



Developing Real Picture Analysis (RPA) -Based Assessment Instruments to Measure Students' Higher-Order Thinking Skills

Inas Aulia Majid^{1*}; Sutarto²; Bea Hana Siswati³

^{1,2,3}Department of Science, Universitas Jember, Indonesia

^{1*}Corresponding Email: inas.aulia2@gmail.com

Article History:

Received: Sep 05, 2025

Revised: Oct 04, 2025

Accepted: Nov 07, 2025

Online First: Nov 27, 2025

Keywords:

4-D Model,

Assessment Instrument,

HOTS,

Real Picture Analysis,

Validity and Reliability.

Kata Kunci:

Analisis Gambar Nyata,

HOTS,

Instrumen Penilaian,

Model 4-D,

Validitas dan Reliabilitas.

Abstract: The purpose of this study is to describe the validity, reliability, discriminating power, and difficulty level of an RPA-based assessment instrument designed to measure students' higher-order thinking skills. Using the 4-D model: Define, Design, Development, and Disseminate the instrument, consisting of 30 questions, was tested on 240 Grade VIII students in several Junior High Schools in Jember during the odd semester of the 2024/2025 academic year. From the 25 essay questions tested, an average score of 81% was achieved, with all questions meeting high validity criteria, indicating that the instrument measured the intended constructs accurately. The instrument demonstrated high reliability, with a coefficient of 0.86, ensuring consistent and dependable results. Its discriminating power was strong, with 20 questions rated as very good and 5 as good in distinguishing student abilities, ensuring meaningful differences in performance. The distribution of question difficulty levels, namely 9 easy, 11 medium, and 5 hard provided a balanced challenge to assess a comprehensive range of student abilities. This study contributes to authentic assessment design by integrating real-world images to foster higher-order thinking skills and can serve as a benchmark for further research or application.

Abstrak: Tujuan dari penelitian ini adalah untuk mendeskripsikan validitas, reliabilitas, daya pembeda, dan tingkat kesulitan instrumen penilaian berbasis RPA yang dirancang untuk mengukur keterampilan berpikir tingkat tinggi siswa. Menggunakan model 4-D: Define, Design, Development, dan Disseminate, instrumen yang terdiri dari 30 pertanyaan tersebut diujicobakan kepada 240 siswa kelas VIII di beberapa SMP di Jember selama semester ganjil 2024/2025. Dari 25 pertanyaan esai yang diujikan, skor rata-rata 81% tercapai, dengan semua pertanyaan memenuhi kriteria validitas tinggi, yang menunjukkan instrumen tersebut mengukur konstruk yang dimaksud dengan tepat. Instrumen tersebut menunjukkan reliabilitas tinggi, dengan koefisien 0,86, memastikan hasil yang konsisten dan dapat diandalkan. Daya pembedanya kuat, dengan 20 pertanyaan dinilai sangat baik dan 5 baik dalam membedakan kemampuan siswa, memastikan perbedaan yang berarti dalam kinerja. Distribusi tingkat kesulitan pertanyaan, yaitu 9 mudah, 11 sedang, dan 5 sulit, memberikan tantangan yang seimbang untuk menilai berbagai kemampuan siswa yang komprehensif. Studi ini berkontribusi pada desain penilaian autentik dengan mengintegrasikan gambar dunia nyata untuk menumbuhkan keterampilan berpikir tingkat tinggi dan dapat digunakan sebagai tolok ukur untuk penelitian atau penerapan lebih lanjut.

How to cite:

Majid, I. A., Sutarto, S., & Siswati, B. H. (2025). Developing Real Picture Analysis-Based Assessment Instruments to Measure Students' Higher-Order Thinking Skills. *Edunesia : Jurnal Ilmiah Pendidikan*, 6(3), 1994-2014.

This is an open-access article under the CC-BY-NC-ND license



A. Introduction

Physics learning focuses on concepts that require a high level of understanding. Physics is a subject that can be quite challenging for students during the classroom learning process, making it less attractive (Meryastiti et al., 2022). Several junior high school students stated that the difficulty they experienced in learning Physics was when working on Physics problems. This is because Physics is often associated with formulas and can be difficult to understand, so students often memorize formulas without grasping the underlying concepts (Hidayati et al., 2021). Furthermore, Physics learning by teachers in schools only focuses on the mathematical aspects rather than the concepts and principles of Physics (Budiarso et al., 2020). Physics learning that is linked to everyday life serves as a means and source of learning, making it more meaningful for students (Kokkonen et al., 2022). Teachers should develop students' abilities to reason logically and abstractly, enabling them to present conceptual knowledge about real phenomena in the classroom (Pursitasari et al., 2022). Therefore, it is necessary to take corrective steps to change students' views and improve the quality of Physics learning.

In the 21st-century learning era, advances in science and technology require human resources to master a range of skills, including higher-order thinking skills. Higher-order thinking skills are defined as critical, logical, reflective, metacognitive, and creative thinking (Limbach & Waugh, 2022). The successful implementation of higher-order thinking skills is evident in the provision of valid explanations, informed decision-making, effective presentations, and high-quality products (Widyaningsih et al., 2021). Through the implementation of the Kurikulum Merdeka and the Profil Pelajar Pancasila, teachers are expected to be able to optimally integrate higher-order thinking skills and student character values (Tawil & Tampa, 2024). Based on the revised Bloom's Taxonomy, higher-order thinking skills include analyzing (C4), evaluating (C5), and creating (C6) (Desiriah & Setyarsih, 2021). When students are assessed on tasks with higher-order thinking characteristics, their thinking skills are honed (Julianingsih et al., 2013). The main characteristic of assessments with higher-order thinking characteristics is the presence of a stimulus. Stimuli can take the form of text, images, scenarios, tables, graphs, discourse, dialogue, videos, or problems (Murwantini, 2022). However, currently, there are no assessments specifically designed to train higher-order thinking skills, so students are still under-trained in working on questions that measure their higher-order thinking abilities (Amir et al., 2024)

In the 21st-century learning era, teachers need to understand and present assessment concepts that can optimize classroom learning. Physics assessment concepts can be represented verbally, in pictures, in equations, in diagrams, in tables, and in graphs (Nikat et al., 2021). Qualitative representations, such as pictures and diagrams, can help in understanding problems before solving them quantitatively (Aisyah & Sudarti, 2021). Visual representations such as diagrams and images significantly assist students in understanding abstract physics concepts, including force, work, and energy. Diagrams of levers and inclined planes are used as examples of effective visual representations in illustrating the

relationship between force, distance, and work (Lamanepa et al., 2022). Assessments represented through images can aid student understanding.

Images used in learning will be more meaningful if the concepts, relationships between concepts, and other elements contained in the image are analyzed or studied. Real Picture Analysis (RPA) is used to provide students with an understanding of science concepts and relate them to their surrounding environment, making science learning more meaningful and training students' critical thinking skills. RPA is defined as solving physics problems through real images or images that actually occur around students (Putri et al., 2023). Real picture media can be considered an alternative to improve student learning outcomes, as students may become more interested and enthusiastic in the learning process (Nikat et al., 2021). Physics learning accompanied by images is easier to accept and understand, making it more applicable to other relevant concepts (Idris et al., 2023). In previous research, RPA has been implemented in the form of a Teaching Module to improve students' critical thinking skills (Putri et al., 2023). Therefore, in this study, RPA was used as a medium and basis for compiling assessment instruments.

Although previous studies have successfully applied Real Picture Analysis (RPA) in the development of learning modules to enhance conceptual understanding and critical thinking skills, a significant gap remains in the literature regarding its application as the foundational basis for standardized assessment instruments specifically designed to measure Higher Order Thinking Skills (HOTS). Most existing research has focused on the experiential learning benefits of RPA; yet, little is known about the psychometric validation and systematic design of RPA-based instruments capable of reliably and validly assessing HOTS in physics. This gap underscores the necessity for the rigorous development and empirical evaluation of RPA-based assessment tools that can meet both educational and measurement standards.

This study addresses this gap by presenting the first attempt to design, validate, and implement RPA-based HOTS assessment instruments tailored to physics content, particularly on the topics of work and energy. By integrating real-world imagery into complex problem scenarios, this instrument not only tests students' analytical, evaluative, and creative abilities in physics but also provides a novel, authentic approach that bridges the gap between abstract scientific concepts and tangible experiences. Through thorough validity and reliability testing, this research offers an innovative contribution to physics education assessment and expands the potential applications of RPA in formal evaluation contexts.

The purpose of this study is to describe the validity, reliability, discriminatory power, and difficulty level of a Real Picture Analysis (RPA)-based assessment instrument in measuring eighth-grade students' higher-order thinking skills in the topic of work and energy. The expected benefits of this study include theoretically increasing knowledge regarding the development of Real Picture Analysis (RPA)-based assessment instruments to improve students' higher-order thinking skills. For teachers, it can provide a new alternative for measuring students' higher-order thinking skills. Furthermore, for future research, it can

serve as a source of information and a scientific reference for developing learning instruments to enhance students' higher-order thinking skills, particularly in the topic of Work and Energy.

B. Method

The type of research employed was Research and Development (R&D), with a research design based on the 4D development research model, which encompasses Define, Design, Develop, and Disseminate, as illustrated in Figure 1.

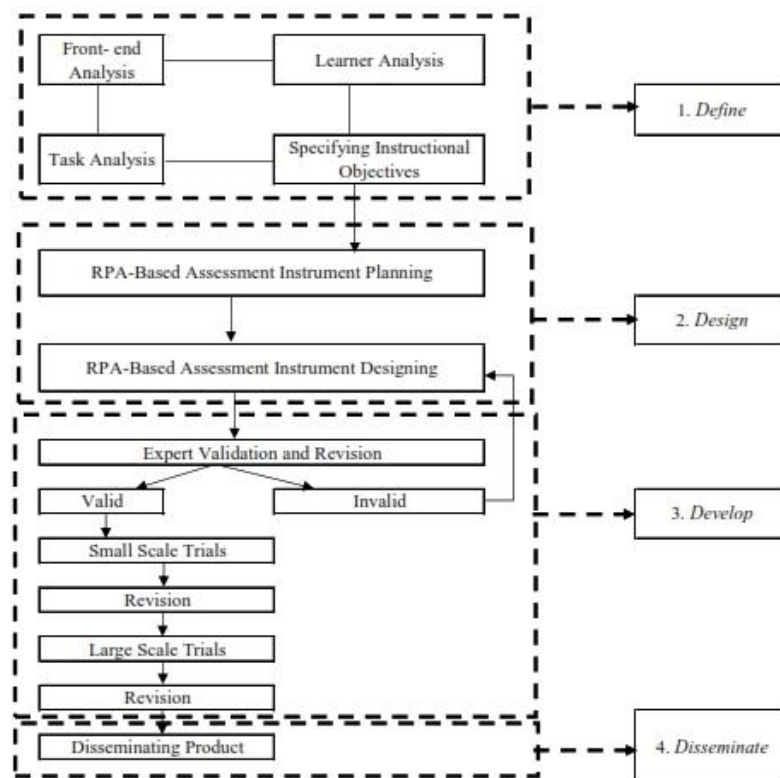


Figure 1. Research Model

The Define phase consists of needs analysis, student analysis, task analysis, and learning outcomes analysis. The needs analysis is conducted by first reviewing and collecting data on Physics learning assessment through various literature sources and distributing questionnaires to teachers. The purpose of the needs analysis is to identify factors that inhibit learning and assessment, particularly in measuring students' higher-order thinking skills. Questionnaires were distributed to several Physics teachers regarding the Physics learning assessment process, obstacles and challenges encountered during the learning process, the learning media used, and the strategies employed by teachers in conducting the assessment.

Student analysis involves assessing the competencies and actual conditions of students who will undergo the assessment process, including the competencies they already possess upon entering the course. Student analysis is conducted by reviewing and collecting

data on Physics learning assessment through various literature and distributing questionnaires to students. Next, a task analysis is conducted to identify the key skills students must acquire. This analysis also ensures the scope of the assessment instrument being developed. Next, a learning outcomes analysis is conducted, which will form the basis for developing the assessment instrument to be used by teachers and students. Learning outcomes analysis was conducted by identifying learning outcomes in accordance with the Decree of the Head of the Education Standards, Curriculum, and Assessment Agency of the Ministry of Education, Culture, Research, and Technology Number 032/H/Kr/2024 concerning Learning Outcomes in Early Childhood Education, Elementary Education, and Secondary Education in the Independent Curriculum. This was done to ensure that assessment instruments align with the applicable curriculum requirements. The results of the needs analysis, student analysis, task analysis, and learning outcomes analysis formed the basis for formulating the research problem background, which then led to the formulation of the research title.

The Design stage of the RPA-based assessment instrument to measure students' high-level thinking skills in the material of work and energy involves determining the assessment objectives, designing the item grid, and compiling high-level thinking skills items along with their scoring guidelines. The RPA-based assessment instrument developed integrates real images around students to be analyzed conceptually in Physics.

The Development stages include expert and practitioner validation of the RPA-based assessment instrument, as well as limited trials with students. Prior to the trial, students are first given Real Picture Analysis (RPA)-based learning to familiarize them with the thinking context and visual representations that will appear in the assessment. This is done to ensure the instrument truly measures higher-order thinking skills validly and in an authentic learning context.

Experts and practitioners carry out validation. If the product is valid, it can proceed to the next stage. However, if the product is not valid, it must be revised and revalidated until it is valid. Once the developed product is declared valid, it proceeds to limited trials. Limited trials were conducted with 30 eighth-grade students from a junior high school to determine whether the developed RPA-based assessment instrument performed as expected. The results of student responses in the small-scale trial were then analyzed for validity, reliability, discrimination power, and difficulty level. The results of the limited trials will be used for evaluation and then used as the basis for further revisions if any aspects do not meet expectations. Trials and revisions were conducted repeatedly to achieve an effective and consistent instrument.

The results of the revised small-scale limited trial were retested on a large group of 180 eighth-grade junior high school students to determine whether the developed product, an RPA-based assessment instrument, performed as expected. Student responses from the large-scale trial were then analyzed for validity, reliability, discriminatory power, and difficulty level to obtain a final product for dissemination.

The expert validity analysis of test items is conducted based on three review aspects: content, construction, and language. Content validity is calculated using Aiken's V formula, which produces a content validity coefficient derived from expert assessments. The assessment instrument is further validated using the product-moment correlation formula, where an item is considered valid if the calculated r_{value} is greater than the r_{table} value. The reliability of the RPA-based essay assessment instrument is analyzed using the Cronbach-Alpha formula to measure internal consistency. The difficulty level is determined by the ratio of students answering correctly to the total test takers, while the discrimination index is calculated from the proportion of upper- and lower-group students answering correctly. Overall, this process ensures that the developed assessment instrument is valid, reliable, and has appropriate difficulty and discrimination levels to measure student ability in an academic context accurately.

The dissemination stages to expand the use of the developed RPA-based assessment instruments. Dissemination will be carried out by providing RPA-based assessment instrument packages to science teachers for use in measuring students' higher-order thinking skills.

This research, developing an RPA-based assessment instrument to measure students' higher-order thinking skills, was conducted in four public junior high schools in Jember Regency: SMPN 3 Tanggul, SMPN 7 Jember, SMPN 1 Panti, and SMPN 1 Rambipuji. The study was conducted in the odd semester of the 2024/2025 academic year, with a sample of two eighth-grade students from each school. All participants in this study, including students and teachers, provided informed consent prior to their participation. This study guaranteed the confidentiality and anonymity of all participants, and data were handled in accordance with strict ethical guidelines to protect participants' rights and privacy. The data collection technique used in this study was a test administered to students. The data collection instruments used were a validation sheet and an assessment instrument package.

C. Result

1. Define Stage Results

The results of the Define stage analysis include needs analysis, student analysis, task analysis, and learning achievement analysis as follows:

a. Front-end analysis

The data for the needs analysis were collected through a questionnaire completed by junior high school science teachers to identify existing problems in the field. The questionnaire was administered to 15 science teachers in Jember Regency; the results of the needs analysis questionnaire are presented in Figure 2.

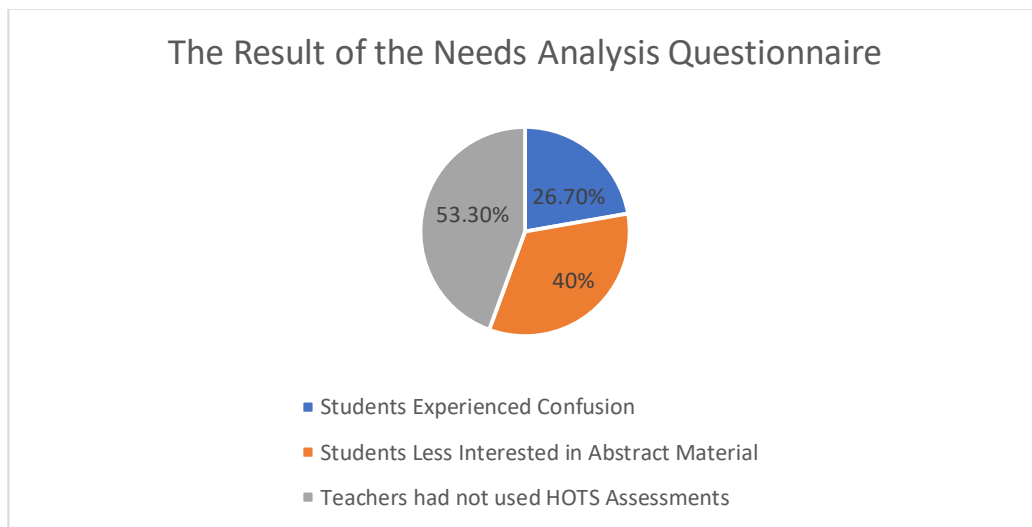


Figure 2. The Result of the Needs Analysis Questionnaire

Based on the needs analysis, the data reveals a notable challenge in student engagement and comprehension related to abstract physics concepts; Specifically, over one quarter of students experienced confusion, and 40% expressed low interest in abstract material. This indicates a critical need for instructional strategies and assessment tools that provide concrete, contextualized representations to facilitate understanding. Concurrently, more than half of the surveyed teachers reported not using higher-order thinking skills (HOTS) assessments, primarily due to difficulties in formulating questions, a lack of adequate references, and a limited understanding of HOTS criteria. Furthermore, the minimal use of real-image media in assessments – used by only about a quarter of teachers – underscores a gap in integrating authentic stimuli that can bridge conceptual abstractness with tangible contexts. This data supports the strategic development of Real Picture Analysis (RPA)-based assessment instruments, which utilize real-world visual stimuli to enhance cognitive engagement and better align with the demands of 21st-century education.

b. Learner analysis

Data collection for student analysis information was obtained through a questionnaire completed by junior high school students to identify existing problems in the field. The questionnaire was administered to 180 junior high school students in Jember Regency, as illustrated in Figure 3.

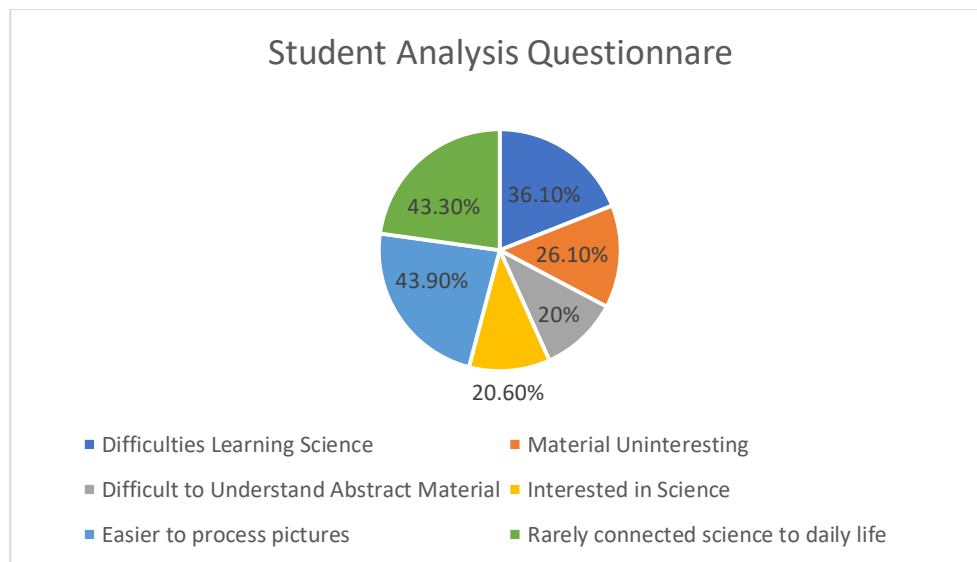


Figure 3. Student Analysis Questionnaire

The data suggest that a notable proportion of students face challenges in learning science due to the abstract nature of the material and the predominance of calculations and formulas, which can hinder engagement and understanding. Although nearly 60% of students expressed a positive interest in science, a considerable gap remains in connecting scientific concepts to everyday experiences, as indicated by the low frequency of students relating the material to real-life situations. The finding that over half of the students benefit from visual aids, such as simple pictures or illustrations, highlights the importance of integrating visual and contextual supports into assessment and instructional design to enhance comprehension. This underscores the necessity for assessment tools like Real Picture Analysis (RPA), which utilize concrete and authentic visual stimuli to bridge abstract scientific content with tangible understanding, ultimately fostering deeper cognitive engagement.

c. Task analysis

In the 21st century, students need higher-order thinking skills to compete, solve problems, and face challenges in everyday life. Real Picture Analysis (RPA)-based assessment instruments are an effective alternative to traditional evaluation tools for measuring students' higher-order thinking skills. By presenting questions based on real-life images, this instrument encourages students to go beyond memorizing concepts to analyzing relationships between concepts, critically evaluating information, and applying Physics knowledge in relevant contexts. Real Picture Analysis (RPA) serves as a stimulus that activates higher-order thinking skills such as analyzing (C4), evaluating (C5), and creating (C6). With this approach, the assessment process becomes more meaningful, contextual, and able to reflect the expected 21st-century competencies.

d. Specifying instructional objectives

Learning outcomes analysis is conducted to identify, detail, and organize concepts relevant to the applicable curriculum. The scope of the material aligns with the curriculum used, specifically the independent curriculum for junior high school level, which is included in Phase D. The scope of the material focuses on work and energy. Learning objectives are aligned with the learning outcomes stipulated in the applicable Independent Curriculum.

2. Design Stage Results

The design phase involved creating an RPA-based assessment instrument to measure students' higher-order thinking skills in the subject of work and energy. The RPA-based assessment instrument was developed with integrated real-world images around students for analysis based on physics concepts. The design process yielded an RPA-based assessment instrument package comprising question indicators, assessment tools, and scoring guidelines.

The assessment instrument was structured based on the learning outcomes and objectives of the Work and Energy subject to ensure a link between the question items, the competencies being measured, and the real-life context. The question indicators were developed by incorporating aspects of higher-order thinking skills, namely the cognitive levels based on the revised Bloom's Taxonomy, including analyzing (C4), evaluating (C5), and creating (C6). The descriptive question items were designed using real-world image stimuli, specifically Real Picture Analysis (RPA), selected based on contextual phenomena relevant to students. This aimed to measure the ability to analyze, evaluate, and apply physics concepts in a meaningful and effective manner. The scoring guidelines were developed analytically to provide clear and objective guidance in assessing students' responses.

3. Development Stage Results

The Develop stage involves the expert and practitioner validation of RPA-based assessment instruments, as well as limited trials with students. Prior to pilot testing, students are first introduced to Real Picture Analysis (RPA)-based learning to familiarize them with the thinking context and visual representations that will appear in the assessment. This is done to ensure the instrument truly measures higher-order thinking skills validly and appropriately for the learning context.

The RPA-based assessment instrument consists of 30 descriptive questions. The descriptive questions are subjectively scored based on the completeness of the student's answers, with a score ranging from 1 to 10. The test items must meet valid criteria. Three types of validation determine the validity of the RPA-based assessment instrument: construct, content, and language validation. The expert validation results determine the validity of the higher-order thinking skills assessment instrument before it is tested on students. If the validation results are declared valid, the instrument is ready to be tested on students.

The validation results indicate that the developed instrument has a good level of validity. The total percentages of assessments from the three validators were 84.75%, 81.04%, and 89.04%, respectively, with Aiken's V scores ranging from 0.70 to 0.87. Overall, the average Aiken's V score from the three validators for all questions was 0.81, indicating that the majority of questions have high validity.

a. Small-Scale Trials

Prior to the trial of the Real Picture Analysis (RPA)-based assessment instrument, students first participated in a lesson using the RPA approach. The learning activities were conducted over three sessions, focusing on the topic of work and energy. A small-scale trial was conducted with 30 students in small groups, and observers monitored whether the learning activities aligned with the steps outlined in the teaching module. Following the learning activities, students were assessed using the Real Picture Analysis (RPA)-based instrument.

Using the product-moment correlation formula, data obtained can be considered valid if the calculated r-value is greater than or equal to the corresponding r-value in the table. Based on the validity analysis of 30 questions, 25 were declared valid and five were declared invalid. The declared invalid items were 4, 19, 23, 24, and 28. Therefore, for the reliability test, these five items were excluded to maintain the quality of the assessment instrument. Next, a reliability test was conducted using the Cronbach's Alpha formula. The results of the reliability test on the 25 valid questions are shown in Table 4.1.

Table 1. Small-Scale Reliability Trial

Items	B	Category
Essay	0.78	High

The reliability test yielded a Cronbach's alpha score of 0.78, indicating that the Real Picture Analysis (RPA)-based assessment instrument demonstrated a high level of internal consistency. This score suggests that the instrument's items are sufficiently correlated, ensuring that the assessment reliably measures the intended construct of higher-order thinking skills. While a value of 0.70 is commonly accepted as the minimum threshold for acceptable reliability, a score of 0.80, as observed here, reflects a dependable level of measurement consistency suitable for research and practical educational use. Consequently, the findings support the conclusion that the instrument can provide stable and reproducible results when administered under similar conditions. Next, discriminant power and difficulty levels were tested on a small-scale trial. The results of the small-scale discriminant power and difficulty levels test are shown in Table 2.

Table 2. Small-Scale Discriminant Power and Difficulty Levels Trial

Items	Discriminant Power	Difficulty Levels
1	Good	Easy

Items	Discriminant Power	Difficulty Levels
2	Very Good	Easy
3	Good	Easy
4	Very Good	Medium
5	Good	Easy
6	Good	Medium
7	Good	Easy
8	Good	Easy
9	Very Good	Medium
10	Good	Medium
11	Good	Easy
12	Good	Medium
13	Good	Easy
14	Good	Easy
15	Good	Medium
16	Good	Medium
17	Very Good	Medium
18	Good	Difficult
19	Very Good	Medium
20	Good	Medium
21	Good	Difficult
22	Good	Medium
23	Good	Difficult
24	Very Good	Difficult
25	Very Good	Difficult

The results reported in Table 2 indicate that the assessment instrument demonstrates strong discriminatory power, with seven questions showing very good discrimination and 18 questions showing good discrimination. This means that these items effectively differentiate between students who have a higher level of ability and those with a lower level of ability, making them suitable for assessing student performance. Additionally, the distribution of item difficulty across the 25 questions shows a balanced mix of easy, medium, and difficult items (9 easy, 11 medium, and five difficult). Such a balance is crucial for an effective assessment instrument, as it ensures a comprehensive evaluation of students' skills across various ability levels. Overall, these findings suggest that the Real Picture Analysis (RPA)-based assessment instrument is well-constructed and meets the criteria necessary to be considered a reliable measurement tool for assessing student abilities.

The final results of the development phase indicate that the Real Picture Analysis (RPA)-based assessment instrument meets the criteria for validity, reliability, discriminatory power, and a good level of difficulty.

b. Large Scale Trials

Prior to the large-scale trial of the Real Picture Analysis (RPA)-based assessment instrument, students first participated in a lesson using the RPA approach. This lesson aimed to familiarize students with analyzing real-world images as representations of

physics concepts, so that when faced with RPA-based assessment questions, they would not experience confusion or misconceptions due to unfamiliarity with this type of presentation format. The learning activities were conducted over three sessions, covering the main topic of work and energy. An observer was assigned to monitor the alignment of the learning activities with the steps in the teaching module. The observer completed an observation sheet for the implementation of the learning and provided an evaluation.

After the lesson, students completed the assessment instrument, and the results were analyzed for validity, reliability, discriminatory power, and difficulty level. A summary of the instrument test results is presented in Table 3.

Table 3. A Summary of The Instrument Test Results

Items	Validity	Reliability	Discriminant Power	Difficulty Level
1	Valid	Reliable	Good	Easy
2	Valid	Reliable	Good	Easy
3	Valid	Reliable	Good	Easy
4	Valid	Reliable	Good	Medium
5	Valid	Reliable	Good	Easy
6	Valid	Reliable	Good	Medium
7	Valid	Reliable	Good	Easy
8	Valid	Reliable	Good	Easy
9	Valid	Reliable	Good	Medium
10	Valid	Reliable	Good	Medium
11	Valid	Reliable	Good	Easy
12	Valid	Reliable	Good	Medium
13	Valid	Reliable	Good	Easy
14	Valid	Reliable	Good	Easy
15	Valid	Reliable	Good	Medium
16	Valid	Reliable	Good	Medium
17	Valid	Reliable	Good	Medium
18	Valid	Reliable	Very Good	Difficult
19	Valid	Reliable	Good	Medium
20	Valid	Reliable	Good	Medium
21	Valid	Reliable	Very Good	Difficult
22	Valid	Reliable	Good	Medium
23	Valid	Reliable	Very Good	Difficult
24	Valid	Reliable	Very Good	Difficult
25	Valid	Reliable	Very Good	Difficult

The validity of the data was determined by comparing the calculated correlation coefficient (r) to the critical value from the statistical table; items were deemed valid if the correlation coefficient exceeded the critical value. Based on this criterion, 25 items were confirmed as valid. The reliability analysis yielded a Cronbach's alpha coefficient of 0.856, indicating a high level of internal consistency and measurement stability for the Real Picture Analysis (RPA)-based assessment instrument. Furthermore, the discriminatory power evaluation revealed that 20 items exhibited excellent discrimination, while the remaining 5

demonstrated good discrimination, confirming the instrument's efficacy in distinguishing between varying levels of student ability. The difficulty analysis classified the items across easy, medium, and difficult categories, comprising 9, 11, and 5 items, respectively. This balanced distribution of item difficulty levels supports the instrument's ability to assess student competencies comprehensively across a spectrum of challenges. Collectively, these psychometric properties confirm the instrument's suitability and robustness for assessing higher-order thinking skills.

The comparison of instrument test results between small and large scales presents insightful findings with implications for assessment quality and applicability. The higher validity coefficient in the small-scale trial compared to the large-scale trial may indicate that the initial instrument's construct alignment and clarity were stronger under controlled or limited conditions, whereas the scalability introduced variability affected item validity, as shown by Figure 4.

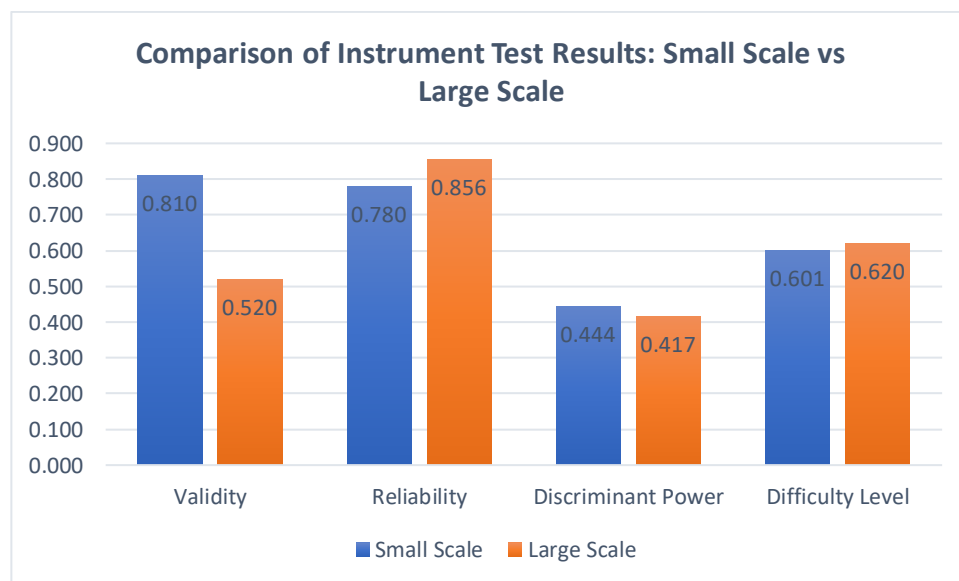


Figure 4. Comparison of Instrument Test Results: Small Scale vs Large Scale

The comparison of instrument test results between small and large scales presents insightful findings with implications for assessment quality and applicability. The higher validity coefficient in the small-scale trial compared to the large-scale trial may indicate that the initial instrument's construct alignment and clarity were stronger under controlled or limited conditions, whereas the scalability introduced by the large-scale trial affected item validity. However, the reliability coefficient improves from small scale to large scale, demonstrating greater internal consistency when applied to a broader, more diverse sample, which supports the instrument's stability across populations. Discrimination indices, which are moderately high and relatively consistent, confirm that the instrument's capacity to differentiate student abilities remains effective regardless of the scale. The comparable and balanced difficulty levels suggest that item challenge remains appropriate and consistent across pilot phases. Taken together, these results indicate that while validity may fluctuate

with scale, the overall instrument maintains strong reliability, effective discrimination, and suitable difficulty, supporting its use as a robust tool for assessing higher-order thinking skills in varied student populations.

4. Disseminate Stage Results

The dissemination of the development product in the form of an RPA-based Assessment Instrument is the final stage, following the product's validation, reliability, and demonstration of good discrimination power and difficulty level, which enables it to measure students' higher-order thinking skills. The development product, in the form of an RPA-based Assessment Instrument, was disseminated to several schools through the provision of assessment instrument packages, including SMPN 3 Tanggul, SMPN 7 Jember, SMPN 1 Rambipuji, and SMPN 1 Panti.

D. Discussion

1. Define Stage Discussion

The Define Stage includes needs analysis, student analysis, task analysis, and learning outcome analysis. The questionnaire results showed that most teachers have not used higher-order thinking assessments. This is due to various obstacles, including difficulty in developing questions, a lack of references, and a limited understanding of the criteria for higher-order thinking questions. Teachers experience significant difficulties in developing higher-order thinking skills questions due to a lack of understanding of the concept of higher-order thinking skills, difficulty determining cognitive levels, choosing appropriate operational verbs, and carrying out technical steps in question development, such as creating grids and contextual stimuli (Pramesti & Riastini, 2023). Other contributing factors include teachers' high workloads, difficulty finding references for questions that require higher-order thinking skills, and a lack of preparedness in developing learning techniques and methods that effectively integrate higher-order thinking skills into the classroom. Teachers often focus more on material achievement than on developing students' higher-order thinking skills (Astrid et al., 2022). Furthermore, the use of real-world images in assessments is still very minimal. Therefore, the development of a real-picture-based assessment instrument, hereinafter referred to as Real Picture Analysis (RPA), is a strategic step to bridge the gap between abstract concepts and student understanding, while simultaneously encouraging improved assessment quality relevant to the needs of 21st-century learning.

The questionnaire results also indicated that students experienced difficulties and lacked interest in abstract science material. This aligns with research by Rizaldi & Fatimah (2024), which stated that students experience difficulties learning physics, particularly abstract concepts such as dynamics, waves, and optics. This difficulty is also triggered by the large number of formulas that must be mastered, making it difficult for students to grasp the concepts deeply (Rizaldi & Fatimah, 2024). Abstract physics concepts are difficult to grasp because there are no real-world examples in the students' environment.

Understanding these concepts requires higher-order thinking processes, and approaches such as the use of analogies or concrete media are essential to help students grasp these difficult concepts (Zahriah et al., 2025). Despite this, students' learning motivation for science is quite good. Learning motivation is a crucial factor that supports students' academic achievement in science learning. Motivation encompasses intrinsic and extrinsic aspects, task value, learning control beliefs, and self-efficacy, all of which play a significant role in encouraging students to remain motivated despite facing difficulties in learning science (Nacaroglu et al., 2025). One method to help students understand concepts is to provide simple images or illustrations as stimuli, especially those that can relate science material to everyday life. The use of visual media significantly improves students' understanding of complex science concepts (Mayarita et al., 2023). Appropriate and relevant images, especially those that relate the material to students' everyday contexts, can reduce the level of abstraction of the material and make learning more engaging and effective (Hanif, 2020). These findings suggest the need for assessment media that can reduce the abstract nature of the material and enhance the connection between concepts and everyday reality. Therefore, the development of a real-picture-based assessment instrument, known as Real Picture Analysis (RPA), is seen as a strategic step to bridge the conceptual gap, improve student understanding, and support the quality of assessment relevant to the demands of 21st-century learning.

2. Design Stage Discussion

The Design Stage in developing a Real Picture Analysis (RPA)-based assessment instrument plays a crucial role in ensuring the instrument aligns with learning objectives and student competency outcomes. This instrument is designed to measure students' higher-order thinking skills, referring to the cognitive levels in the revised Bloom's Taxonomy, which include analyzing (C4), evaluating (C5), and creating (C6) (Wilson, 2016). The emphasis on these cognitive skills is based on the needs of 21st-century learning, which requires students to think critically and creatively, beyond simply memorizing facts. By using clear question indicators aligned with learning objectives, this instrument is expected to measure higher-order thinking skills in a valid and relevant manner (Desiriah & Setyarsih, 2021).

The use of real-life images taken from contextual phenomena surrounding students represents a significant innovation in the design of this instrument. The Real Picture Analysis (RPA) approach aims to bridge the gap between abstract physics concepts and students' concrete experiences in everyday life. This aligns with research indicating that visualization and real-world contexts significantly enhance students' understanding of complex physics concepts (Oktaviana et al., 2021). With real-world image stimuli, it is hoped that students will be able to connect physics theories more meaningfully with objects and events in their environment, thereby enhancing conceptual understanding and higher-order thinking skills.

Furthermore, detailed and objective analytical scoring guidelines were developed during this stage of the design. These guidelines ensure that student responses can be scored consistently and fairly, facilitating teachers' assessment of the analytical, evaluative, and creative skills developed through the instrument. Transparent analytical scoring also supports the instrument's reliability, ensuring reliable feedback for instructional improvement (Zahriah et al., 2025). Therefore, the overall design of this instrument focuses not only on the question content but also on the technical aspects of the assessment, which support the comprehensive measurement of higher-order thinking skills.

3. Develop Stage Discussion

After the learning process was implemented, a validity test was conducted using the product-moment correlation formula, and the validity of 25 test items was determined to be valid. Validity is a measure of the extent to which an instrument actually measures what it is supposed to measure (Amelia & Erita, 2024). An instrument is considered valid when it demonstrates that the measuring tool used to obtain data is accurate and can be used to measure what it is intended to measure (Desiriah & Setyarsih, 2021). Next, a reliability test was conducted using the Cronbach's Alpha formula. The higher the reliability value of an instrument, the higher the level of trustworthiness it possesses (Okyanida et al., 2024). Therefore, the results of the reliability test for the Real Picture Analysis (RPA)-based assessment instrument were declared to have a high level of reliability.

The ability of test items to detect differences in student abilities is evident from the high or low discriminatory power criteria for the items. Discriminatory power tests were also conducted on the pilot test results. Based on the results of the discriminatory power test, 20 test items met the very good discriminatory power criteria, and 5 test items met the good criteria. All items tested for discriminatory power met the criteria and were therefore suitable for use in assessment activities.

Next, a difficulty test was conducted, yielding results for 25 items tested, each with easy, medium, and difficult levels. A good instrument must strike a balance between easy, medium, and difficult items to accurately measure student abilities. This aligns with Okyanida et al (2024), who stated that a quality instrument must meet the criteria of validity, reliability, a difficulty index, and good discriminatory power.

The final results of the development phase indicated that the RPA-based assessment instrument met the criteria of validity, reliability, discriminatory power, and a good level of difficulty. The developed Real Picture Analysis (RPA)-based assessment instrument was found to be valid, as demonstrated by the validation of 25 test items through product-moment correlation, indicating the instrument accurately measures the intended higher-order thinking skills. Reliability analysis, using Cronbach's Alpha, yielded a high coefficient, confirming the instrument's consistency and reliability for repeated assessments. Additionally, the instrument effectively discriminated among students' abilities, with most items meeting the criteria for very good or good discriminatory power, ensuring sensitivity to varied cognitive levels. The items were explicitly designed to align with higher-order

thinking skills, targeting cognitive processes such as analysis, evaluation, and creation, thereby affirming their appropriateness in measuring HOTS competencies.

The effectiveness of RPA as an assessment medium lies in its use of contextualized, concrete visual stimuli derived from real-world images, which helps students connect abstract physics concepts, such as work and energy, to tangible experiences, facilitating deeper comprehension and critical thinking. However, a limitation of this approach is its dependence on students' visual literacy; those with lower visual interpretive skills may struggle to engage fully with the instrument, which could affect their performance. This highlights the necessity of complementing RPA-based assessments with interventions to develop visual literacy or incorporate multimodal supports, thereby ensuring equitable assessment outcomes across diverse learner profiles.

4. Disseminate Stage Discussion

The dissemination of the development product in the form of an RPA-based Assessment Instrument is the final stage, following the product's validation, reliability, and demonstration of good discrimination power and difficulty level, which enables it to measure students' higher-order thinking skills. The development product, in the form of an RPA-based Assessment Instrument, was disseminated to several schools through the provision of assessment instrument packages, including SMPN 3 Tanggul, SMPN 7 Jember, SMPN 1 Rambipuji, and SMPN 1 Panti.

E. Implication

Theoretically, this research makes a significant contribution to the development of authentic assessment, particularly in the context of measuring higher-order thinking skills through the Real Picture Analysis (RPA) approach. This approach enriches authentic assessment studies by providing an instrument capable of linking abstract physics concepts to real phenomena around students, thereby adding a dimension of authenticity to the assessment process (Desiriah & Setyarsih, 2021). Practically, the results of this study provide guidelines that teachers can adapt to develop and implement RPA-based assessments, not only on topics related to work and energy, but also on other materials according to learning needs. This allows teachers to effectively integrate visual media in assessments that stimulate higher-order thinking skills in students (Murwantini, 2022). From a social perspective, this RPA-based assessment instrument supports the Independent Curriculum policy by strengthening students' scientific literacy through a contextual approach and active engagement in learning. Thus, this research also supports the government's efforts to improve the quality of education and prepare students to face the challenges of the 21st century through more meaningful and authentic scientific literacy.

F. Limitation and Suggestion for Further Research

The limitation of this research is the development of RPA-based assessment instruments using the 4D research method, specifically Define, Design, Develop, and

Disseminate. The dimensions of high-order thinking skills applied in the development of this essay-based assessment instrument comprise C4 (Analysis), C5 (Evaluation), and C6 (Creation). This assessment instrument is also limited to paper-and-pencil tests and has not been tested in the digital realm. Science learning in this research is limited to the material of Work and Energy for Class VIII SMP Negeri Jember in the Odd Semester using the Independent Curriculum.

For further research, this RPA-based assessment instrument is expected to provide an innovative alternative solution in the evaluation process, utilizing questions that demonstrate higher-order thinking skills. Furthermore, further research should lead to the development of more interactive and adaptive digital assessments for widespread application in various learning contexts. Cross-regional trials are also important to identify the acceptability and effectiveness of this instrument across diverse student characteristics. Furthermore, similar developments can include other physics topics to enrich the application of RPA-based assessment instruments in comprehensively measuring higher-order thinking skills. Thus, this research is expected to contribute to the advancement of educational technology and serve as a reference for further studies in the field of technology-based assessment.

G. Conclusion

This study confirms that the Real Picture Analysis (RPA)-based assessment instrument is valid, reliable, and effective in measuring students' higher-order thinking skills (HOTS) in physics, specifically in the context of the Work and Energy topic. The instrument demonstrates strong construct validity and high reliability, ensuring that it consistently measures students' analytical and reasoning abilities as intended. Moreover, the balanced distribution of item difficulty and good discriminatory power indicates that the instrument can effectively differentiate students' levels of understanding. Overall, this study contributes to the development of innovative assessment models in physics education by integrating real-picture contexts to promote and evaluate higher-order thinking skills more authentically.

References













- Aisyah, O. N., & Sudarti. (2021). Analisis Kemampuan Multirepresentasi Verbal dan Gambar pada Mahasiswa Pendidikan Fisika dalam Memahami Konsep Reaksi Inti Matahari. *Silampari Jurnal Pendidikan Ilmu Fisika*, 3(1), 29–38. <https://doi.org/10.31540/sjpif.v3i1.1136>
- Amelia, N., & Erita, S. (2024). Eksplorasi Validitas dan Reliabilitas Soal Pemahaman Konsep dalam Asesmen Pembelajaran. *Jurnal Bima*, 2(1), 222–232.
- Amir, F., M, M., & Nur, A. A. S. (2024). Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) Peserta Didik pada Mata Pelajaran Fisika. *Jurnal Fisika dan Pembelajarannya (Phydogic)*, 6(2). <https://doi.org/10.31605/phy.v6i2.3653>

- Astrid, A., Hasanah, A., & Syafradin. (2022). Integrating Higher Order Thinking Skills (HOTS) Into English Language Teaching for Elementary School Students: Teachers' Perspectives and Challenges. *3L: Language, Linguistics, Literature*, 28(3), 217-230. <https://doi.org/10.17576/3L-2022-2803-14>
- Budiarso, A. S., Sutarto, & Rohmatillah, S. (2020). Analisis Kemampuan Siswa dalam Menjelaskan Fenomena IPA di Sekitar Lingkungan. *Webinar Pendidikan Fisika 2020*, 5(1), 27-32.
- Desiriah, E., & Setyarsih, W. (2021). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Fisika di SMA. *ORBITA: Jurnal Kajian, Inovasi dan Aplikasi Pendidikan Fisika*, 7(1), 79. <https://doi.org/10.31764/orbita.v7i1.4436>
- Hanif, M. (2020). The Development and Effectiveness of Motion Graphic Animation Videos to Improve Primary School Students' Sciences Learning Outcomes. *International Journal of Instruction*, 13(3), 247-266.
- Hidayati, A. R., Fadly, W., & Ekapti, R. F. (2021). Analisis Keterampilan Berpikir Kritis Siswa pada Pembelajaran IPA Materi Bioteknologi. *Jurnal Tadris IPA Indonesia*, 1(1), 34-48. <https://doi.org/10.21154/jtii.v1i1.68>
- Julianingsih, S., Rosidin, U., & Ismu Wahyudi. (2013). Pengembangan Instrumen Asesmen HOTS untuk Mengukur Dimensi Pengetahuan Ipa Siswa di SMP. *Jurnal Pembelajaran Fisika*, 1, 59-68.
- Kokkonen, T., Lichtenberger, A., & Schalk, L. (2022). Concreteness Fading in Learning Secondary School Physics Concepts. *Learning and Instruction*, 77. <https://doi.org/10.1016/j.learninstruc.2021.101524>
- Lamanepa, G. H., Maing, C. M., Mukin, M. U. J., & Naen, A. B. (2022). The Role of Visual Representation for High School Physics in Teaching of Classical Mechanics. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 8(1), 105-114. <https://doi.org/10.21009/1.08110>
- Limbach, B., & Waugh, W. (2022). Developing High-Level Thinking. *Journal of Instructional Pedagogies*, 125-143. https://doi.org/10.1142/9789811252617_0009
- Idris, S., Muliani, M., Novita, N., Sakdiah, H., Ginting, F. W., Mellyzar, M., & Fatwa, I. (2023). Aksi Bersih Pantai dan Edukasi Peduli Lingkungan di Pantai Wisata Gampoeng Geulumpang Sulu Timue. *JALIYE: Jurnal Abdimas, Loyalitas, Dan Edukasi*, 2(1), 5-12.
- Mayarita, L. R., Dafik, D., & Prastati, T. (2023). The Development of Environment-Based Visual Media to Enhance Learning Outcomes and Student Motivation in Science Course. *International Journal of Current Science Research and Review*, 06(07), 4420-4428. <https://doi.org/10.47191/ijcsrr/v6-i7-56>
- Meryastiti, V., Rasyid Ridlo, Z., & Supeno. (2022). Identifikasi Kemampuan Berpikir Kritis pada Pembelajaran IPS Siswa SMP Negeri 1 Glenmore Kabupaten Banyuwangi. *Saintifika*, 24(1), 20-29.

- Murwantini, S. (2022). Implementasi Keterampilan Berpikir Tingkat Tinggi (HOTS) dalam Pembelajaran. *BALANGA: Jurnal Pendidikan Teknologi dan Kejuruan*, 10(2), 49–54. <https://doi.org/10.37304/balanga.v10i2.8090>
- Nacaroglu, O., Kizkapan, O., & Demir, H. (2025). Middle School Students' Motivations and Learning Competencies in Science: Mediating Role of Digital Literacy. *Psychology in the Schools*, 62(5), 1475–1487.
- Nikat, R. F., Loupatty, M., & Zahroh, S. H. (2021). Kajian Pendekatan Multirepresentasi dalam Konteks Pembelajaran Fisika. *Jurnal Pendidikan dan Ilmu Fisika*, 1(2), 45. <https://doi.org/10.52434/jpif.v1i2.1449>
- Oktaviana, J., Sutarto, S., & Mahardika, I. K. (2014). Paket Bahan Ajar dengan Analisis Kejadian Riil dalam Foto dan Wacana Isu dalam Pembelajaran Fisika di SMA (Kajian Pada: Konsep Fluida Statis). *Jurnal Pembelajaran Fisika*, 3(3), 230-234.
- Okyranida, I. Y., Mayanty, S., & Widiyatun, F. (2024). Analisis Butir Soal Kemampuan Berpikir Kritis Siswa SMAIT Nururrohmah Depok. *Jurnal Penelitian Pembelajaran Fisika*, 15(1), 73–79. <https://doi.org/10.26877/jp2f.v15i1.17057>
- Pramesti, N. K. A., & Riastini, P. N. (2023). Professional Status: Teachers' Difficulties in Compiling Questions Based on Higher Order Thinking Skills. *International Journal of Elementary Education*, 7(3), 516–523. <https://doi.org/10.23887/ijee.v7i3.61625>
- Pursitasari, I. D., Permanasari, A., & Jaenudin, D. (2022). Pelatihan Penyusunan E-Asesmen Literasi Sains Berbasis AKM Bagi Guru IPA SMP di Kabupaten Bogor. *Jurnal Pemberdayaan Masyarakat*, 1(1), 26–33. <https://doi.org/10.46843/jmp.v1i1.265>
- Putri, A. N. L., Sutarto, S., Wahyuni, D., Indrawati, I., & Wicaksono, I. (2023). Real Picture Analysis Modules: Empowering Critical Thinking in Science Education. *Advances and Challenges in Science and Technology*, 5, 68–83. <https://doi.org/10.9734/bpi/acst/v5/6128e>
- Rizaldi, D. R., & Fatimah, Z. (2024). Analysis of Various Contexts of Difficulties Experienced by Students in Learning Physics. *Physics Education Research Journal*, 6(2), 103–108. <https://doi.org/10.21580/perj.2024.6.2.14122>
- Tawil, M., & Tampa, A. (2024). Pelatihan Pembuatan Asesmen Portofolio Berorientasi Pembelajaran Kooperatif Bagi Guru IPA. *Jurnal Altifani Penelitian dan Pengabdian kepada Masyarakat*, 4(1), 85-96.
- Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The Development of The Hots Test of Physics Based on Modern Test Theory: Question Modeling Through E-Learning of Moodle LMS. *International Journal of Instruction*, 14(4), 51–68. <https://doi.org/10.29333/iji.2021.1444a>
- Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy Revised Understanding the New Version of Bloom's Taxonomy. *The Second Principle*, 1(1), 1-8.

Zahriah, Z., Annisa, F., Ahmad, N. H. B., & Fahira, A. (2025). Design and Validation of a Diagnostic Test Instrument for Identifying Misconceptions in Basic Physics. *ORBITA: Jurnal Pendidikan dan Ilmu Fisika*, 11(1), 1-11. <https://doi.org/10.31764/orbita.v11i1.27782>

Author's Biography

	<p>Inas Aulia Majid, S.Pd.    Was born in Malang on September 17, 1996. She is currently a student at Universitas Jember (Unej). She completed her Bachelor's degree (S1) in Physics Education at Universitas Negeri Malang in 2019 and is now pursuing his Master's degree (S2) in Science Education at Universitas Jember. Email: inas.aulia2@gmail.com</p>
	<p>Prof. Dr. Sutarto, M.Pd.    He is a Professor and a Lecturer in the Physics Education Study Program at the Faculty of Teacher Training and Education, University of Jember, Indonesia. His research focuses on science education, STEM learning, and process images. Email: sutarto.fkip@unej.ac.id</p>
	<p>Dr. Bea Hana Siswati, M.Pd.    She is a Lecturer in the Biology Education Study Program at the Faculty of Teacher Training and Education, University of Jember, Indonesia. Her research interests include metacognition and biology education. Email: beahana.fkip@unej.ac.id</p>