



ASSESSMENT OF SCIENTIFIC REASONING-COMMUNICATION SKILLS (SR-CS) TEST ON WORK AND ENERGY CONCEPT: DEVELOPMENT, CONTENT VALIDITY AND RASCH MODEL ANALYSIS

Fajar Fanika^{1*}, Selly Feranie², Parlindungan Sinaga³

Departemen Pendidikan Fisika, Universitas Pendidikan Indonesia, Indonesia

fajarfanika@gmail.com

INFO ARTIKEL

ABSTRACT

Keywords: Scientific reasoning skills; scientific communication skill; test instrument; rasch model.

Studies indicate that scientific reasoning assessments are still rarely integrated as higher-level assessments and new science education standards in school instruction. This study aims to develop a scientific reasoning skills test instrument that is integrated with a scientific communication skills test instrument (SR-CS test) on work and energy concept. This study used the ADDIE procedure. The research procedure consists of five stages including analyzing, designing, developing, implementing, and evaluating. The initial draft of SR-CS test consisted of 14 multiple choice scientific reasoning skills (SRS) questions and 8 open-ended scientific communication skills (SCS) questions. The results of the expert judgement were analyzed using the content validity index (CVI) and obtained a value of 0,94 (very suitable) for the SRS instrument and 0.97 (very suitable) for the SCS instrument. After being revised based on the expert suggestions, the test instrument was tested on 25 students (15 girls, 10 boys) aged 16-17 years. The trial data were analyzed using the Rasch Model to obtain item fit (validity), reliability, distinction level, and difficulty level. The results show that 14 questions of SRS instrument have item validity and 8 questions of SCS instrument have item validity. Besides that, the item reliability of the SRS and SCS test instrument is 0.79 and 0.91, respectively. Meanwhile, the person reliability is 0.82 (SRS) and 0.91 (SCS). Therefore, the SR-CS test is valid and reliable so that it can be used to measure scientific reasoning skills and scientific communication skills of students in further research.

Introduction

Scientific reasoning is one of the 21st-century skills that students must learn. One of the science process skills that must be mastered is reasoning, which is necessary for the planning and interpretation of experimental outcomes (Coleman et al., 2015). Critical thinking skills even frequently include reasoning (Birgili, 2015; Tiruneh et al., 2017). Critical thinking skills and other higher-level cognitive skills cannot fully develop without powerful reasoning skills.

Previous research has provided information on Indonesian high school students' capacity for scientific thinking, particularly in physics classes. Research on high school students' capacity for scientific reasoning reveals that students' scientific reasoning skill is low

(Ayuningtyas & Pramudya, 2019; Khoirina & Cari, 2018; laela Ermaya & Mashuri, 2018). According to early research, 49% of participants are thought to be at the Concrete Operational level, 49% to be at the Early Transitional level, and 2% to be at the Late Transitional level. The assessment of students' scientific reasoning on the subject, regardless of the learning method employed, reveals that students' scientific reasoning level falls into the low category.

Scientific reasoning skills can be trained with good communication efforts from students in conveying their arguments. A person can communicate well if they have good arguments. This can be realized if it is familiarized with the atmosphere of discussion, exchanging opinions in conveying ideas or scientific arguments. This is also what is emphasized during the learning process in the classroom so that there is always social interaction between students and students, students and teachers, and students and the environment in conveying their thinking process. Knowledge that has been formed by students actively, not only passively received from the teacher but also must communicate their thinking process both orally and in writing (Fadly, 2014). Training science communication skills to students makes students able to express the science ideas they have. However, a report prepared by (McInnis et al., 2000) for the Australian Council of Deans of Science said that communication skills (oral, interpersonal, and written) consistently did not meet the predetermined criteria.

Based on this, researchers will develop instruments that can measure students' scientific reasoning skills and scientific communication skills. Assessment is the activity of interpreting measurement data based on certain criteria or rules (Widoyoko, 2017). A good assessment instrument must meet reliability criteria. Reliability criteria include validity, reliability, difficulty level, and differentiation/discrimination. To test the reliability/quality of the learning outcomes assessment instrument, an instrument trial was conducted. Instrument testing can be done internally and externally. Internal trials are conducted to experts to see content and construct validity, as well as grammar (Widoyoko, 2012). Expert judgment is needed to consider the structure of the instrument is correct or not with the structure or scientific arrangement used in compiling the instrument. To stabilize the validity of the assessment instrument, it is necessary to conduct an external trial, namely a field trial (Arikunto, 2017). Field trials can be carried out on subjects who are similar / equivalent to the subjects to be assessed. Then after the trial, it is necessary to analyze the results of the trial, including analyzing the validity, reliability, difficulty level, and differentiation analysis of the questions.

Analysis of learning outcomes assessment items from field trials / empirical trials can be done in a classic way known as classical test theory (CTT) or modern with item response theory (IRT). Classical test theory is based on the observed score which is the sum of the true score and the measurement error score. The quality of the items is determined by the level of difficulty and distinguishing power (Hardianti et al., 2021), but the characteristics of the items are inconsistent depending on the ability of the students (Erfan et al., 2020). The modern method with the Rasch Model, which was first proposed by Dr. Georg Rasch, a mathematician from Denmark, is here to overcome the weaknesses of the classical method. The Rasch Model uses raw scores in different ways to produce a measurement scale with the same interval, so that it can provide accurate information about test takers and question quality (Sumintono & Widhiarso, 2015).

This article tries to describe the results of analyzing the quality of test instruments in aspects of validity, reliability, difficulty level, and differentiability through Rasch Model analysis. The test instrument analyzed in this study is a scientific reasoning skill measurement test instrument integrated with scientific communication skills (SR-SC).

Research Methods

Research and Development (R&D) is the type of research being done to create the physics assessment, which is the process for developing and validating product (Ardiyanto &

Fajaruddin, 2019). The five stages of the ADDIE procedure, analyzing, planning, developing, implementing, and evaluating were applied in this investigation. The SR-CS test is a predetermined assessment tool created by the researcher to gauge scientific reasoning and communication abilities. Five experts were given a content validity exam by the researcher. In addition, the instrument was refined in response to professional advice before being given to the pupils. Following deployment, Ministep 9.3.1.0 software was used to assess each student's reaction using Rasch analysis.

The participants of this research are the 11th-grade public high school students who are locating in Balaraja, Tangerang Regency, Indonesia. This A-accreditate public high school has been established since 1995. Most of the students in Tangerang Regency are a mix of Sundanese and Javanese ethnic. The participants were 25 students (15 females, 10 males) aged 16-17 years.

The scientific reasoning skills (SRS) test, which had 14 multiple-choice questions, and the scientific communication skills (SCS) test, which had 8 open-ended questions, were the test instruments that were designed. The SRS test instrument includes seven aspects, there are control of variables, probability reasoning, correlational reasoning, hypothetical-deductive reasoning, deductive reasoning, inductive reasoning, dan causal reasoning. The SCS test instrument consists of information representation and scientific reading. In expansion, there's an expert assessment rubric for content validity

Five experts assess the completed test instrument. The content validity index (CVI) developed by Lynn (1986), was used to analyze the expert assessment results. After that, the test instrument was enhanced according to of the experts' recommendations and comments. 25 students were assigned the test after it had been revised. The Rasch Model was then used to analyze the test results.

The test instruments' item fit, reliability, difficulty level, and distinction level are all analyzed. The outfit z-standard ($-2.0 < ZSTD < 2.0$), outfit means-square ($0.5 < MNSQ < 1.5$), and point measure correlation ($0.4 < \text{Pt Measure Corr} < 0.85$) can all be used to determine the item fit level. If an item satisfies the requirements for MNSQ, ZSTD, and Pt Measure Corr scores, it is very suitable. Items which satisfy at least one of the three scores, however, may still be approved (Rachmadtullah, 2020).

The correlation between each item's difficulty and the overall test difficulty is displayed by the point measure correlation. When the value of 1 is reached, all participants with high ability answered correctly, whereas all pupils with poor ability answered incorrectly. A value of 0 on the other hand, indicates no correlation between the item responds. In other words, a student's response does not necessarily reflect their ability (Smiley, 2015).

The item's degree of difficulty also reveals an instrument's qualities. The standard deviation (SD) value that was determined by the analysis served as the basis for this analysis's standard value. The SD number shows that the logit size in item difficulty is fairly distributed. The item difficulty was categorized into five groups: very difficult ($\text{JMLE Measure} \geq \text{mean logit} + 2\text{SD}$), difficult ($\text{mean logit} + 2\text{SD} > \text{JMLE Measure} \geq 1\text{SD}$), moderate ($1\text{SD} > \text{JMLE Measure} \geq \text{mean logit}$), easy ($\text{mean logit} > \text{JMLE Measure} \geq -1\text{SD}$), and very easy ($\text{JMLE Measure} < -1\text{SD}$) (Soeharto & Csapó, 2022).

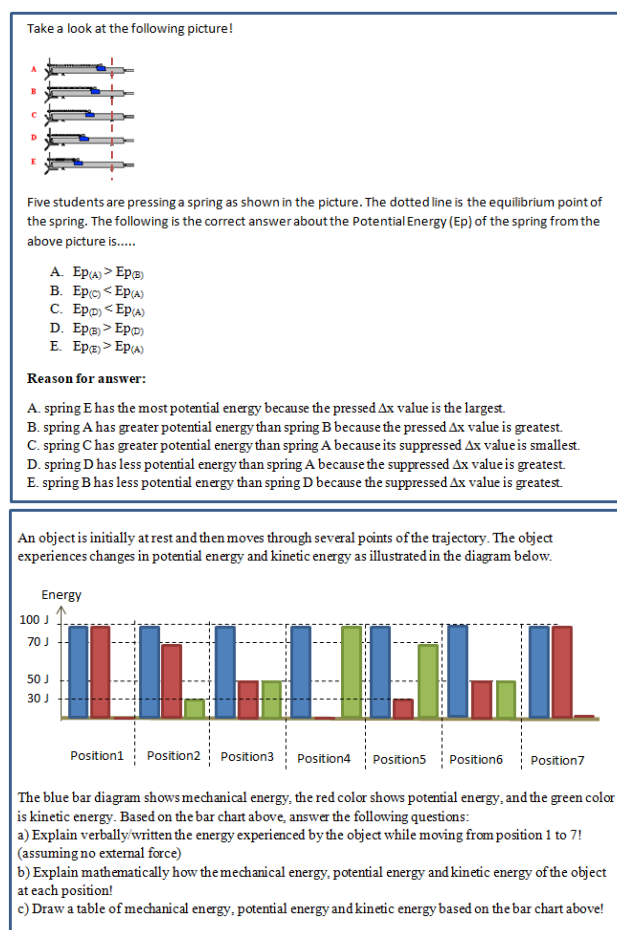
Results and Discussion

Analyzing

The first step of the research is a review of the literature on scientific communication skills (SCS) and scientific reasoning skills (SRS). The researcher discovered how the test instrument was created after reviewing the literature. When creating the SRS test instrument, the researcher refers to the Lawson Classroom Test Scientific Reasoning (LCTSR) rubric. In the meantime, scientific reading and information representation are being considered in the creation of the SCS test instrument. The physics content for which the SRS and SCS test instruments will be created has been determined to be work and energy.

Designing

At this point, the researchers created SRS and SCS test instrument indicators. The aspects noticed serve as the basis for developing the indicators. Control of variables, probabilistic reasoning, correlational reasoning, hypothetical-deductive reasoning, deductive reasoning, inductive reasoning, and causal reasoning are some of the components of science reasoning skills that are observed. In the meantime, scientific reading and information representation were considered components of scientific communication skills. In addition, indicators are used to create the SRS and SCS items. There are 14 multiple-choice questions on the SRS test, and 8 open-ended questions on the SCS test. Figure 1 shows an example of an SRS and SCS item.



Developing

During the development phase, the researcher made enhancements to the SRS and SCS test instruments, which were created using feedback and recommendations from experts. The

content validity index (CVI) is used to determine expert judgment in order to assess the validity of the instrument. The I-CVI and S-CVI are the two components that form the overall CVI score assessment. A question item's validity score is displayed using I-CVI. Conversely, S-CVI displays an instrument's overall validity. For 3-5 experts, the I-CVI value should ideally be 1. However, the S-CVI value does not go below 0.90. According to Lynn (1995), the I-CVI value should be at least 0.78. The overall CVI scores for the SRS and SCS test instruments were determined by an expert evaluation, and the results were 0.94 and 0.97, respectively.

Implementing

Test subjects were 25 public high school students in their 11th grade, using the recently revised SRS and SCS test instruments. SR-CS test instrument testing was carried out at the beginning of the year, in March. Students can use a paper-based testing system to access the SRS and SCS test instruments. The time allocation for doing the SRS and SCS tests (SR-CS test) is 90 minutes.

Evaluating

To ascertain item validity, reliability, distinction level, and difficulty level, each student's results were analyzed using ministep software. Table 3 shows the results for the MNSQ outfit, ZSTD outfit, and Pt-Measure Corr. of scientific reasoning skills items.

Table 4. The interpretation of Scientific Reasoning Skills Item Fit and Distinction Level

SRS Aspect	Item Number	Outfit		Pt. Measure Corr.	Item Fit Interpretation	Distinction level Interpretation
		MNSQ	ZSTD			
Inductive	R1	1.66	1.65	0.74	Accepted	Very Good
Correlational	R2	0.42	-0.72	0.74	Accepted	Very Good
Probability	R3	1.30	0.67	0.77	Accepted	Very Good
Hypothetical-deductive	R4	1.27	0.58	0.74	Accepted	Very Good
Control of Variables	R5	1.14	0.48	0.46	Accepted	Very Good
Correlational	R6	1.17	0.46	0.74	Accepted	Very Good
Causal	R7	1.11	0.50	0.68	Accepted	Very Good
Probability	R8	1.06	0.27	0.77	Accepted	Very Good
Hypothetical-deductive	R9	0.91	0.18	0.72	Accepted	Very Good
Inductive	R10	0.79	-0.43	0.75	Accepted	Very Good
Deductive	R11	0.78	-0.31	0.77	Accepted	Very Good
Hypothetical-deductive	R12	0.79	-0.73	0.69	Accepted	Very Good
Causal	R13	0.74	-0.80	0.72	Accepted	Very Good
Probability	R14	0.68	-1.24	0.57	Accepted	Very Good

According to the table, 12 items satisfy all three criteria, indicating that they are fit. Even though two of the goods only satisfy the ZSTD and Pt Measure Corr requirements, they can still be considered approved (Sumintono & Widhiarso, 2015). Items R1 and R2 indicate that the MNSQ output value surpasses the 1.5 score limit, while item R5 indicates that the Pt. Measure Corr value is less than 0.5. The Pt. Measure Corr value for the remaining items is in the vicinity of one. The degree of discriminating power increases with proximity to one. Table 4 shows the results for the MNSQ outfit, ZSTD outfit, and PT-Measure Corr. of Scientific Communication Skills items.

Table 4. The interpretation of Scientific Communication Skills Item Fit and Distinction Level

SCS Aspect	Item Number	Outfit		Pt. Measure Corr.	Item Fit Interpretation	Distinction level Interpretation
		MNSQ	ZSTD			
Information representation	C1	1.36	1.17	0.50	Accepted	Very Good
	C2	1.35	1.29	0.53	Accepted	Very Good
	C3	0.94	-0.12	0.51	Accepted	Very Good
	C4	0.97	-0.03	0.54	Accepted	Very Good
	C5	0.90	-0.30	0.54	Accepted	Very Good
Scientific reading	C6	0.90	-0.32	0.54	Accepted	Very Good
	C7	0.72	-1.14	0.54	Accepted	Very Good
	C8	0.60	-1.57	0.54	Accepted	Very Good

According to Table 4, all items match the criteria for MNSQ, ZSTD, and Pt-Measure Corr. values, indicating that the items are suitable and can be used in research aimed at detecting students' scientific communication skills. While the MNSQ outfit value of six items is close to 1.00, indicating that the items have a fair level of consistency, the range of PT-Measure Corr. values for eight items is close to between 0.50 and 0.54; this suggests that the items have impoverished differentiating strength.

In addition, Ministep Software will generate Cronbach's Alpha (α), item reliability, and person reliability. Table 5 displays the SRS and SCS test instrument summary statistics.

Table 5. Summary Statistic of Measured Item and Person for each SRS and SCS Test Instrument

	SRS		SCS	
	Item	Person	Item	Person
N	14	25	8	25
Mean	41.6	23.3	74.1	23.7
Mean Measure	0.00	-2.00	0.00	1.04
P.SD	1.36	1.26	0.49	0.91
Mean Outfit MNSQ	0.99	0.93	0.97	0.97
Mean Outfit ZSTD	0.04	0.01	-0.13	-0.07
Reliability	0.89	0.81	0.59	0.61

Assessment of Scientific Reasoning-Communication Skills (SR-CS) Test on Work and Energy Concept: Development, Content Validity and Rasch Model Analysis

Cronbach's alpha	0.92	0.64
---------------------	------	------

For SRS items, the Rasch model yields good person and item reliability scores of 0.89 and 0.81, respectively. In contrast, the SCS test instrument's person and item dependability values were 0.59 and 0.61, respectively, indicating low item reliability (Sumintono & Widhiarso, 2015). The results presented suggest that the science reasoning and scientific communication skills of students can be accurately assessed using the SRS and SCS test instruments. Students' earnestness about taking the SR-CS test is indicated by person reliability ratings that are not significantly different. Furthermore, the quality of the interaction between person and item as illustrated by the Cronbach Alpha value is 0.92 (excellent) for the SRS test instrument and 0.64 (weak) for the SCS test instrument. Furthermore, the results of the analysis of the level of difficulty on the SPS test instrument can be seen in Table 6.

Tabel 6. Difficulty Level of SRS Test Instruments

Entry Number	JMLE Measure	Difficulty Level Interpretation
4	1.73	Difficult
1	1.44	Difficult
11	1.44	Difficult
13	1.44	Difficult
2	0.71	Moderate
3	0.52	Moderate
9	0.34	Moderate
14	0.04	Moderate
8	-0.22	Easy
5	-0.54	Easy
10	-0.82	Easy
6	-0.90	Easy
7	-2.06	Very Easy
12	-3.12	Very Easy
Mean	0.00	
P.SD	1.36	

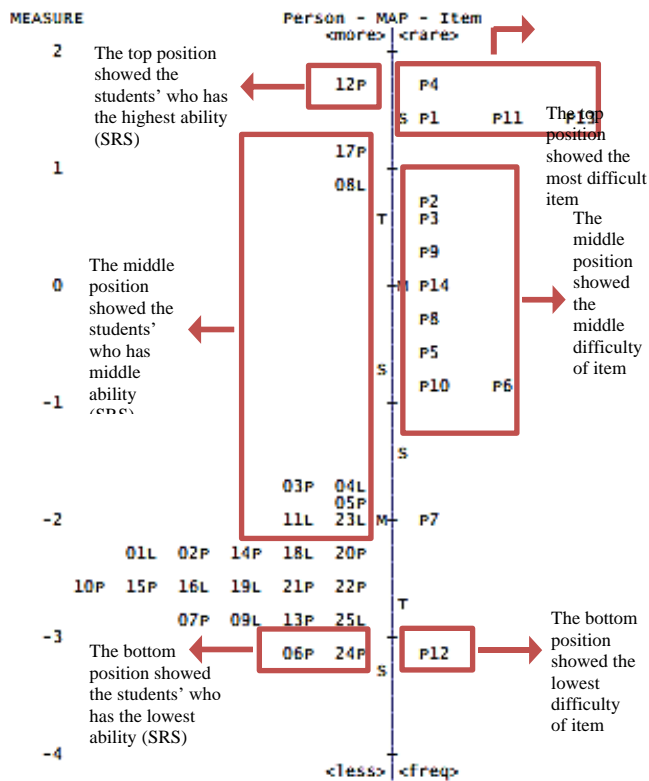
The JMLE Measure scores are shown in Table 6 from highest to lowest. The more challenging the item, and vice versa, the higher the JMLE Measure score. Three difficult items, four moderate items, four easy items, and two very easy items make up the SRS test instrument. In the meantime, Table 7 shows the findings of the difficulty level analysis on SCS test devices.

Tabel 6. Difficulty Level of SCS Test Instruments

Entry Number	JMLE Measure	Difficulty Level Interpretation
8	0.91	Difficult
1	0.19	Moderate
6	0.19	Moderate
7	0.19	Moderate
5	0.02	Moderate
4	-0.14	Easy
3	-0.58	Very Easy
2	-0.77	Very Easy
Mean	0.00	
P.SD	0.49	

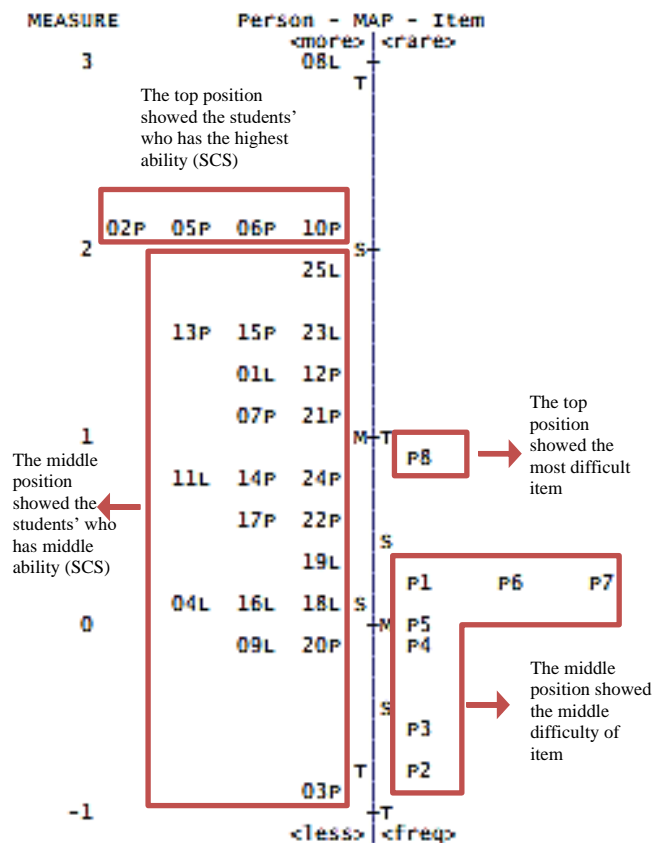
The JMLE Measure score is displayed in Table 7 from highest to lowest. There are one easy, two very easy, four moderate, and one difficult things.

The distribution of students' abilities and the degree of difficulty of items on the same scale can be described by the Rasch model analysis (Sumintono & Widhiarso, 2015). The distribution is shown on a map known as the Wright map. The outcome of the Wright map is displayed in Figure 2.



(a)

Assessment of Scientific Reasoning-Communication Skills (SR-CS) Test on Work and Energy Concept: Development, Content Validity and Rasch Model Analysis



(b)

Figure 2 above illustrates some of the Wright maps of objects and people. The outline person appeared on the left, and the outline item displayed on the right. The participant's proficiency in scientific thinking and scientific communication is nearly medium, according to the plan.

The SRS and SCS test instruments produced can be deemed valid and reliable based on the study of item fit, reliability, distinction level, and difficulty level using the Rasch Model. Thus, for future research, the SRS and SCS test instrument (SR-CS test) can be utilized to assess scientific communication and reasoning abilities

Conclusion

The scientific reasoning skills test instrument designed had 14 multiple-choice questions. The SRS test instrument's distinction level fell into the very good range, and each item exhibited item fit validity. The person and item reliability of the SRS test instrument as a whole obtained a score of 0,81 and 0.89 in the good category. For the scientific communication skills test instrument that was developed, there were 8 open-ended questions, and all of questions had item fit. The distinction level of 8 items is in the very good category. The person and item reliability of the SCS test instrument as a whole obtained a score of 0.61 and 0.59 in the low category. The quality of interaction between the person and items illustrated by Cronbach Alpha value, the SRS test instrument scored 0.92 (very good) and the SCS instrument scored 0.64 (low). Therefore, the SR-CS test which consists of 14 multiple choice questions and 8

open-ended questions is valid and reliable so that it can be used to measure students' scientific reasoning skills and scientific communication skills in further research.

References

- Ardiyanto, H., & Fajaruddin, S. (2019). Tinjauan atas artikel penelitian dan pengembangan pendidikan di Jurnal Keolahragaan. *Jurnal Keolahragaan*, 7(1), 83–93.
- Arikunto, S. (2017). Pengembangan instrumen penelitian dan penilaian program. *Yogyakarta: Pustaka Pelajar*, 53.
- Ayuningtyas, W., & Pramudya, I. (2019). Students' responses to the test instruments on geometry reasoning ability in senior high school. *Journal of Physics: Conference Series*, 1265(1), 12015.
- Birgili, B. (2015). Creative and critical thinking skills in problem-based learning environments. *Journal of Gifted Education and Creativity*, 2(2), 71–80.
- Coleman, A. B., Lam, D. P., & Soowal, L. N. (2015). Correlation, necessity, and sufficiency: common errors in the scientific reasoning of undergraduate students for interpreting experiments. *Biochemistry and Molecular Biology Education*, 43(5), 305–315.
- Erfan, M., Maulyda, M. A., Ermiana, I., Hidayati, V. R., & Widodo, A. (2020). Validity and reliability of cognitive tests study and development of elementary curriculum using Rasch model. *Psychology, Evaluation, and Technology in Educational Research*, 3(1), 26–33.
- Hardianti, H., Liliawati, W., & Tayubi, Y. R. (2021). Karakteristik tes kemampuan berpikir kritis siswa SMA pada materi momentum dan impuls: Perbandingan classical theory test (CTT) dan model Rasch. *WaPFI (Wahana Pendidikan Fisika)*, 8(1), 21–28.
- Khoirina, M., & Cari, C. (2018). Identify students' scientific reasoning ability at senior high school. *Journal of Physics: Conference Series*, 1097(1), 12024.
- laela Ermaya, H. N., & Mashuri, A. (2018). Kinerja Perusahaan dan Struktur Kepemilikan: Dampak terhadap Pengungkapan Lingkungan. *Jurnal Kajian Akuntansi*, 2(2), 225–237.
- McInnis, C., Hartley, R., & Anderson, M. (2000). What did you do with your science degree. *A National Study of Employment Outcomes for Science Degree Holders 1990-2000*.
- Rachmadtullah, R. (2020). Critical Thinking Instrument Test (CTIT): Developing and analyzing Sundanese students' critical thinking skills on physics concepts using Rasch analysis. *International Journal of Psychosocial Rehabilitation*, 24(08).
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikasi.
- Tiruneh, D. T., De Cock, M., Weldeclassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15, 663–682.
- Widoyoko, E. P. (2012). *Teknik penyusunan instrumen penelitian*.
- Widoyoko, E. P. (2017). *Evaluasi program pelatihan*. Yogyakarta: Pustaka Pelajar.



© 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)