Journal of Educational Research and Evaluation
Volume 8, Issue 2, 2024, pp. 338-349
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://doi.org/10.23887/jere.v8i2.69259

Construction of Critical Reasoning Skills Assessment Instruments as Diagnostic Assessments in Physics Learning with Polytomus Scoring

Elvin Yusliana Ekawati[1*], Sabrilla Almas Azzahra[2], Hanun Fithriyah[3]

[1,2,3] Physics Education, Sebelas Maret University, Surakarta, Indonesia

ARTICLE INFO

Article history:
Received October 19, 2023
Accepted March 10, 2024
Available online May 25, 2024

Kata Kunci :
Asesmen Diagnostik, Kemampuan Bernalar Kritis, Politomus

Keywords:
Diagnostic Assessment, Critical Reasoning Skills, Polytomus

ABSTRAK

Ketersediaan tes yang dirancang khusus untuk mengukur kemampuan awal bernalar secara kritis dalam Fisika masih terbatas, baik dari segi ketersediaan maupun inovasinya. Hal ini menunjukkan bahwa, adanya kebutuhan akan instrumen yang lebih inovatif dan efektif untuk mengukur kemampuan penalaran kritis siswa dalam Fisika.Tujuan penelitian ini yaitu untuk menghasilkan instrumen penilaian kemampuan bernalar kritis dengan penskoran politomus pada pembelajaran Fisika Fase E dalam materi Energi Alternatif sebagai asesmen diagnostik. Jenis penelitian ini merupakan penelitian campuran. Metode pengumpulan data yang digunakan adalah kuesioner dengan menggunakan instrumen berupa lembar kuesioner. Sumber data penelitian ini yaitu sebanyak 9 orang ahli dan 62 orang peserta didik dari 3 sekolah. Hasil penelitian menunjukkan bahwa, dari 12 butir instrumen asesmen diagnostik kemampuan bernalar kritis awal peserta didik dengan program Quest menunjukkan bahwa semua butir dalam kategori baik. Kesimpulan penelitian ini memberikan informasi konstruksi instrumen asesmen kemampuan penalaran kritis layak digunakan untuk mendiagnosis kemampuan penalaran kritis siswa. Instrumen asesmen diagnostik ini dapat diimplementasikan dalam pembelajaran fisika terutama pada saat kegiatan asesmen yaitu untuk mengukur kemampuan penalaran kritis siswa di awal pembelajaran sehingga guru dapat memutuskan rancangan pembelajaran yang sesuai dengan profil siswa. Terdapat beberapa implikasi dalam penelitian ini, salah satunya adalah penggunaan instrumen penilaian dengan penilaian politomus memungkinkan penilaian yang lebih rinci dan mendalam terhadap pemahaman siswa.

ABSTRACT

The availability of tests specifically designed to measure initial critical reasoning skills in Physics still needs to be improved in terms of availability and innovation. This shows a need for more innovative and effective instruments to measure students' critical reasoning skills in Physics. This research aims to produce a vital reasoning ability assessment instrument with polytomous scoring in Phase E Physics learning in Alternative Energy material as a diagnostic assessment. This type of research is mixed research. The data collection method used was a questionnaire using an instrument as a questionnaire sheet. The data sources were nine experts and 62 students from 3 schools. The results showed that the 12 items of the diagnostic assessment instrument of students' initial critical reasoning ability with the Quest program were in a suitable category. The conclusion of this study provides information on the construction of critical reasoning ability assessment instruments worth using to diagnose students' critical reasoning abilities. This diagnostic assessment instrument can be implemented in physics learning, especially during assessment activities, to measure students' essential reasoning skills at the beginning of learning so that teachers can decide on learning designs by student profiles. This study has several implications, one of which is that using assessment instruments with a polynomial assessment allows for a more in-depth and detailed evaluation of student understanding.

1. INTRODUCTION

The implementation of diagnostic assessment in educational institutions has not reached its full potential due to several challenges (Arrohman &; Lestari, 2023; Martanti et al., 2022; Nugroho et al., 2023). The implementation of diagnostic assessments still confuses educators, in terms of preparation, implementation, and follow-up. The form of delivering information or feedback has not been found in the field, educators are only limited to providing diagnostic assessments at the beginning and conducting joint evaluations. In addition, the implementation and design of diagnostic assessments are not based on good quality schools and rely more on educators' understanding of the importance of these assessments (Antika et al., 2023; Arrohman & Lestari, 2023; Firmanzah & Sudibyo, 2021). Educators need to prepare diagnostic

*Corresponding author.
E-mail addresses: elvin_fisika@staff.uns.ac.id (Elvin Yusliana Ekawati)

assessment activities carefully and do them well before learning activities so that they can match the initial abilities of students.

Diagnostic assessment is carried out before starting the learning process to identify potential, parts that need to be strengthened, understanding, skills, and learning preferences of students, to assist educators in providing efficient learning guidance to increase the potential of students according to their level of development. Providing diagnostic assessments at the beginning of learning is a good action in anticipating and overcoming potential problems such as low literacy skills, lack of numeracy skills, and limited understanding of science. Diagnostic assessment provides significant data about the situation of learners, such data can be used as a foundation for designing more efficient learning activities, which will hopefully contribute to improved learning outcomes for learners (Fan et al., 2021; Okta Susilawati et al., 2023; Puspita Rakhmi et al., 2023) The application of diagnostic assessment can be in the form of assessing and measuring students' critical reasoning skills at the beginning of learning.

Critical reasoning is an essential skill because it can help solve complex problems in the 21st century. Critical reasoning is included in the process of rational and reflective thinking to determine appropriate beliefs or actions and is part of critical thinking. The skills to reason critically include the skills to critically assess each statement, not accept information without evidence, and at the same time remain open to new methods and ideas (Febliza et al., 2023; Kahn-Horwitz & Goldstein, 2023; Suhirman & Prayogi, 2023). This skill is very important for learners because it allows them to be able to solve social, scientific, and practical problems effectively. The skills to reason critically have an important role in various aspects of life, including in the context of education, and have become very relevant in the development of critical reasoning skills. The development of critical reasoning skills in educational institutions becomes an initial milestone for learners in the learning process, which will enable them to overcome challenges in the classroom environment and ultimately contribute to problem solving in the community sphere. These skills are essential in facing the challenges that exist in the 21st century and play a fundamental role in achieving success in learning about science, one of which is in the subject of Physics (Leniati & Indarini, 2021; Prahani et al., 2023; Suharno et al., 2022; Suresman et al., 2023). Learning Physics involves active reasoning and is used to make decisions as well as evaluate complex problems in modern life.

Physical science is included in one of the disciplines that focuses on understanding how the universe works, so it takes the skills to think critically to achieve a deep understanding of the concept of Physics. One of the materials that discusses Alternative Energy requires students to use their critical reasoning skills. Alternative Energy material requires critical reasoning of students to be able to respond to the latest energy phenomena. These skills have a good impact on perfecting the learning cycle and student achievement in learning (Arini & Juliadi, 2018; Razak, 2017; Wijayanti & Siswanto, 2020). Therefore, it is necessary to assess the initial skills of critical reasoning of students to map student profiles by developing appropriate diagnostic assessment instruments.

However, the availability of tests specifically designed to measure the initial skills to reason critically in Physics is limited both in availability and innovation (Holiqi et al., 2021). The results of bibliometric analysis related to the development of diagnostic assessment instruments, and critical reasoning skills, show that the form of instruments used is limited to two-tier diagnostic tests and four-tier diagnostic tests (Dirman &; Mufit, 2022; Sunarti et al., 2021). The use of multilevel multiple-choice based instruments has several disadvantages (Rukmana, 2019). Both tests use design multiple-choice dichotomy so that it only allows right or wrong responses, which can limit respondents from coming up with answers that may be authentic and creative. This system does not allow the assessment of behavior at a deeper level of synthesis or understanding, since it only focuses on the right answer from a predetermined point of view. The system is also less able to measure respondents' skills to think more deeply or creatively about a problem. This can hinder the measurement of higher critical and analytical reasoning skills in learners. The dichotomous scoring model is also less able to distinguish mistakes made by students and does not appreciate the process of solving questions due to different error rates but is given the same score of 0 (Atılgan et al., 2020; Saepuzaman et al., 2021). Based on these shortcomings, the solution offered is through the development of diagnostic assessment instruments using a polytomous scoring system that can provide solutions to the limitations of dichotomous scoring.

Graded Response Model (GRM) is one form of indirect approach from the IRT model. This model can be used for data with polytomous responses that group in order of difficulty in each category of test items, ranging from basic to complex levels (Andayani et al., 2019; Wiyoko, 2019). For instrument items with ordinal response categories that allow a different number of response categories on each item. GRM models allow the drawing of parameters for each item to accurately estimate the skills of students (Herwin et al., 2023; van der Mark et al., 2023). The assessment in this model is in the form of a description, and each item is arranged according to its level of difficulty (Mirunnisa &; Razi, 2021). The use of this approach in instrument development can require learners to perform critical reasoning.

Development instruments of critical reasoning skills are still rare that use the development model Gable. The development of critical reasoning skills instruments mostly uses the 4D development model and the plomp development model (Hasan et al., 2020; Suarjana et al., 2020). The development model used is not appropriate because the plomp development model is intended for the development of learning tools, such as books, teaching modules, LKPD, and so on (Arianatasari &; Hakim, 2018; Fauziah et al., 2021; Fitri et al., 2020). The 4D development model is also a development model intended for the development of learning devices such as learning media and learning modules (Kosassy, 2019). To be able to develop instrument products well, a development model that matches the instrument product is needed. The development model Gable aimed at instrument development, so the development of diagnostic assessment instruments is adapted based on the development model by Gable.

The development of critical reasoning skills assessment instruments in Physics learning has been developed many times before (Mappalesye et al., 2021; Sayekti & Wasis, 2021; Yanti et al., 2019). However, the development of the instrument is still in the form of multiple-choice questions similar to cognitive learning outcomes tests with dichotomous scoring and has not referred to the instrument development model by Gable, so it has not been optimal in describing students' critical reasoning skills (Suradipa et al., 2022). Based on this, it is necessary to develop diagnostic assessment instruments that can help educators identify critical reasoning skills that are important in the industrial era 5.0. Therefore, this research aims to conclude conceptual and operational definitions and produce critical reasoning ability assessment instruments with polytomous scoring in Phase E Physics learning in Alternative Energy material as a diagnostic assessment which refers to the instrument development model by Gable.

2. METHODS

The research method applied is in the form of a mixed method. This study used qualitative data reinforced with quantitative data. This research is preliminary research from a series of development research projects with models (Gable & Wolf, 1993) to develop an instrument of assessment of critical reasoning skills with stages such as in Figure 1.

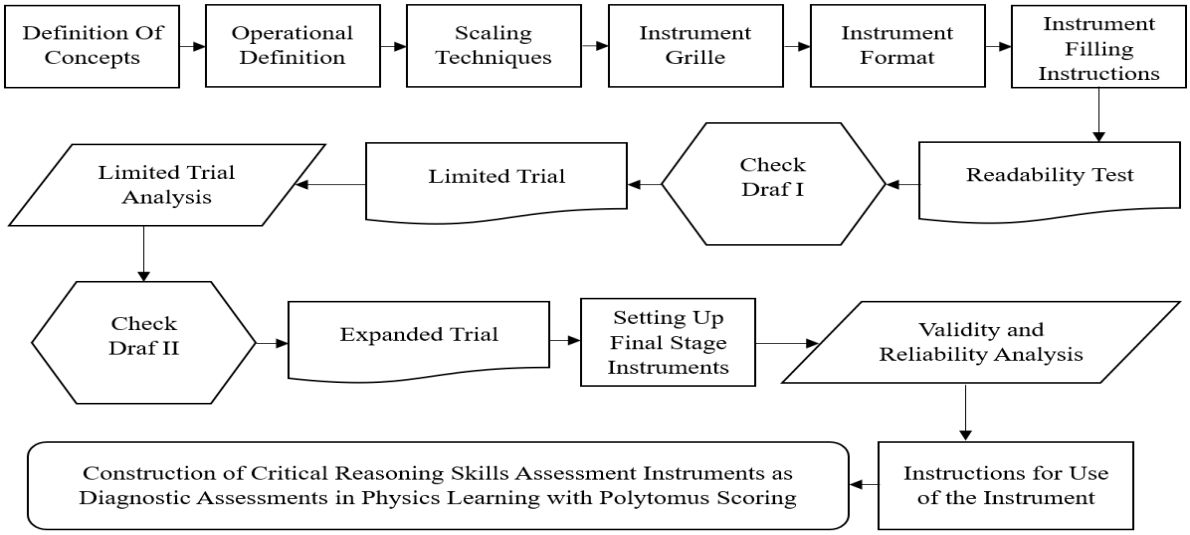


Figure 1. Flowchart Research Procedure

Preliminary research includes literature review activities and empirical research. Study activities literature includes information gathering activities, such as conceptual and operational definitions of critical reasoning abilities. The results of the literature study produce indicators of critical reasoning skills which are then assessed indicators in the form of questionnaires with methods Delphi by 9 experts. The results of the assessment questionnaire are analyzed using a formula aiken (Aiken, 1985) to determine the essential indicators that will be used in the developed instrument. An indicator can be said to be essential if it meets the criteria that the indicator can be measured and can diagnose early the critical reasoning skills of learners. The indicators analyzed by such experts have a grid in Table 1.

Table 1. Instrument Grille

No.	Aspects	Indicator No.	Indicators
1	Clarification	1	Defining terms, assessing definitions, and dealing with misconceptions
		4	Identify a problem
		11	Criticizing an argument
2	Decision	2	Demonstrate evidence from the literature to support analysis, reflection, and research
		15	Reason abstractly and quantitatively
3	Analysis	3	Identify linkages between statements, data facts, and concepts
		16	Leverage data for solutions
4	Troubleshoot Problems	5	Find the cause of the problem
		7	Assess the impact or consequences of a problem
		8	Predict the next outcome of the impact of a problem
		10	Compare similarities and differences of a condition
5	Self-Regulation	6	Provide a solution to a problem
		9	Organize self-existence in the face of problem-solving
6	Explanation	12	Present arguments and formulate them rationally based on the information or data obtained
7	Inference	13	Summing up the results of the analysis
8	Metacognitive	14	Make and defend decisions

The instruments that have been prepared, then get validation by 3 Physics Education lecturers who are experts in the field of educational measurement and experts in the field of Physics learning. The assessment is carried out by the interview method. The aspects assessed include material aspects, aspects of measuring critical reasoning skills, aspects of problem construction, and language aspects. Furthermore, a readability test was carried out by 3 class X students at SMA Negeri 4 Surakarta using the interview method. The aspects assessed include aspects of problem construction and aspects of language. The finished assessment instrument media was tested for navigation of 3 grade X students at SMA Negeri 4 Surakarta using the interview method. The aspects assessed include media aspects and display aspects. The instrument was then tested on a limited basis (initial trial) into 3 different schools, SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta, and SMA Islam 1 Surakarta. The results of the critical reasoning assessment instrument are analyzed with the Quest program to determine the estimation of the difficulty index of an item, the estimation of item fit to the Rasch model, and the estimation of item acceptance.

3. RESULT AND DISCUSSION

Results

Construction instrument for measuring the initial skills of critical reasoning of students begins with conducting theoretical studies related to aspects and indicators of critical reasoning. The basic theory of instrument construction of the initial skills to reason critically is based on theories proposed by Ennis (1993), Greenstein (2012), Facione (2010), and Glaser (1941). The results of theoretical studies, aspects, and indicators of critical reasoning skills. According to the results of the literature review obtained regarding the collection of aspects and indicators of critical reasoning, it was concluded that there are 8 aspects and 16 indicators of critical reasoning presented in Table 2.

Table 2. Critical Reasoning Aspects and Indicators

Theory	Aspects	Indicators
Ennis (1993)	Clarification	Identify the center of the problem, question, or conclusion
		Analyze arguments
		Ask and provide answers to questions to clarify or face challenges
		Decipher expressions, evaluate understanding and overcome confusion with terms
		Identify unstated assumptions
	Decision	Evaluate the trustworthiness of a particular source
		Observe and assess observation reports

Theory	Aspects	Indicators
Greenstein (2012)	Conclusion Drawing	Summing up and assessing deductions
		Generalizing and explaining conclusions as well as hypotheses
		Compile and evaluate assessments
	Metacognitive Abilities	Review and describe the basis of basic statements, explanations, allegations, stands, and other ratios that are not agreed with or in doubt
		Integrate other abilities and attitudes in making and maintaining decisions
		Using some type of reasoning appropriate to the situation
		Respond to audience needs, tasks, and objectives
		Construction of viable arguments and critiquing other people's reasoning
		Demonstrate evidence and literature to support analysis, reflection, and research
		Analyze the purpose of the information provided
		Reasoning abstractly and quantitatively
		Using random data from the results of other experiments to compare two treatments
Facione (2010)	Interpretation	Describe and interpret a problem
	Analysis	Examine and recognize the relationship between expressions, information, and ideas, and be able to present conclusions
	Inference	Identify and explore concepts or elements to reach conclusions
	Evaluation	Can filter the truth of an expression or description and take advantage of the relationship between statements, data, facts, concepts, or other ideas
	Explanation	Can present arguments and formulate them rationally based on the information or data obtained
	Self-Regulation	Skills to supervise oneself in the application, analysis, and evaluation of the results of previous thoughts when solving a problem
Glaser (1941)	Problem-facing behavior	Identify central issues
		Compare similarities and differences
		Create and formulate critical questions
	Troubleshooting	Find the cause of the problem
		Assess the impact or consequences
		Predict the subsequent impact of the event
	Application of the Method	Explain the problem and make conclusions
		Design a simple solution
		Reflect the values or attitudes of an event

These aspects and indicators were selected to determine the essential critical reasoning aspects and indicators through the questionnaire dissemination process using the Delphi method. The Delphi method aims to collect systematic opinions from a group of 9 experts anonymously (not knowing the judgments between one expert and another). The results of the questionnaire were analyzed using a formula aiken, with the results of the analysis stated in Table 3. Critical reasoning aspects and indicators are considered essential if they have a value $V_{calculate}$ greater than or equal to 0.8. Table 3 shows that from 8 aspects and 16 indicators featured hotels 5 aspects and 6 essential critical reasoning indicators. These aspects include clarification, decision, analysis, explanation, and inference. Essential critical reasoning indicators include indicators that identify a problem, indicators that show evidence from the literature to support analysis, reflection, and research, indicators that identify linkages between statements, data facts, and concepts, indicators that utilize data to obtain solutions, indicators that present arguments and formulate them rationally based on obtained information or data and indicators summing up the results of the analysis.

Table 3. Content Validity Test Results with V Aiken

No.	Aspects	Indicator Number	Calculate	V_{tabel}	Information
1	Clarification	1	0.741	0.8	Non-Essential
		4	0.926	0.8	Essential

No.	Aspects	Indicator Number	Calculate	V_{tabel}	Information
2	Decision	11	0.778	0.8	Non-Essential
		2	0.963	0.8	Essential
		15	0.778	0.8	Non-Essential
3	Analysis	3	0.926	0.8	Essential
		16	1.000	0.8	Essential
4	Troubleshoot Problems	5	0.778	0.8	Non-Essential
		7	0.778	0.8	Non-Essential
		8	0.778	0.8	Non-Essential
		10	0.630	0.8	Non-Essential
5	Self-Regulation	6	0.778	0.8	Non-Essential
		9	0.630	0.8	Non-Essential
6	Explanation	12	0.889	0.8	Essential
7	Inference	13	0.963	0.8	Essential
8	Metacognitive	14	0.778	0.8	Non-Essential

The initial draft of the instrument was developed from 5 aspects with 6 indicators above, resulting in 15 points of questions with polytomous scoring on Alternative Energy material. The content of the question instrument was discussed with 3 experts in the field of Physics Education and obtained 12 questions that were used in the next trial. To know that the problem can be understood by students, a readability test was performed on 3 learners who are studying Physics in phase E, with The result is that the instrument is easy to understand and easy to use by learners. The final instrument has been developed, packaged in electronic media, and carried out navigation tests by 3 personal learners. The results of the navigation test state that students are easy and effective in operating the instrument through electronic media. The diagnostic assessment instrument was carried out in a limited trial of three schools including SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta, and SMA Islam 1 Surakarta with a total of 62 respondents. This trial aims to determine the quality of the question items that have been developed by knowing the estimation of the difficulty index of an item, estimation of item fit to the Rasch model, and estimation of item acceptance. Analysis of the quality of the question items was carried out using the grain response theory with the 1-PL logistic model or the model Rasch assisted by the Quest program.

Based on the results of the analysis, the difficulty index of the item on the diagnostic assessment instrument of students' critical reasoning skills in the column threshold item estimate. The criteria for instrument items for diagnostic assessment of students' critical reasoning skills are expected to meet easy criteria, with follow categorization criterion. Based on the results of the analysis, information is obtained that all items have values threshold between -0.13 to 1.27 which is in the easy criteria. The results of the grain difficulty index can be interpreted as the conformity of the item to the purpose of measurement. Based on the results of the analysis, information is obtained that all items have values Infit MNSQ between 0.92 to 1.18 are included in the criteria to match the Rasch model. These results provide information that all question items that have been made are appropriate to measure students' initial critical reasoning skills.

Value Outfit t in the Quest program is used to determine which question items are accepted or rejected from the analysis process. Criterion the item received is the item that has a value Outfit t that is smaller than 2.00 or follows the criteria. Items received are items that can be used for the next trial process, while rejected items mean items that are not used in the next trial process. The results of the analysis using the Quest program show that all values Outfit t in the interval between –0.5 to 1.00 which is included in the criteria for question items accepted. Value summary Outfit t and the decision on receipt of items based on the results of the initial trial. Conclusions about the quality of the questions can be seen from the estimation of the difficulty index of an item, estimation of item fit to the Rasch model, and estimation of item acceptance with criteria.

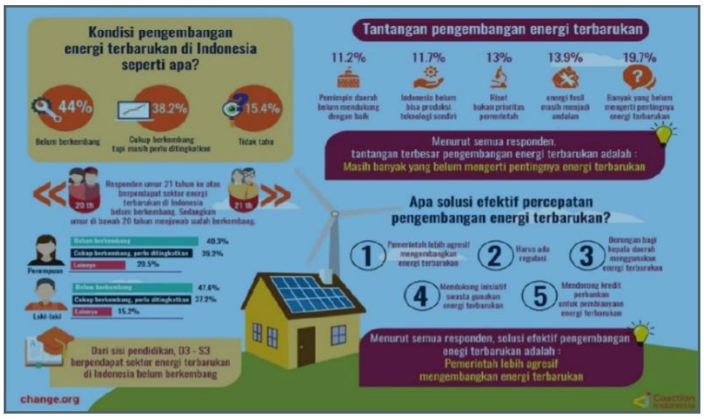
The results of the analysis using the Quest program show that all items have values threshold (b) between -0.13 to 1.27 which falls into the easy category, value Outfit t in the range of -0.6 to 1.0 which falls into the category of question items is accepted, and grades Infit MNSQ between 0.94 to 1.18 which falls into the question item category matches the Rasch model. Overall, the items on the assessment instrument diagnose students' critical reasoning abilities as seen from the threshold value (b), Outfit t value, and MNSQ Infit value. The results of the analysis using the Quest program show that all question items have a threshold value (b) between -0.13 to 1.27, which is included in the easy category; outfit t value is in the range -0.6 to 1.0, which is included in the question item category accepted, and the MNSQ Infit value is between 0.94 to 1.18 which is included in the item category that fits the Rasch model. All question items have met the criteria for a good instrument and can be tested on a broader scale.

Discussion

As a diagnostic assessment in Physics learning, the critical reasoning skills assessment instrument was developed using polytomous scoring, as presented in Figure 2.

1. Perhatikan infografis tantangan dan solusi energi terbarukan pada *Gambar 1*!

Gambar 1. Tantangan dan Solusi Energi Terbarukan



Sumber: www.change.org

Berdasarkan *Gambar 1* dapat disimpulkan bahwa…

a. Energi fosil masih menjadi andalan sampai saat ini (2 poin)
b. Energi terbarukan perlu ditindak lanjuti dengan cepat (4 poin)
c. Perlu adanya biaya dalam mengolah energi terbarukan (1 poin)
d. Sektor energi terbarukan di Indonesia belum berkembang (3 poin)

Figure 2. Critical Reasoning Skills Assessment Instrument with Polytomus Scoring

In the Figure 2, each answer choice option on the assessment instrument has a different level of score gradation so that it can assess the complexity of answers from students, which is an essential aspect of critical reasoning. This assessment system allows educators to diagnose students' understanding at the beginning of learning in more depth and provide more targeted tutoring. The development of diagnostic assessment instruments for students' critical reasoning skills uses a multilevel multiple-choice system with a dichotomous multiple-choice design. Dichotomous assessment instruments, such as multiple-choice tests, are not sufficient to measure essential skills of reasoning (Dirman & Mufit, 2022; Facione & Facione, 1996; Sunarti et al., 2021). Dichotomous assessments can only measure learners' recall of facts and cannot measure their critical reasoning skills. Multiple-choice assessment instruments tend to lead learners to strategies of remembering facts and memorizing answers rather than critically analyzing problems. The use of more authentic assessment instruments is more appropriate for measuring learners' critical skills. Thus, the use of a polytomous scoring system for diagnostic assessment instruments of learners' essential skills of reasoning is appropriate (Ennis, 1993; H. Putri et al., 2022; Thomas & Thorne, 2009).

Diagnostic assessment aims to identify the level of mastery of student competencies that will affect the effectiveness of the learning process, so the assessment given must be in the form of questions related to the abilities mastered by students (Nur Budiono &; Hatip, 2023). The results of the analysis of the item difficulty index criteria on the diagnostic assessment instrument of students' critical reasoning skills show that all items have a threshold value between -0.13 to 1.27, and the 12 questions as a whole have a low level of difficulty or are on easy criteria. These results are in accordance with the purpose of diagnostic assessment, which is to measure students' initial abilities so that the question instruments given to students must have a low or easy level of difficulty. Other researchers also revealed that the results of the analysis of diagnostic assessments in science subjects showed that 60% of the total questions had a manageable level of difficulty. The results of the study of the difficulty level of questions that are classified as easy indicate that the question instrument is still within the scope of mastery of students so that it is suitable for mapping the initial ability of students' critical reasoning (Forniawan & Wati, 2023; Husnaini et al., 2018; A. S. Putri et al., 2021).

This is different from instruments for assessing critical reasoning abilities from learning outcomes, which have varying levels of difficulty. Research on the development of instruments for evaluating essential abilities of reasoning shows different levels of difficulty and is dominated by moderate levels of difficulty (Mappalesye et al., 2021; Sayekti & Wasis, 2021; Yanti et al., 2019). Instruments that measure critical reasoning abilities from learning outcomes students are used to measure essential reasoning abilities as well as a form of training in developing critical reasoning abilities. The difficulty level of the instrument will

vary where the distribution of difficulty levels approaches the normal distribution(Satria & Istiyono, 2018; Sumarni et al., 2018). The instrument for measuring student learning outcomes is also carried out at the end of learning, where its function is to measure learning outcomes and critical reasoning abilities so that students have better competencies than before learning took place. Meanwhile, in diagnostic assessments, measurement activities are carried out before learning is carried out with the aim of mapping student profiles before being given treatment. The results of this assessment will be used as a basis for designing the learning process based on the student's initial competencies so that the difficulty level of the questions used must be in the easy category or within the scope of the student's mastery.

The MNSQ Infit value is used to determine which question items are fit or compatible with the Rasch model. If there are items that are not suitable, they can be removed (Mardapi, 2016; Setyawarno, 2017). The characteristics of the diagnostic assessment instrument items for students' critical reasoning abilities in physics lessons meet the criteria according to the Rasch model, namely that the instrument items have a uniform response distribution, have high instrument discrimination; there is no correlation between items and the instrument can be used consistently to measure the same ability. The results of the analysis of the suitability of the items on the diagnostic assessment instrument for students' critical reasoning abilities using the Rasch model have an Infit MNSQ value between 0.92 and 1.18, which is included in the criteria for matching the Rasch model. These results provide information that all the question items that have been created are suitable for measuring students' initial critical reasoning abilities. The Outfit t value is used to determine whether the question item is accepted or rejected from the analysis process (Hasan et al., 2020; Setyawarno, 2017). This value is used in the Rasch model and several other models to identify items that do not fit the model, improve instrument quality, detect Differential Item Functioning (DIF), and analyze data in a multidimensional Rasch model (Azizah &; Wahyuningsih, 2020).

The results of the analysis of the acceptance of items on the diagnostic assessment instrument for students' critical reasoning abilities show that all Outfit t values are in the interval between -0.5 and 1.00, which is included in the criteria for accepted items. These results provide information that all the question items that have been created can be used to measure students' initial critical reasoning abilities. Furthermore, there is other research regarding analysis of the quality of instruments for assessing students' essential skills of reasoning using a classical theory approach, where analysis using classical theory still has weaknesses, such as measurement results will depend on the characteristics of the test used, item parameters depend on the test taker's abilities and measurement errors. It can only be known for groups, not individuals (Mappalesye et al., 2021; Sayekti & Wasis, 2021; Yanti et al., 2019). This deficiency can be overcome by using the Rasch model, which will produce more comprehensive information about the instrument and better meet the definition of measurement. This model is designed to build instruments that comply with the theory and basic requirements of objective measurement (Mardapi, 2016; Ridwan et al., 2023; Wibisono, 2018). Thus, the use of the Rasch model as an analysis of instrument quality is appropriate.

However, this study has some limitations. First, the assessment instrument used in this study had a low difficulty level, with all items in the easy category. The purpose of the diagnostic assessment is to measure students' initial abilities. Still, it may make the instrument more challenging for students with higher critical reasoning abilities. Also, the instrument was developed specifically for Physics learning, meaning its findings and validity may only partially apply to other subjects requiring essential reasoning. Therefore, it is possible to create an assessment instrument with a wider variety of difficulty levels to overcome the limitation of the low difficulty level. This includes adding questions with medium to high difficulty levels to better test students' critical reasoning skills. Another solution is that further development needs to be done for other subjects that require essential reasoning skills. This will broaden the application and benefits of this assessment instrument and ensure that students in different subject areas benefit from a comprehensive diagnostic assessment.

The research implications are as follows: First, using assessment instruments with polytomous scoring allows for a more detailed and in-depth assessment of student understanding. With various levels of score gradations, teachers can evaluate the level of complexity of student answers, which is an essential aspect of critical reasoning. This provides a more comprehensive view of students' critical thinking abilities from the beginning of learning so that learning can be tailored to the needs and initial abilities. Secondly, the results showed that the polytomous assessment system is more effective than the previously used dichotomous assessment system. Dichotomous assessment tends only to measure memory and memorization, while polytomous assessment is able to measure critical analysis and problem-solving skills. Therefore, implementing polytomous scoring in assessment instruments is expected to encourage teachers to integrate further learning activities that emphasize developing students' critical thinking skills. Third, using the correct diagnostic assessment instrument, teachers can make learning designs more focused on the student's ability profile. The results of this diagnostic assessment provide accurate information about students' initial abilities so that teachers can adjust more effective teaching strategies and methods to

improve learning outcomes. Finally, the developed assessment instrument can map students' initial competencies well, especially regarding critical reasoning ability. This is crucial in physics learning, which requires analytical and critical thinking skills. By knowing students' initial level of mastery, teachers can more easily design appropriate learning interventions to address students' weaknesses and strengthen their strengths.

4. CONCLUSION

The limited trial analysis of 12 items on the diagnostic assessment instrument for students' initial critical reasoning abilities using the Quest program showed that all items were in the easy category, by the Rasch model, and were acceptable (no items were rejected). These findings indicate that the 12-question instruments prepared are in a suitable category and can be used to measure students' initial critical reasoning abilities. Constructing this critical reasoning ability assessment instrument is ideal for diagnosing students' critical reasoning abilities. To ensure the instrument's reliability, trials involving a wider variety of student respondents are needed on a broader scale. This diagnostic assessment instrument can be implemented in physics learning, especially during assessment activities, to measure students' critical reasoning abilities at the beginning so that teachers can design learning that suits the student's profile. The context and operational definitions developed in this research can be used as a basis for developing critical reasoning instruments in different materials and other subjects.

5. ACKNOWLEDGE

Thanks to Universitas Sebelas Maret for providing a group research grant for RG Management and Physics Learning Innovation from non-PNPB funds with contract number 228/UN.27/PT.01.03/2023, so that this research can be carried out properly. Thank you to SMA Negeri 4 Surakarta, SMA Negeri 7 Surakarta, and SMA Islam 1 Surakarta for their cooperation in the research process.

6. REFERENCES

Aiken, L. R. (1985). Three Coefficients for Analyzing The Reliability and Validity of Ratings, Educational and Psychological Measurument. *Journal Articles; Reports - Research; Numerical/Quantitative Data*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012.

Andayani, A., Purwanto, & Ramalis, T. R. (2019). Kajian Implementasi Teori Respon Butir Dalam Menganalisis Instrumen Tes Materi Fisika. *Prosiding Seminar Nasional Fisika 5.0*, *1*(1), 37–42.

Antika, W., Sasomo, B., & Rahmawati, A. D. (2023). Analisis Asesmen Diagnostik Pada Model Pembelajaran Project Based Learning di Kurikulum Merdeka SMPN 3 Sine. *Pedagogy*, *8*(1), 250–263. https://doi.org/10.30605/pedagogy.v8i1.2512.

Arianatasari, A., & Hakim, L. (2018). Penerapan Desain Model Plomp Pada Pengembangan Buku Teks Berbasis Guided Inquiry. *Jurnal Pendidikan Akuntansi*, *6*(1), 36–40.

Arini, W., & Juliadi, F. (2018). Analisis Kemampuan Berpikir Kritis pada Mata Pelajaran Fisika untuk Pokok Bahasan Vektor Siswa Kelas X SMA Negeri 4 Lubuklinggau, Sumatera Selatan. *UAD Journal Management System*, *10*(1), 1–11.

Arrohman, D. A., & Lestari, T. (2023). Analisis Keragaman Peserta Didik dan Implementasi Kurikulum Merdeka pada Mata Pelajaran Fisika. *JSER Journal of Science and Education Research*, *2*(2), 1–11. https://doi.org/10.62759/jser.v2i2.29.

Atılgan, H., Demir, E. K., Ogretmen, T., & Başokçu, T. O. (2020). The Use of Open-ended Questions in Large-Scale Tests for Selection: Generalizability and Dependability. *International Journal of Progressive Education*, *16*(5), 216–227. https://doi.org/10.29329/ijpe.2020.277.13.

Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, *3*(1), 45–50. https://doi.org/10.30598/jupitekvol3iss1pp45-50.

Dirman, H. M., & Mufit, F. (2022). Analisis Penggunaan Instrument Penilaian Pemahaman Konsep dan Literasi Sains di SMA Kabupaten Solok. *Jurnal Penelitian Pembelajaran Fisika*, *13*(2), 251–256. https://doi.org/10.26877/jp2f.v13i2.12923.

Ennis, R. H. (1993). Critical Thinking Assessment. *Theory Into Practice*, *32*(3), 179–186. https://doi.org/10.1080/00405849309543594.

Facione, N. ., & Facione, P. . (2010). *Externalizing, The Critical Thinking in Knowledge Development and Clinical Judgment*. Nursing Outlook.

Facione, N. C., & Facione, P. A. (1996). Externalizing the Critical Thinking in Knowledge Development and

Clinical Judgment. *Nursing Outlook*, *44*(3), 129–136. https://doi.org/10.1016/S0029-6554(06)80005-9.

Fan, T., Song, J., & Guan, Z. (2021). Integrating Diagnostic Assessment into Curriculum: A Theoretical Framework and Teaching Practices. *Language Testing in Asia*, *11*(1), 2. https://doi.org/10.1186/s40468-020-00117-y.

Fauziah, N., Putri, I. I., & Oktariani, O. (2021). Analisis Preliminery Research Phase Menggunakan Model Pengembangan Plomp Sebagai Dasar Pengembangan Modul Belajar Dan Pembelajaran Pendidikan Biologi Bermuatan Hasil Riset Untuk Mahasiswa. *Jurnal Bioterdidik: Wahana Ekspresi Ilmiah*, *9*(3), 226–233. https://doi.org/10.23960/jbt.v9i3.23072.

Febliza, A., Afdal, Z., & Copriady, J. (2023). Improving Students' Critical Thinking Skills: Is Interactive Video and Interactive Web Module Beneficial? *International Journal of Interactive Mobile Technologies (IJIM)*, *17*(03), 70–86. https://doi.org/10.3991/ijim.v17i03.34699.

Firmanzah, D., & Sudibyo, E. (2021). Implementasi Asesmen Diagnostik dalam Pembelajaran IPA pada Masa Pandemi Covid-19 Di SMP/MTS Wilayah Menganti, Gresik. *Pensa E-Jurnal: Pendidikan Sains*, *9*(2), 165–170.

Fitri, M., Yuanita, P., & Maimunah, M. (2020). Pengembangan Perangkat Pembelajaran Matematika Terintegrasi Keterampilan Abad 21 Melalui Penerapan Model Problem Based Learning (PBL). *Jurnal Gantang*, *5*(1), 77–85. https://doi.org/10.31629/jg.v5i1.1609.

Forniawan, A., & Wati, D. R. (2023). Analysis and Follow-Up of Cognitive Diagnostic Assessment Result in Natural Science Subjecys. *Al-Jahiz: Journal of Biology Education Research*, *4*(2), 164–179. https://doi.org/10.32332/al-jahiz.v4i2.7962.

Gable, R. K., & Wolf, M. B. (1993). *Instrument Development in the Affective Domain (2nd Edition): Evaluation in Education and Human Services* (G. F. Madaus & D. L. Stufflebeam (eds.); 2nd ed.). Springer Science+Business Media, LLC. https://doi.org/10.1007/978-94-011-1400-4

Glaser, E. M. (1941). *An Experiment in the Development of Critical Thinking* (2nd ed.). Teachers College, Columbia University.

Greenstein, L. (2012). *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning*. Corwin Press.

Hasan, S. W., Auliah, A., & Herawati, N. (2020). Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Siswa SMA. *Chemistry Education Review (CER)*, *3*(2), 185–193. https://doi.org/10.26858/cer.v3i2.13769.

Herwin, H., Nurhayati, R., Lidyasari, A. T., & da Costa, A. (2023). Graded Response Models on the Curiosity Measurement of Elementary School Students. *International Journal of Educational Methodology*, *9*(1), 53–62. https://doi.org/10.12973/ijem.9.1.53.

Holiqi, E. S., Handono, S., & Prastowo, B. (2021). Development of Contextual Assessment Instruments for Students ' Critical Thinking Skills Using Kahoot Quiz. *Jurnal Pendidikan Fisika*, *10*(2), 176–186. https://doi.org/10.24252/jpf.v10i2.30274.

Husnaini, G., Sutrisno, L., & Silitonga, H. T. M. (2018). Pengembangan Tes Diagnostik Pada Materi Suhu dan Kalor di SMA. *Jurnal Pendidikan Dan Pembelajaran*, *7*(3), 1–10.

Kahn-Horwitz, J., & Goldstein, Z. (2023). English Foreign Language Reading and Spelling Diagnostic Assessments Informing Teaching and Learning of Young Learners. *Language Testing*, *1*(1), 1–29. https://doi.org/10.1177/02655322231162838.

Kosassy, S. O. (2019). Mengulas Model-Model Pengembangan Pembelajaran dan Perangkat Pembelajaran. *Jurnal PPKn Dan Hukum*, *14*(1), 152–173.

Leniati, B., & Indarini, E. (2021). Meta Analisis Komparasi Keefektifan Model Pembelajaran Kooperatif Tipe Jigsaw dan TSTS (Two Stay Two Stray) Terhadap Kemampuan Berpikir Kritis pada Pembelajaran Matematika Siswa Sekolah Dasar. *Mimbar Ilmu*, *26*(1), 149–157. https://doi.org/10.23887/mi.v26i1.33359.

Mappalesye, N., Sari, S. S., & Arafah, K. (2021). Pengembangan Instrumen Tes Kemampuan Berpikir Kritis dalam Pembelajaran Fisika. *Jurnal Sains Dan Pendidikan Fisika (JSPF)*, *17*(1), 69–82. https://doi.org/https://doi.org/10.35580/jspf.v17i1.19091.

Mardapi, D. (2016). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Parama Publishing.

Martanti, F., Widodo, J., Rusdarti, R., & Priyanto, A. S. (2022). Penguatan Profil Pelajar Pancasila Melalui Pembelajaran Diferensiasi Pada Mata Pelajaran IPS di Sekolah Penggerak. *Prosiding Seminar Nasional Pascasarja*, 415–417.

Mirunnisa, M., & Razi, Z. (2021). Analysis of High School Students 'Mathematic Critical Thinking Ability with Graded Response Models. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, *4*(1), 1108–1116. https://doi.org/10.33258/birci.v4i1.1719.

Nugroho, D., Febriantania, P., & Ridaningsih, I. (2023). A Sistematic Literature Review: Implementasi

Asesmen Diagnostik pada Kurikulum Merdeka. *Journal of Islamic Education ANNABA*, *9*(2), 1–12. http://dx.doi.org/10.37286/ojs.v9i2.197.

Nur Budiono, A., & Hatip, M. (2023). Asesmen Pembelajaran Pada Kurikulum Merdeka. *Jurnal Axioma : Jurnal Matematika Dan Pembelajaran*, *8*(1), 109–123. https://doi.org/10.56013/axi.v8i1.2044.

Okta Susilawati, W., Yulia Friska, S., & Rohmawanti. (2023). Pengembangan Asesmen Diagnostik Mata Pelajaran Pendidikan Pancasila Kelas Iv Dalam Kurikulum Merdeka Di Sekolah Dasar. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, *9*(2), 3129–3144. https://doi.org/10.36989/didaktik.v9i2.1183.

Prahani, B. K., Jatmiko, B., Pristianti, M. C., Satriawan, M., Amelia, T., & Makahinda, T. (2023). The dHOTLearn Model to Improve Critical Thinking Skills of Physics Education Program Undergraduate. *Journal of Higher Education Theory and Practice*, *23*(19), 113–124. https://doi.org/10.33423/jhetp.v23i19.6676.

Puspita Rakhmi, M., Purwo Yudi Utomo, A., Ashfiya Adiratna Salim Putri, A., & Ghufron, W. (2023). Pemanfaatan Google Form dalam Asesmen Diagnostik di SMA Negeri 11 Semarang. *Concept: Journal of Social Humanities and Education*, *2*(1), 115–126. https://doi.org/10.55606/concept.v2i1.236.

Putri, A. S., Silitonga, H. T. M., & Hamdani, H. (2021). Pengembangan Tes Diagnostik Two Tier Menggunakan Aplikasi Google Form Pada Materi Tekanan Di SMP. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa (JPPK)*, *11*(1), 1–8. https://doi.org/10.26418/jppk.v11i1.51712.

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, *4*(2), 139–148. https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649.

Razak, F. (2017). Hubungan Kemampuan Awal terhadap Kemampuan Berpikir Kritis Matematika pada Siswa kelas VII SMP Pesantren Immim Putri Minasatene. *Mosharafa: Jurnal Pendidikan Matematika*, *7*(1), 117–128.

Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramalis, T. R., Mahmudah, I. R., & Vilmala, B. K. (2023). Rasch Modeling: Developing a Critical Thinking Ability Test for Students on Climate Change Material. *Momentum: Physics Education Journal*, *7*(2), 247–258. https://doi.org/10.21067/mpej.v7i2.8310.

Rukmana, A. P. dkk. (2019). Pengembangan Four-Tier Diagnostic Test untuk Mendeteksi Miskonsepsi pada Fisika SMA. *SNPF (Seminar Nasional Pendidikan Fisika)*, 1–5.

Saepuzaman, D., Istiyono, E., Haryanto, H., Retnawati, H., & Yustiandi, Y. (2021). Analisis Karakteristik Butir Soal Fisika Dengan Pendekatan IRT Penskoran Dikotomus dan Politomus. *Radiasi : Jurnal Berkala Pendidikan Fisika*, *14*(2), 62–75. https://doi.org/10.37729/radiasi.v14i2.1200.

Satria, M. aufiq, & Istiyono, E. (2018). Instrumen Penilaian Kemampuan Berpikir Kritis Peserta Didik SMA Negeri Mata Pelajaran Biologi Kelas X. *BIOEDUKASI: Jurrnal Pendidikan Biologi*, *11*(2), 105–110. https://doi.org/10.20961/bioedukasi-uns.v11i2.21624.

Sayekti, A. A., & Wasis, W. (2021). Instrumen Tes Berbasis Web Menggunakan Framework Codeigniter untuk Menilai Kemampuan Berpikir Kritis Siswa SMA. *PENDIPA Journal of Science Education*, *5*(2), 210–217. https://doi.org/10.33369/pendipa.5.2.210-217.

Setyawarno, D. (2017). *Upaya Peningkatan Kualitas Butir Soal dengan Analisis Aplikasi Quest*. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.

Suarjana, I. M., Lasmawan, I. M., & Gunamantha, I. . (2020). Pengembangan Instrumen Kemampuan Berpikir Kritis dan Sikap Peduli Lingkungan Tema 8 Peserta Didik Kelas IV SD. *Jurnal Pendidikan Dasar Indonesia*, *4*(2), 101–111. https://doi.org/10.23887/jpdi.v4i2.3345.

Suharno, S., Selviana, A. S., & Sunarno, W. (2022). The Effectiveness of Using Physics Module with Problem-Based Learning to Enhance Critical and Creative Thinking Skills. *Journal of Education Research and Evaluation*, *6*(1), 19–25. https://doi.org/10.23887/jere.v6i1.35476.

Suhirman, & Prayogi, S. (2023). Problem Based Learning Utilizing Assistive Virtual Simulation in Mobile Application to Improve Students' Critical Thinking Skills. *International Journal of Education and Practice*, *11*(3), 351–364. https://doi.org/10.18488/61.v11i3.3380.

Sumarni, W., Supardi, K. I., & Widiarti, N. (2018). Development of Assessment Instruments to Measure Critical Thinking Skills. *IOP Conference Series: Materials Science and Engineering*, *349*(1), 1–11. https://doi.org/10.1088/1757-899X/349/1/012066.

Sunarti, T., Wasis, W., Setyarsih, W., Zainudin, A., & Jauhariyah, M. N. R. (2021). Analysis of the Development of Critical Thinking Instruments Test in Physics. *International Joint Conference on Science and Engineering 2021 (IJCSE 2021)*, *209*(IJCSE), 627–634. https://doi.org/10.2991/aer.k.211215.105.

Suradipa, I. A., Putrayasa, I. B., & Gunamantha, I. M. (2022). Pengembangan Instrumen Literacy Humanistic dan Kemampuan Berpikir Kritis pada Pembelajaran IPA Kelas V SD. *Jurnal Pendidikan Dasar Indonesia*, *6*(1), 105–117. https://doi.org/10.23887/jurnal_pendas.v6i1.566.

Suresman, E., Febrianti, F., & Dallyono, R. (2023). Implementation of i-Spring Suite to Improve Darul Muta'alimin Students' Learning for Critical Thinking Skills in Natural Science. *Jurnal Cakrawala*

Pendidikan, *42*(2), 433–446. https://doi.org/10.21831/cp.v42i2.53646.

Thomas, A., & Thorne, G. (2009). *How to Increase Higher Order Thinking* (2nd ed.). Center for Development and Learning.

van der Mark, C., Bitter, J., Hendriks, P. H. J., Vermeulen, H., & van Oostveen, C. J. (2023). The Nurse Perceived Adequacy of Staffing Scale for General Hospital Wards: A Development and Psychometric Validation Study. *International Journal of Nursing Studies Advances*, *5*(1), 1–12. https://doi.org/10.1016/j.ijnsa.2023.100138.

Wibisono, S. (2018). Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *5*(1). https://doi.org/10.15408/jp3i.v5i1.9239.

Wijayanti, R., & Siswanto, J. (2020). Profil Kemampuan Berpikir Kritis Siswa SMA pada Materi Sumber-sumber Energi. *Jurnal Penelitian Pembelajaran Fisika*, *11*(1), 109–113. https://doi.org/10.26877/jp2f.v11i1.5533.

Wiyoko, T. (2019). Analisis Profil Kemampuan Berpikir Kritis Mahasiswa PGSD Dengan Graded Response Models Pada Pembelajaran IPA. *IJIS Edu : Indonesian Journal of Integrated Science Education*, *1*(1), 25–32. https://doi.org/10.29300/ijisedu.v1i1.1402.

Yanti, T. D., Suana, W., Maharta, N., Herlina, K., & Distrik, I. W. (2019). Development of Critical Thinking Instrument of Electricity for Senior High School Students. *Journal of Physics: Conference Series*, *1157*(3), 1–5. https://doi.org/10.1088/1742-6596/1157/3/032007.