, 9, and 10 are declared valid. In the next stage, the questions that are declared valid can be used for field tests.

The second question point analysis is the reliability analysis of assessment instruments. Reliability according to the reliability obtained by researchers from limited trials conducted is 0.641. The reliability value of the limited trial was then criterion based on Sumardi's (2020) criteria, namely in the range of 0.60-0.79 included in the high-reliability criteria. Based on limited test results, it is known that the assessment instruments developed by researchers have high reliability. So that if the assessment instrument is used, the same results will be obtained, even though the subject and time are different.

The next analysis is the analysis of the level of difficulty of each question. The results of the calculation of the difficulty level in limited trials are presented in the following Table 10.

Table 10. Limited trial difficulty

Question Number	Difficulty Level	Information
1	0.60	Sufficient
2	0.41	Sufficient
3	0.67	Sufficient
4	0.47	Sufficient
5	0.27	Difficult
6	0.49	Sufficient
7	0.41	Sufficient
8	0.56	Sufficient
9	0.35	Sufficient
10	0.23	Difficult

The results of obtaining the level of difficulty are interpreted based on Kurniawan's criteria (2021). From the interpretation carried out by the researcher, it is known that questions number 1, 2, 3, 4, 6, 7, 8, and 9 are classified as questions that have a moderate level of difficulty because in the difficult test that has been carried out, the difficulty level results on these numbers in the range of 0.3 < the difficulty level < 0.7. As for numbers 5 and 10, it is quite difficult because the test of the level of success on these numbers is located in the range of difficulty levels < 0.3.

The limited trial in its implementation was observed by three observers, where each observer assessed the implementation sheet on a scale of 1-5. The results of the assessment conducted by the three observers are presented in the following Table 11.

Table 11. The practicality of limited trials

Activities	Meeting (%)			Average
	1	2	3	(%)
Preliminary activities	100	93.33	93.33	95
Receive a question *link*	100	100	100	100
Do the problem	100	100	100	100
Collect answers	93.33	93.33	93,33	93.33
Concluding activities	100	100	100	100
Overall average	98.66	97.33	97.33	97.66
Category	Very practical			

Furthermore, the results of obtaining assessment data by observers were criterion according to (Wahyuni* et al., 2022) criteria, namely in small group tests, an implementation score of 97.66% was obtained, so if the implementation score is the criterion, it is classified as very practical. Based on small group tests that have been conducted, it can be seen that critical thinking skills assessment instruments based on nearpod can be used in assessment activities because of their very practical use. Researchers at the pilot study stage analyzed students' critical thinking skills. Analysis of critical thinking skills is detailed on each critical thinking indicator. The results of obtaining critical thinking skills analysis are presented in the following Table 12.

Table 12. Analysis of students' critical thinking skills on limited trials

Indicators of critical thinking skills	Percentage	Criterion
Interpretation	23.50	Very Low
Analysis	26.67	Very Low

Indicators of critical thinking skills	Percentage	Criterion
Inference	37.67	Very Low
Explanation	54.53	Low

The indicators of critical thinking skills used by researchers to create assessment instruments are critical thinking indicators proposed by Facione (2020) including interpretation, analysis, *inference*, and *explanation*. However, not all indicators from Facione are used in making question items. This is because based on the slices that have been made, researchers at the stage of defining content for evaluation and *self-regulation* indicators do not enter the wedge between the *subskills* of critical thinking indicators and the learning objectives of the substance material and its changes. So that for the analysis of critical thinking skills, indicators of *evaluation* and *self-regulation* cannot be raised. Then the results of the analysis in the table above are criteria based on the criteria of Setyowati (2011), namely for *interpretation, analysis, and inference* classified as very low criteria, and for explanation classified as low criteria.

Evaluation

Researchers at the evaluation stage analyze the question items that have been developed, their practicality when used in field tests, and analysis of students' critical thinking skills. For analysis of questions, items include tests of validity, reliability, and level of difficulty. As for the results of the validity of the question items, they are presented in the following Table 13.

Table 13. Field Validity Test

Question Number	r calculate	r table	Information
1	0.632	0.176	Valid
2	0.471	0.176	Valid
3	0.471	0.176	Valid
4	0.402	0.176	Valid
5	0.393	0.176	Valid
6	0.533	0.176	Valid
7	0.307	0.176	Valid
8	0.265	0.176	Valid
9	0.712	0.176	Valid
10	0.365	0.176	Valid

The validity of the field test can be known from the correlation between the r table and the r count. If the r table is greater than the r count then the question item is declared invalid and if the r count is greater than the r table then the question item is declared valid. To be able to interpret the question items as valid or not can be seen in Table 13. From table 13 it is known that questions number 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are declared valid because r is calculated on each question item more than r in table 0.176. To find out r table from this study can be seen from n. n is the number of students who took the field test. The results of the difficulty level in the field test, it is presented in the following Table 14.

Table 14. Field Test Difficulty Level

Question Number	Difficulty Level	Information
1	0,57	Sufficient
2	0,35	Sufficient
3	0,85	Easy
4	0,29	Sufficient
5	0,31	Sufficient
6	0,34	Sufficient
7	0,47	Sufficient

Question Number	Difficulty Level	Information
8	0,29	Difficult
9	0,40	Sufficient
10	0,29	Sufficient

Based on field tests that have been conducted by researchers, it is known that questions number 1, 2, 4, 6, 7, 9, and 10 are classified as questions that have a moderate level of difficulty because the difficult test that has been carried out the results of the difficulty level on these numbers in the range of 0.3 < the difficulty level <0.7 and for question number 3 has an easy level of difficulty because in the difficulty level test that has been carried out the difficulty level on the number is more than 0.7. As for number 8, it is classified as difficult because the difficulty level test on these numbers is located in the range of difficulty levels less than 0.3.

Field tests in its implementation were observed by three observers, where each observer assessed the implementation sheet on a scale of 1-5. The results of the assessment by the three observers on the implementation sheet are presented in the following Table 15.

Table 15. The practicality of field tests

Activities	Observer (%)			Average
	1	2	3	(%)
Preliminary activities	100	93.33	93.33	93.33
Receive a question *link*	100	100	100	100
Do the problem	100	100	100	100
Collect answers	93.33	93.33	93.33	93.33
Concluding activities	100	100	100	100
Overall average	98.66	97.33	97.33	97.33
Category	Very practical			

Furthermore, the results of obtaining assessment data by observers were criterion according to (Wahyuni* et al., 2022) criteria, namely in field tests, an implementation score of 97.33% was obtained, so if the implementation score is the criterion, it is classified as very practical. Based on small group tests that have been conducted, it can be seen critical thinking skills assessment instruments based on nearpod can be used in assessment activities because of their very practical use.

Data on student scores in field tests are then analyzed based on indicators of critical thinking skills. The results of the analysis of students' critical thinking skills are presented in the following Table 16.

Table 16. Analysis of Students' critical thinking skills on field tests

Indicators of critical thinking skills	Percentage	Criterion
Interpretation	28.68	Very Low
Analysis	31.20	Very Low
Inference	43.44	Very Low
Explanation	47.80	Low

From the acquisition of data on students' critical thinking skills test results after analysis, it is known that indicators of critical thinking skills have different magnitudes, namely for *interpretation* by 28.68%, *analysis* by 31.20%, *inference* by 43.44%, and *explanation* by 47.80%. Then the results of the analysis are criteria based on the criteria of Setyowati et al. (2011), namely for *interpretation, analysis, and inference* classified as very low criteria, and for explanation classified as low criteria.

CONCLUSION

The instrument developed by researchers consisting of 10 questions is classified as valid when reviewed from the validation of experts and the validity of the question items. Then the instrument developed is also classified as a reliable instrument with a level of difficulty dominated by medium questions with details in limited trials consisting of 2 difficult questions and 8 medium questions. While the field test consists of 1 difficult question, 1 sufficient question, and 9 medium questions.

RECOMMENDATION

The development of the questions pays more attention to the proportion of difficulty level and in future research it is expected to choose schools with good internet access.

ACKNOWLEDGMENT

Thank you to the bachelor of science education University of Jember and the supervisors, both the main supervisor and member supervisors who have guided the preparation of the article from start to finish so that this article can be published.

REFERENCES

Bilad, M. R., Anwar, K., & Hayati, S. (2022). Nurturing Prospective STEM Teachers' Critical Thinking Skill through Virtual Simulation-Assisted Remote Inquiry in Fourier Transform Courses. *International Journal of Essential Competencies in Education*, *1*(1), 1–10. https://doi.org/10.36312/ijece.v1i1.728

Ekayanti, B. H., Prayogi, S., & Gummah, S. (2022). Efforts to Drill the Critical Thinking Skills on Momentum and Impulse Phenomena Using Discovery Learning Model. *International Journal of Essential Competencies in Education*, *1*(2), Article 2. https://doi.org/10.36312/ijece.v1i2.1250

Facione, P. a. (2011). Critical Thinking: What It Is and Why It Counts. *Insight Assessment*, *ISBN 13: 978-1-891557-07-1.*, 1–28.

Fika Amalia, N., & Endang Susilaningsih, dan. (2014). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Siswa Sma Pada Materi Asam Basa. *Jurnal Inovasi Pendidikan Kimia*, *8*(2), 1380–1389.

Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan Instrumen Penilaian Berbasis Hots Menggunakan Aplikasi "Quizizz." *Efisiensi : Kajian Ilmu Administrasi*, *18*(1), 105–124. https://doi.org/10.21831/efisiensi.v18i1.36997

Jember, F., Lampu, P., Sdn, D. I., & Banyuwangi, S. (2018). *Copyright © anggyet al, 2018 , this is an open access article distributed under the terms of the FKIP E-Proceding license, which permits unrestricted use, distribution PENDAHULUAN and reproduction in any medium, provided the original work is properly cite*. 101–107.

Martanti, N., Malika, E. R., & Setyaningsih, A. (2021). Pengaruh Metode Pembelajaran Eksperimen Virtual Menggunakan PhET Terhadap Hasil Belajar Kognitif Siswa. *Konstelasi: Konvergensi Teknologi Dan Sistem Informasi*, 83–92.

Minalti, M. P., & Erita, Y. (2021). Penggunaan Aplikasi Nearpod untuk Bahan Ajar Pembelajaran Tematik Terpadu Tema 8 Subtema 1 Pembelajaran 3 Kelas IV Sekolah Dasar. *Journal of Basic Education*, *4*(1), 2231–2246.

Nugroho, A. N., & Airlanda, G. S. (2020). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Pembelajaran IPA Kelas 4 SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, *3*(3), 400. https://doi.org/10.23887/jippg.v3i3.29712

Nurhamidah, D. (2021). Pengembangan Instrumen Penilaian Berbasis Media Nearpod dalam Mata Kuliah Bahasa Indonesia. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, *4*(2), 80–90.

Nyoman Putriadi, D., & dkk. (2020). Pegembangan Asesmen Kinerja Pada Praktikum IPA Berbasis Pendekatan Saintifik Dalam Meningkatkan Kemampuan Berpikir Kritis Siswa Kelas VII SMP. *Oktober*, *14*(2), 1858–0629.

Putra, I. P. S., Suastra, I. W., & Suarni, N. K. (2021). Pengembangan instrumen kemampuan berpikir kritis dan kemampuan literasi sains siswa kelas IV SD. *PENDASI: Jurnal Pendidikan Dasar Indonesia*, *5*(2), 203–213.

Redhana, I. W. (2019). Mengembangkan Keterampilan Abad Ke-21 Dalam Pembelajaran Kimia. *Jurnal Inovasi Pendidikan Kimia*, *13*(1).

Reta, K. I. (2012). Pengaruh Model Pembelajaran Berbasis Masalah Terhadap Keterampilan Berpikir Kritis Ditinjau Dari Gaya Kognitif Siswa. *Pengaruh Model Pembelajaranberbasis Masalah Terhadap Keterampilan Berpikir Kritis Ditinjau Dari Gaya Kognitif Siswa*, *2*, 1–17.

Ridlo, Z. R., Dafik, & Nugroho, C. I. W. (2020). The effectiveness of implementation research-based learning model of teaching integrated with Cloud Classroom (CCR) to improving critical thinking skills in an astronomy course. *Journal of Physics: Conference Series*, *1563*(1). https://doi.org/10.1088/1742-6596/1563/1/012034

Rosmalinda, N., Syahbana, A., & Nopriyanti, T. D. (2021). Analisis Kemampuan Berpikir Kritis Siswa Smp Dalam Menyelesaikan Soal-Soal Tipe Pisa. *Transformasi : Jurnal Pendidikan Matematika Dan Matematika*, *5*(1), 483–496. https://doi.org/10.36526/tr.v5i1.1185

Setiana, D. S. (2018). Pengembangan Instrumen Tes Matematika Untuk Mengukur Kemampuan Berpikir Kritis. *Jurnal Pendidikan Surya Edukasi (JPSE)*, *4*(2), 35–48.

Sulianto, J., Cintang, N., & Azizah, M. (2018). Analisis Korelasi dan Regresi Berpikir Kritis Terhadap Kemampuan Pemecahan Masalah Matematika Siswa SD Kota Semarang. *Mengembangkan Kompetensi Pendidik Dalam Menghadapai Era Disrupsi*, *2009*, 237.

Wahyuni*, S., Ridlo, Z. R., & Rina, D. N. (2022). Pengembangan Media Pembelajaran Interaktif Berbasis Articulate Storyline Terhadap Kemampuan Berpikir Kritis Siswa SMP pada Materi Tata Surya. *Jurnal IPA & Pembelajaran IPA*, *6*(2), 99–110. https://doi.org/10.24815/jipi.v6i2.24624

Zubaidah Siti. (2019). Memberdayakan keterampilan Abad ke-21 melalui Pembelajaran Berbasis Proyek. . *Seminar Nasional Nasional Pendidikan Biologi*, *October*, 1–19.