

Using scientific reasoning test as assessment tool differentiated science learning for grade VIII junior high school

Afida Nurul Sabilla^{1*}, Purwo Susongko², Muriani Nur Hayati³

^{1,2,3}Study Program of Science Education, University of Pancasakti Tegal, Indonesia

*Correspondence: afidanurulsabilla3@gmail.com

Abstract

Keywords:
Assessment;
Profile;
Science learning;
Scientific reasoning;
Validity test

The process of learning science in class VIII junior high school teachers only provide direct observation of a problem to students, so that the level of scientific reasoning of students is low. This study evaluates the efficacy of junior high school students' grade VIII scientific reasoning test as a measure of differentiated science instruction. A quantitative exploratory descriptive study was conducted. Scientific reasoning instrument tests based on rudimentary, basic, and intermediate levels are being collected using Google Forms. The conclusions were formulated from the data obtained from the instrument test. Research data is analyzed using tests for construct, psychometric, and content validity. The outcomes demonstrated that the scientific reasoning test was deemed practical for use in evaluating differentiated science learning based on psychometric tests and content validity. Construct validity test on item suitability obtained 28 questions said to be in accordance with the scientific reasoning test and 4 questions said to be inappropriate. 1 out of 252 students did not have the ability to answer the scientific reasoning test on the measurement of person fit statistics, and the dominant students answered at the rudimentary level number 10 and basic number 11.

To cite this article:

Sabilla, A.N., Susongko, P., & Hayati, M.N. (2024). Using scientific reasoning test as assessment tool for differentiated science learning for grade VIII junior high school. *Thabiea: Journal of Natural Science Teaching*, 7(2), 238-249.

Introduction

The 21st century is seeing a rapid advancement in science. The advancement of time has an impact on the lives of students. Among the many challenges of the twenty-first century is the requirement that students apply critical thinking abilities when obtaining information from their surroundings. These challenges can create students who thrive when engaged in learning activities (Nofianti et al., 2022).

The PISA 2022 results that have been released on Tuesday, 5 December 2023 show that Indonesia has improved its literacy and numeracy scores compared to PISA 2018. Indonesia rose 5 places from the previous year in literacy and numeracy, and rose 6 places in science literacy. The tests conducted in PISA include tests measuring skills in scientific reasoning (OECD., 2023).

Skills in scientific reasoning are closely related to scientific education. All disciplines are built on scientific reasoning, which contributes to the advancement of other scientific knowledge. Improving scientific literacy and environmental awareness are the main goals of

scientific education, especially in Indonesia. The improvement of scientific concepts that form the basis of environmental issues relies heavily on scientific education (Susongko et al., 2021).

With effect from 2022, the Ministry of Education, Culture, Research, and Technology will be substituting an autonomous curriculum for the 2013 curriculum. One curriculum in the independent curriculum makes use of diversity in the classroom to boost students' proficiency and broaden their understanding of the material. The independent curriculum needs to include learning steps that can help students recover from their experiences during the crisis. This step involves reading and writing skills that students can acquire (Marsela Yulianti et al., 2022).

The step in science learning used in the independent curriculum is science differentiated learning. Differentiated science education comprises instruction that changes to meet the requirements of every learner. One of the issues with the autonomous curriculum is the discrepancy between the content acquired and the understanding of the students. Students find it difficult to develop their interests and learning abilities. Assessment of student learning, solely considering their interests and character development is another problem that arises. In the subject of education, assessments that gauge pupils' capacity for scientific reasoning are the most crucial. (Syahbana et al., 2024).

When interviews were conducted with several teachers in one of the junior high schools in Tegal City, teachers constantly push students to use scientific reasoning as part of the classroom curriculum. Teachers use a contextual approach and students to directly observe the problem. Therefore, students in grade VIII at the junior high school had poor scientific reasoning skills. This is evidenced by the analysis of student needs described in Table 1.

Table 1. Student Needs Analysis

Number	Activity	Percentage
1.	Basic Understanding	54%
2.	How it works	77%
3.	Technique	85%
4.	Experiment	92%
5.	Experimental Design	31%
6.	Solving Problems	62%
7.	Tendency of Science Literacy	46%
8.	Formulating Questions	46%

Due to the low interest in science literacy among students, the researchers used a scientific reasoning test instrument to assess the understanding of grade VIII junior high school students in order to implement differentiated science teaching. The interview results also relate to the research findings of Yucel (2023) who explained that students' scientific reasoning depends on a knowledge of scientific practices. So that the low value of students' scientific practical knowledge has a critical effect on scientific reasoning. This is related to table 1 which explains that the percentage value of students' experiments is at 31%. This study was conducted to measure the validity of the scientific reasoning test so that it can be used as an assessment action plan in the context of differentiated science learning. This study is a new research because it develops a new test specifically for grade VIII junior high school students and assesses scientific reasoning abilities at the basic, basic, and intermediate levels that have been identified by Wenning & Vierya's research. Another gap in this study is that the material tested was taken

from the first semester science learning material in class VIII junior high school, and the research sample conducted in the data collection process did not look at gender and was carried out in class VIII junior high school. This allows different science lessons to use the results as assessment data.

Method

Students in class VIII junior high school were given scientific reasoning exam instruments as part of a quantitative descriptive study design. A research technique that provides a methodical, factual, and correct explanation, picture, or description of the topic under study is descriptive research. Descriptive approaches provide relationships between concepts and offer initial estimates for a problem that needs to be solved. Exploratory descriptive research looks at the elements that naturally influence an event to address the problem. Exploratory descriptive research reporting uses narrative sentences, not numerical. The research process was conducted using scientific reasoning tests based on Rudimentary, Basic, and Intermediate levels. The results obtained are then presented and conclusions are formulated (Zukhruf., 2020).

The results of student answers were then analysed using the Rash Model with the help of R studio software. The examination that was administered to the pupils was a dichotomous test, meaning that it was a multiple-choice exam with 32 test items. Correct answers yielded a score of 1, while erroneous answers yielded a score of 0. Testing of scientific reasoning tests was carried out on all class VIII junior high school students in one of the schools in Tegal City, approximately 252 students with an age range of 13-16 years are present. The number of male students is 113, and the number of female students is 139. In answering the research objectives, this test used content validity test, psychometrics validity test, and construct validity test.

Categories in content validity and psychometrics are taken based on the journal Ashrof et al., (2024) which explains that the results of content validity and psychometrics in the 80%-100% range are suitable for use as test instruments. The last measurement is construct validity. The Rasch Model, with assistance from the R computer application, is used to perform construct validity. Measuring item fit, person fit statistic, and unidimensional test are examples of construct validity. The following is explained in table 2. Related to the standard assessment of construct validity according to Messick in the journal (Natanael et al, 2023; Fauzi et al, 2022).

Table 2. Construct Validity Assessment Standards

Indicators	Assessment Standards
Item Fit Test	<i>P- Value > 0.01</i>
<i>Person fit Statistic</i>	<i>P-Value > 0.01</i>
Unidimensional Test	<i>Screeplot</i> picture factor analysis and there is one main factor

Scientific Reasoning Test Instrument

Study uses research by Susongko et al. (2023) which explains that the scientific reasoning test is a test used to determine how students think in solving a problem in the scientific world. This test uses indicators according to Wenning & Vierya which consists of 6 levels including

rudimentary, basic, intermediate, integrated, culminating, and advanced levels. In the research process, researchers only used 3 levels including rudimentary, basic, and intermediate. This is because to focus on the application of scientific reasoning in junior high school grade VIII and prevent the research from being too broad. The scientific reasoning test given to students is based on the indicators described in table 3 below.

Table 3. Indicators of Scientific Reasoning Test Instruments

Ladder	Indicators	Question Indicators
Beginning	Classification	Q ₁ Showing the difference in food type pictures
	Conceptualization	Q ₂ , Q ₃ Showing the concept of body and blood shaping
	Conclusion	Q ₄ , Q ₅ Providing conclusions on food trials and cigarette effects
	Contextualization	Q ₆ Indicates excretory processes in the body Q ₇ Mention of pulley application
	Generalization	Q ₈ , Q ₉ Showing the statement of the phenomenon of light wave propagation and simple plane
	Order	Q ₁₀ Sorting frequencies from largest to smallest Q ₁₁ Sorting periods from largest to smallest
	Problem Formulation	Q ₁₂ , Q ₁₃ Formulating problems on vibration experiments
Basis	Estimation	Q ₁₄ , Q ₁₅ Calculating the experimental data of object weight and frequency
	Explanation	Q ₁₆ , Q ₁₇ Describe the chemical digestive processes and the respiratory system
	Predictions	Q ₁₈ , Q ₁₉ Predicting values on business charts and levers
	Conditional Thinking	Q ₂₀ , Q ₂₁ Showing the conclusion of the food and wave test statement
	Application Information	Q ₂₂ , Q ₂₃ Describe the application description of optical and vibration devices
Intermediate	Relationship Explained	Q ₂₄ , Q ₂₅ Explain the relationship of natural event stories to digestive and circulatory system materials
	Quantitative Understanding	Q ₂₆ , Q ₂₇ Interpreting vibration experiments
	Applying Combinatorial Reasoning	Q ₂₈ , Q ₂₉ Describe the relationship of data statements in the chapter on potential energy and fast propagation of waves
	Applying Relationship Theory	Q ₃₀ Illustrates a graph of the relationship between 2 variables that affect power
	Clearly Defining Research Problems	Q ₃₁ , Q ₃₂ Explain the problems that arise in the concept of vibration and tilt plane in everyday life

Based on Table 3, 32 scientific reasoning test items were then analysed including content, psychometric, and construct validity analyses. Content validity according Utami et al. (2024) namely the validity that fixes the test items and shows that the test measures the things that

should be measured. This study was measured by 2 expert validators on content validity by researchers providing a sheet containing indicators and scoring according to the test instrument made in order to be given suggestions for the next. In addition to content validity, researchers also conducted psychometric validity conducted on 2 experts in their fields. According to Riyadi & Susongko. (2024) explained that psychometric validity includes validity that analyses test items for test validity includes how well the test items are prepared so that they follow psychometric guidelines and consider factors such as language, construction, and content. The last measuring construct validity test.

Results and Discussion

Validity of Content

Content validity conducted by 2 expert validators in their fields provides the results in table 4 below.

Table 4. Results of Content Validity

Validators	Aspect I			Aspect II				Aspect III				Sum	%
	a	b	c	a	b	c	d	a	b	c	d		
I	4	3	3	4	3	4	4	3	4	3	4	39	88,6%
II	3	3	3	3	3	3	3	3	3	3	3	33	75%
AVERAGE												36	81,8%
Information												Valid (Usable)	

Table 4's data demonstrate that the two validators provided an average of 36 assessments, or 81.8% of the total number. The total number of total content validity scores is 44 with a percentage value of 100%. It can be said that the scientific reasoning test instrument is valid and can be used with students as an assessment of differentiated learning in science. This is consistent with the findings of Ashrof et al. (2024), which states that a test instrument can be considered viable if its validity score falls between 80% and 100%.

Validity of Psychometrics

Psychometric validity conducted by 2 expert validators in their fields provides the results in table 5 below.

Table 5. Results of Psychometric Validity

Validators	Aspect I				Aspect II				Aspect III								Sum	%
	a	b	c	d	a	b	c	d	a	b	c	d	e	f	g			
I	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3		49	81,6%
II	3	4	3	4	3	4	3	3	3	3	4	3	3	4	3		50	83,3%
AVERAGE																	99	82,45%
Information																	Valid (Usable)	

Table 5 shows that the two validators provided an average of 99 assessments, or 82.45% of the total, on the percentage basis. The total number of total content validity scores is 120 with a percentage value of 100%. It can be said that the scientific reasoning test instrument is

valid and can be used with students as an assessment of differentiated learning in science. This is consistent with the findings of Ashrof et al. (2024), which states that a test instrument can be considered viable if its validity score falls between 80% and 100%.

The validity of psychometrics and content is used to find out how valid the scientific reasoning test instrument that will be given to students is. The granting of the validity of psychometrics and content is related to the validity of the material used on the validity of the content and the use of educational measurements on the validity of psychometrics. The explanation related to the validity of the content and psychometrics is also in line with research (Dewi & Haryani, 2022), a viable test instrument is a test instrument that goes through the validity process through experts. Factors that determine the validity of the test instrument include (1) the student's condition, (2) the educator's ability in evaluation, and (3) the student's adequate ability. Considering the learning aspect, the validity process goes a step further to guarantee that the test instruments are appropriate, appropriate, and beneficial to students. The formulation of test instruments to be more advanced in accordance with their intended use, validity is the center of attention (Ruiz Mendoza et.al., 2024).

Differentiated learning is measured by the test tool's viability. Education that is tailored to each student's interests is implemented through differentiated learning evaluations. The assessment's goal is to give decision-makers in junior high school grade VIII the data they need to choose between several learning options. This may enhance the lesson plan that pupils are following Assessment includes associations related to learning achievement in students (Setiawan et.al., 2024).

Construct Validity

To assess the construct validity of this study, 252 grade VIII junior high school students were administered the scientific reasoning test. Using SPSS 25 for the construct validity process and the R Studio program, the researchers modified the Rasch Model while analysing the scientific reasoning test. Numerous question items on the scientific reasoning test were found to be valid based on the study of the test results.

Dewi & Sudaryanto (2020:74) define construct validity as a type of assessment that measures how well a research instrument can capture a conceptual framework. Examination of the difficulty level of test items, item fit test, person fit statistic, and unidimensional test are some of the assessments conducted as part of the construct validity test. The outcomes of data analysis with the Rasch Model are as follows:

a. Test Item Difficulty Parameters

The article's criteria shed light on the limitations of the question items' degree of difficulty Pratama, (2020). In order to evaluate the results of the question items' difficulty level, certain criteria must be met. In particular, if the difficulty value is greater than 2, the question is classified as extremely difficult; if the value is less than 1, the question is classified as tough; if the value is greater than 1, the question is classified as easy; and if the value is greater than 2, the question is classified as very easy. Table 6 shows the outcomes of using the eRM software to determine the question items' degree of difficulty.

Table 6. Test Item Difficulty Level Parameters

Number	Value	Criterion	Number	Value	Criterion
1.	0,612	Keep	17	0,271	Keep
2.	-0,266	Easy	18	-0,392	Easy
3.	-0,611	Easy	19	0,634	Keep
4.	-0,760	Easy	20	-0,316	Easy
5.	0.036	Keep	21	0,251	Keep
6.	-0,993	Easy	22	-1,142	Easy
7.	0,188	Keep	23	-0,699	Easy
8.	1,017	Difficult	24	-0,526	Easy
9.	-0,099	Easy	25	0,434	Keep
10.	0,313	Keep	26	0,313	Keep
11.	1,638	Difficult	27	-0,169	Easy
12.	-1,066	Easy	28	0,271	Keep
13.	0,166	Keep	29	-0,008	Easy
14.	-0,933	Easy	30	0,768	Keep
15.	-0,008	Easy	31	0,534	Keep
16.	0,251	Keep	32	0,354	Keep

Based on the measurement results, it was found that as many as 32 scientific reasoning tests given to students could be concluded that 2 (6.25%) were considered to meet the difficult criteria, 15 (46.87%) questions were considered to meet the moderate criteria, and 15 (46.87%) questions were considered to meet the easy criteria. This indicates that the scientific reasoning exam is not deemed suitable for use as a test for junior high school students in grade VIII. This is consistent with the findings of Purba & Silaban (2024) which explains that questions can be said to be good and ideal if they meet the proportion of difficulty levels of test items which include 25% difficult, 50% moderate and 25% easy.

It is crucial for test developers to make sure their instruments measure students' application of problem-solving techniques and skills in addition to their factual knowledge, and that they align with educational standards and the relevant curriculum. (Hulwani et al., 2024). This is consistent with the study Asbupel et.al., (2024) which revealed that students who have good cognitive skills as seen from the results of student tests, do not determine that students are also good in the psychomotoric and affective fields. such that creating a student assessment is required.

b. Item Fit Test

Construct validity in the content area was measured using the item fit test. A test is considered fit if the p-value > 0.01 according to the item fit criteria. Questions with a p-value above 0.01 are questions 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 27, 28, 29, 31, and 32, according to the results of the item fit analysis. Question numbers 10, 23, 24, and 26 had a p-value < 0.01. Each component in the level of scientific reasoning has different indications. The following are the results of fit or not fit test items per indicator in table 7

Table 7. Item Fit Results per Indicator

Ladder	Indicators	Question	P-Value	Information
Beginning	Classification	1	0,702	Suitable
		2	0,786	Suitable
	Conceptualization	3	0,906	Suitable
		4	0,925	Suitable
	Conclusion	5	0,799	Suitable
		6	0,655	Suitable
	Contextualization	7	0,981	Suitable
		8	0,478	Suitable
	Generalization			

Ladder	Indicators	Question	P-Value	Information
Basis	Order	9	0,267	Suitable
		10	0,002	Not Suit
		11	0,088	Suitable
	Problem Formulation	12	0,787	Suitable
		13	0,967	Suitable
	Estimation	14	1,000	Suitable
		15	0,978	Suitable
	Explanation	16	0,882	Suitable
		17	0,853	Suitable
	Predictions	18	0,982	Suitable
		19	0,721	Suitable
		20	0,956	Suitable
		21	0,941	Suitable
Intermediate	Application	22	0,991	Suitable
	Information	23	0,000	Not Suit
	Relationship	24	0,000	Not Suit
	Explained	25	0,929	Suitable
	Quantitative	26	0,001	Not Suit
	Understanding	27	0,747	Suitable
	Combinatorial	28	0,794	Suitable
	Reasoning	29	0,094	Suitable
	Applying Theory	30	0,027	Suitable
	Clearly Defining	31	0,216	Suitable
	Research Problems	32	0,010	Suitable

Out of 32 test items, 4 (12.5%) are deemed atypical or not fit to be utilized as an initial assessment of differentiated learning, according to the results of the item fit test that was conducted. About 28 (87.5%) test items are deemed fit or normal and can be utilized as a starting point for evaluating differentiated learning. Each indicator is represented by a question in the scientific reasoning test findings obtained from the item fit test. This is in line with the research of Mustaqim & Sulisti (2024) which explains that to make a test instrument, indicators are needed that are sufficient for all components. This is also in line with the research of Tesio et.al., 2024 which argues that the exam tool must be adjusted to indicators that are in accordance with the scientific reasoning test and contain items taken from the student's learning experience.

c. Person Fit Statistic

In the substantive field, the construct validity test is measured using the person fit statistical test. If the p-value > 0.01 on the scientific reasoning test, then the response can be said to be fit or there are no students who give deviant responses. The results of the person fit statistic analysis were conducted with the eRm software package R Studio. One student did not fit the model, and 251 students fit the model, based on the findings of the person fit statistic analysis conducted on 252 students who completed the scientific reasoning test. A person fit statistical study revealed that 99.60% of junior high school students in class VIII had the right values and 0.40% did not. This means that there is 1 student who deviates in answering the scientific reasoning test. This may be brought on by students' poor comprehension on scientific reasoning assessments, hence raising students' level of comprehension of scientific reasoning requires good instruction. This is consistent with the findings of Timu et al. (2024) who clarified

that proper teaching strategies are required to accommodate different levels of student understanding in relation to kids who have weaknesses in scientific reasoning. This opinion is in line with research by Rao & Elias-Medina (2024) which explains that special training intended to develop student understanding and engagement is one way to improve student understanding. Practice in developing students' understanding of scientific reasoning is by practicing reading and solving a problem in the scientific world.

d. Unidimensional test

In the structural field, unidimensional tests are used to measure construct validity. The scatter plot generated by the SPSS 25 software programme shows the scoring standards for the unidimensional test. The unidimensional test consists of approximately 32 questions on scientific reasoning, with 13 questions covering the rudimentary level, 10 questions covering the basic level, and 9 questions covering the intermediate level. The factor analysis test conducted on SPSS 25 application produced the results of the unidimensionality test. Table 8. total variance explained below presents the results of the factor analysis conducted on the scientific reasoning test given to students in grade VIII.

Table 8. Total Variance Explained

Element	Initial Eigenvalues		
	Total	% of Variance	Cumulative%
1	5,209	16,278	16,278
2	2,607	8,146	24,424
3	1,903	5,946	30,370
4	1,787	5,584	35,954
5	1,523	4,761	40,715
6	1,296	4,051	44,765
7	1,282	4,007	48,772
8	1,206	3,769	52,541
9	1,130	3,530	56,071
10	1,014	3,170	59,241
11	1,002	3,131	62,372

11 of the 32 components have a very dominant nature or Eigen value above 1.00, according to the table above which describes the results of the Eigen value obtained in the SPSS 25 factor analysis test for unidimensional analysis. The other 21 components are not dominant or have an Eigen value below 1.00. 34.36% of the components have dominant properties, while 65.64% of the components have non-dominant properties, as shown by the percentages obtained.

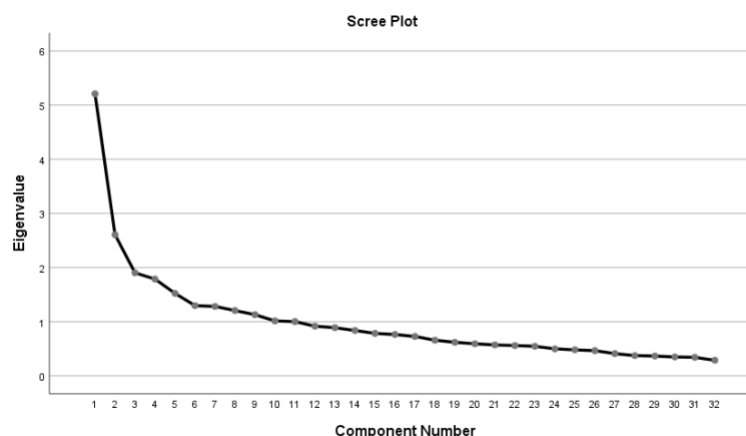


Figure 1. Scree Plot Factor Analysis

From the scatter plot above, it is clear that components 10 and 11 have the most prominent results, indicating that the unidimensionality test can be passed. The results of the unidimensionality test above explain that the rudimentary and basic levels of scientific reasoning are the most prevalent in the scientific reasoning test as an initial assessment of differentiated learning. It can be concluded that the scientific reasoning test is unidimensional. This is in line with Lestari (2023) research which explains that if there is one dominant component, the unidimensional test can be fulfilled. This is in line with Cho (2024) research which explains that unidimensional tests show one for all. Based on the assumption that the test items should measure the same combined abilities.

The validity of scientific reasoning exams, which can be used to evaluate differentiated science instruction, is a novel aspect of this study. Researchers also understand that the results of research that are applied directly have limitations, such as the use of 3 levels according to Wenning and Vierya including early, basic, and intermediate levels (Nikula et al., 2024). The limitations of the research were carried out in accordance with the school level used as the object of research, so that students could understand the scientific reasoning test easily and not too broadly.

Another limitation is seen from the material used for scientific reasoning tests. The material used only uses natural science subject material starting from chapter 1 to chapter 4 in semester I of the 2023/2024 school year.

Conclusion

Based on the research findings, it can be said that by using Rasch modeling and R studio software, it is possible to measure the validity of scientific reasoning assessments as a tool to evaluate different science teaching. Content, psychometric, and construct validity testing were included in the measurement. The findings show that the scientific reasoning assessment is reliable and practical to use as an evaluation tool for different science teaching, after passing the content and psychometric validity tests given by two professionals in the field. The construct validity test includes item fit test, person fit statistics, and unidimensionality test. The item fit test found 4 questions (12.5%) that were not suitable and could not be used for differentiated science learning assessment and 28 questions (87.5%) were suitable and could be used as a

differentiated science learning assessment. The statistical test of person fit on construct validity resulted in 1 out of 252 students who took the scientific reasoning test having a poor understanding of the scientific reasoning test. This is because 1 student who took the test did not fit. The unidimensional test gives the result that the scientific reasoning test is unidimensional because there is one element that is fulfilled. The findings of this study in the academic field can be used as an assessment in differentiated learning of science that can be given to students in grade VIII junior high school.

Credit Authorship Contribution Statement

Afida Nurul Sabilla: Conceptualization, Methodology, Visualization, Formal analysis, Writing-original draft, Project administration, Leader of data collection. **Purwo Susongko:** Conceptualization, Methodology, Software, Formal analysis, Supervision. **Muriani Nur Hayati:** Conceptualization, Writing -review & editing, Assistant of Project administration.

References

- Asbupel, F., Angela, L., & Alfian, M. (2024). *Development of Leaner Attitude Assessment Instruments in Chemistry Learning with Peer-Assessment Techniques*. *Journal of Education, Science and Technology*, 3(1), 248. <https://doi.org/10.47233/jpst.v3i1.1664>
- Ashrof, M., Zakaria, A., & Author, C. (2024). *Content Validity of Building Information Modelling (BIM) Assessment Criteria for Framework Formation for*. 11(1), 293. <https://doi.org/10.24191/myse.v11i1.Content>
- Cho, U. H. (2024). *Impact of Multidimensionality on unidimensional IRT Linking and Equating Methods*. University of North Carolina at Greensboro, 2.
- Dewi, A. M., & Haryani, S. (2022). *Development of Three-Tier Test Instrument Based on Scientific Literacy in Electrolyte and Non-Electrolyte Solution Topics*. *Thabiea : Journal of Natural Science Teaching*, 5(1), 1. <https://doi.org/10.21043/thabiea.v5i1.11634>
- Fauzi, A. A., Susongko, P., & ... (2022). *Critical Thinking Ability Test on Science Learning in Junior High School Based on Rasch Model*. *PSEJ (Pancasakti Science* <https://www.scienceedujournal.org/index.php/PSEJ/article/view/146>
- Hulwani, A. Z., Yuhana, Y., Studi, P., Pendidikan, M., Sultan, U., & Tirtayasa, A. (2024). *Development of Mathematics Test Instruments Based on Higher Order Thinking Skills (Hots) for Junior High School Students*. *Scientific Journal of Mathematics Education, Mathematics, and Statistics*, 5(2), 1075.
- Kusuma, M., Susongko, P., & Arfiani, Y. (2023). *Three-Dimensional Scientific Literacy Test*. Publishing Agency of Pancasakti University Tegal.
- Lestari, N. A. (2023). *Development of Science Knowledge Test with the Theme of Light and Optical Devices in Science Lessons in Junior High School*. In *Pancasakti Mandiri Group*. <http://dx.doi.org/10.1016/j.pbi.201>
- Marsela Yulianti, Divana Leli Anggraini, Siti Nurfaizah, & Anjani Putri Belawati Pandiangan. (2022). *The Role of Teachers in Developing Merdeka Curriculum*. *Journal of Education and Social Sciences*, 1(3), 293–294. <https://doi.org/10.58540/jipsi.v1i3.53>

- Mustaqim, & Sulisti, H. (2024). *Item Analysis of Specialisation Mathematics PAS Questions: Distinguishing Power, Level of Difficulty, and Exception Quality*. *Al-'Adad: Journal of Tadris Mathematics*, 3, 45. <https://doi.org/https://doi.org/10.24260/add.v3i1.3011>
- Natanael, Y., Fahmi, I., & Mulyaningsih, D. U. (2023). *Evaluating Psychometric Properties of Raven's Coloured Progressive Matrices Test in Indonesian Sample using the Rasch Model*. *Journal of Indonesian Psychological and Educational Measurement*, 12(2), 98. <https://doi.org/10.15408/jp3i.v12i2.27838>
- Nikula, T., Sobhy, N. N.-, Minardi, S., Glle, T., Yalm, Œ., Duman, S. K., Bozbıyık, M., Džanić, N. D., Ellison, M., Gmez, E., Hanušová, S., Kováčiková, E., Käntä, L., Lin, A., Llinares, A., Lo, Y. Y., Lopriore, L., Meyer, O., Morton, T., ... Źmavc, J. (2024). Towards an Initial Operationalisation of Disciplinary Literacies. *European Cooperation in Science & Technology*, 1–2. <http://urn.fi/URN:NBN:fi:jyu-202404243062>
- Nofianti, E., Nurhidayanti, A., Handayani, N. A., Rosana, D., & Wilujeng, I. (2022). *Critical Thinking Profile of Junior High School Learners on Human Excretory System Material*. *Indonesian Journal of Science Education*, 10(3), 479–491. <https://doi.org/10.24815/jpsi.v10i3.23877>
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/https://doi.org/10.1787/53f23881-en>.
- Purba, R., & Silaban, A. (2024). *The Effect of Big Book Media Use on Science Learning Outcomes of Ecosystem Theme for Grade V Students of SDN INPRES Perumnas 2 Waena*. 4(2), 210.
- Rao, A. R., & Elias-Medina, A. (2024). *Designing an Internet of Things Laboratory to Improve Student Understanding of Secure Iot Systems*. *Internet of Things and Cyber-Physical Systems*, 4, 156. <https://doi.org/10.1016/j.iotcps.2023.10.002>
- Riyadi, A., & Susongko, P. (2024). *Science Literacy Assessment Model in Elementary School Learners with Rasch Model Application*. 01(3), 3044–3054.
- Ruiz Mendoza, K. K., Pedroza Zniga, L. H., & Lpez Garcıa, A. Y. (2024). *Validity in Evaluation: Where is the Argument-Based Approach Heading? Sapienza (International Juournal of Interdisciplinary Studies)*, 5(3), 1–16. <https://doi.org/10.51798/sijis.v5i3.792>
- Setiawan, R., Wagiran, & Alsamiri, Y. (2024). *Construction of an Instrument for Evaluating The Teaching Process in Higher Education : Content and Construct Validity*. *REiD (Research and Evaluation in Education)*, 10(1), 50–63. <https://doi.org/https://doi.org/10.21831/reid.v10i1.63483>
- Susongko, P., Ratu, I. M., & Hayati, M. N. (2021). *A High-Level Thinking Ability Test to Measure Environmental Literacy for Junior High School Students with the Application of the Rasc Model*. *PSEJ (Pancasakti Science Education Journal)*, 6(2), 93–101. <https://doi.org/10.24905/psej.v6i2.127>
- Syahbana, A., Asbari, M., Anggitia, V., & Andre, H. (2024). *Education Revolution: Analysis of the Independent Curriculum as an Educational Innovation*. *Journal Of Information Systems and Management (JISMA)*, 3(2), 28. <https://jisma.org/index.php/jisma/article/view/935>

- Tesio, L., Caronni, A., Simone, A., Kumbhare, D., & Scarano, S. (2024). *Interpreting Results from Rasch Analysis 2. Advanced Model Applications and the Data-Model Fit Assessment. Disability and Rehabilitation*, 46(3), 604–617. <https://doi.org/10.1080/09638288.2023.2169772>
- Timu, D., Nahak, K. E. N., Soinbala, H. V., Tasuib, M., Lesik, N. M., & Koy, R. (2024). *Analysis of the Application of Truth Or Dare Media to Improve Learners' Learning Outcomes in Grade II Primary School. Indonesian Journal of Education and Learning*, 4, 528. <https://doi.org/https://doi.org/10.53299/jppi.v4.i2.542>
- Utami, L., Festiyed, Dian Purnama Ilahi, Arista Ratih, Elvi yenti, & Lazulva. (2024). *Aiken Index Analysis to Determint the Content Validity of the Scinetific Habbits of Mind Instrument. Journal of Research and Education Chemistry*, 6(1), 59. [https://doi.org/10.25299/jrec.2024.vol6\(1\).17430](https://doi.org/10.25299/jrec.2024.vol6(1).17430)
- Yucel, R. (2023). *Enhancing Powerful Knowledge in Undergraduate Science Curriculum for Social Good. Teaching in Higher Education*, 35–50. <https://doi.org/10.1080/13562517.2023.2215696>