JPPIPA 11(10) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

HOTS Assessment Development to Assess High School Physics Problem-Solving Skills

Veni Wahyuni[1*], Ratnawulan[2], Fatni Mufit[2]

[1] Magister of Physics, Faculty of Mathematics and Natural Sciences, Padang State University, Padang, Indonesia.
[2] Department of Physics, Faculty of Mathematics and Natural Sciences, Padang State University, Padang, Indonesia.

Received: June 14, 2025
Revised: September 17, 2025
Accepted: October 25, 2025
Published: October 31, 2025

Corresponding Author:
Veni Wahyuni
wahyuniveni9@gmail.com

DOI: 10.29303/jppipa.v11i10.12880

Abstract: This study aims to develop an assessment instrument based on Higher Order Thinking Skills (HOTS) to assess high school students' problem-solving abilities in the subject of physics on renewable energy using the Plomp development model. The study was conducted through three main stages, namely (1) preliminary research to analyze needs (2) prototyping phase to design and validate the assessment instrument, and (3) assessment phase to conduct limited trials and product revisions. The assessment instrument was developed by containing essay questions based on real contexts that require critical, analytical, and creative thinking skills. The results of expert validation showed that the assessment was in the valid category with an average percentage score of 0.93. Trials on high school students showed that the assessment was able to measure variations in problem-solving abilities more comprehensively. Thus, this HOTS assessment based on the Plomp model is effective for evaluating physics problem-solving abilities.

Keywords: High school physics; HOTS assessment; Plomp model; Solution to problem

Introduction

The 21st century demands skills that include critical thinking, creativity, effective communication, collaboration, and complex problem-solving. In education, these skills are essential to empowering students to adapt and succeed amidst rapid change (Rohman & Hendra, 2023). Students need more than just mastering and memorizing theory to survive in the 21st century; they also need various higher-order thinking skills, one of the most important of which is problem-solving (Rusmin & Misrahayu, 2024). This aligns with the characteristics of physics, which demands a deep understanding and the ability to analyze various natural phenomena.

Physics is a discipline within the natural sciences that plays a crucial role in helping students understand the various events and natural phenomena that occur around them (Mahardika et al., 2023; Ramadhan et al., 2019). According to Aprilia et al. (2022) and Aina et al. (2023), physics learning is concerned with mastering concepts, laws, theories, principles, and their applications in everyday life. To understand and apply physics concepts in real life, students require not only theoretical mastery but also higher-order thinking skills, one of which is problem-solving ability.

Problem solving is a process that a person carries out to resolve the problems they face until the problem is no longer a problem for them (Aisyah et al., 2018; Nurmaliati et al., 2024). The ability to solve physics problems is the main thing that students must have as a provision to get maximum learning results (Hussin et al., 2018; Kusumaningtyas et al., 2023; Abdullah et al., 2015). The concepts received by students must be truly understood by students so that these concepts can be used in solving physics problems.

HOTS assessment is an evaluation process designed to measure students' higher-order thinking

How to Cite:
Wahyuni, V., Ratnawulan, & Mufit, F. (2025). HOTS Assessment Development to Assess High School Physics Problem-Solving Skills. *Jurnal Penelitian Pendidikan IPA*, *11*(10), 536–546. https://doi.org/10.29303/jppipa.v11i10.12880

skills, such as analyzing, evaluating, and creating. The goal is to determine the extent to which learning can improve students' higher-order thinking skills (Masitoh & Aedi, 2020; Ropi'i et al., 2025; Oktavianty et al., 2021; Zaeni et al., 2025). According to Litna et al. (2021), developing HOTS assessment instruments in physics learning can significantly improve students' problem-solving abilities.

Observations conducted during physics lessons in grade X of SMA N 4 Sungai Penuh show that students are not yet accustomed to facing questions that require higher-order thinking skills, especially in the context of problem-solving. The results of the percentage of students' problem-solving abilities based on the four indicators can be seen in the following table.

Table 1. Percentage of Observation Results of Problem Solving Ability

				Indicator
Class	Understanding the problem	Planning a solution	Problem solving	Check back
XA	67%	50%	47%	39%
XB	61%	47%	44%	36%
XC	72%	56%	53%	44%
XD	64%	53%	47%	42%
XE	69%	56%	50%	44%
XF	53%	44%	44%	33%

Based on table 1, it can be concluded that the problem-solving ability of students in class X of SMA N 4 Sungai Penuh is still at a low level. In the indicator of understanding the problem, the majority of the class shows a moderate category of achievement of 53%-72%, which indicates that students are quite capable of recognizing problem situations presented in contextual-based questions. However, in the other three indicators, namely planning solutions, problem solving, and re-examining, many students are still in the low category, especially in terms of re-evaluating solutions, an average of only 33%-44% of students show achievement. This indicates that most students have not been able to develop high-level thinking skills optimally in dealing with questions that require problem-solving abilities.

Based on the results of interviews conducted with 10th grade physics teachers at SMA N 4 Sungai Penuh, the assessments used in physics learning are still dominated by questions that measure low-level cognitive aspects, such as remembering and understanding. Most of the questions given to students only focus on the ability to remember facts, formulas, and definitions, and understand basic concepts without requiring students to conduct analysis, synthesis, or evaluation. In addition, teachers also revealed that the availability of practice questions at school is still very limited. The school has not provided a question bank or varied and challenging assessment tools to support the measurement of students' higher-order thinking skills. In practice, teachers only rely on questions available in Student Worksheets (LKS) as a source of practice.

There are two factors that influence students' problem-solving abilities: internal and external. One internal factor, or one that originates from within the student, is the student's interest in learning (Hermaini & Nurdin, 2020). Another factor is students' low ability to understand ideas or concepts in problems, and many students are unable to solve new problems that differ from the examples given during the learning process (Sari et al., 2020; Hartiani et al., 2022; Azis et al., 2025). One of the main factors contributing to this difficulty is students' unfamiliarity with problems that require in-depth analysis, resulting in difficulties in understanding the information presented in the problems (Nuryana & Rosyana, 2019; Yee et al., 2011; Sulistyorini et al., 2020; Hamidah & Wulandari, 2021; Intan et al., 2020).

Based on the background of the problem, an approach is needed that can train students to think in a more complex and systematic way, one of which is through the application of questions based on Higher Order Thinking Skills (HOTS).

Method

The development model used in this research is the Plomp model. This model consists of three stages: preliminary research, development or prototyping, and assessment (Plomp & Nieveen, 2013). The evaluation method used in this Plomp model research is as follows.

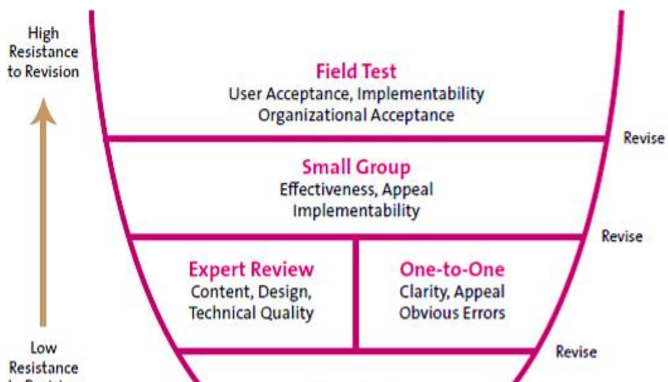

Figure 1. Formative evaluation layer

Based on Figure 1, the evaluation methods used in this study are: Self-evaluation, namely the independent assessment process carried out by researchers using a checklist based on the initial specifications of the HOTS assessment design that has been developed. Expert review, namely a study from a group of experts. One-to-

One Evaluation, namely asking for input from three students, each representing a low, medium, and high ability group, regarding the product being developed by filling out a questionnaire. Small group evaluation, namely a trial conducted on 6 students, where each ability level (low, medium, and high) is represented by 2 students each. Large group trials (field tests), including testing activities on large groups of students (36 students) using HOTS assessments to determine the practicality and effectiveness of the product.

Development Procedures

Preliminary Research Phase: To identify basic learning needs and problems. Activities carried out included document analysis and interviews with high school physics teachers regarding HOTS assessments and problem-solving skills.

Development/Prototype Phase

Prototype I	:	Developing question grids, HOTS questions, and assessment rubrics. Formative evaluation with self-evaluation by the researcher, followed by revision.
Prototype II	:	Expert validation (construction, content, language) through expert review. Then, revisions are made until it is declared valid by the experts.
Prototype III :	:	One-to-one evaluation trial with three students with varying abilities. Then, revisions were made based on student comments regarding the material, language, and presentation.
Prototype IV	:	A small group evaluation trial on six 10th grade students with varying abilities. To test its practicality before being implemented on a larger scale (field test).

Data Collection Technique

Observations were conducted to obtain direct information regarding the condition of physics learning in class X. Through observations, researchers observed student engagement, their ability to understand and solve problems, and the role of teachers in facilitating learning based on higher-order thinking skills. The results of these observations were used as a reference in designing HOTS assessments that are appropriate to student needs.

Questionnaires were used to gather respondents' opinions and experiences. The teacher questionnaire focused on assessing the practicality and effectiveness of HOTS assessments in supporting the learning process. The student questionnaire aimed to determine the ease of use, time efficiency, and benefits of HOTS assessments in learning activities.

A HOTS-based essay test is used as the primary technique to measure students' problem-solving abilities. The test is designed around four indicators: understanding the problem, planning a solution, solving the problem, and reviewing the solution. Each question is accompanied by a scoring rubric to ensure objective and systematic assessment of test results and to serve as a basis for assessing the validity, reliability, and effectiveness of the HOTS assessment developed.

Data Analysis Techniques

Data analysis techniques relate to calculations to answer the problem formulation. This study involved three stages: preliminary research analysis, instrument development analysis, and final product analysis, which included calculations of construct validity, reliability, discriminatory power, and the instrument's difficulty index.

Development Stage

Self-Evaluation

At this stage, the researcher conducts a self-assessment of the initial prototype's incompleteness and errors by examining the instrument in detail and adjusting it to the theory proposed by previous experts. The results of the completeness evaluation are then expressed using the following percentage technique:

$$\% = \frac{scores\ obtained}{total\ score}\ x\ 100\% \tag{1}$$

Expert Validity

Validity is obtained from the total number of points given by the validator minus the lowest validity assessment number (in this case 1) divided by the number of validators multiplied by the highest validity assessment number (in this case 4). Determining the value obtained with provisions based on the validity category as seen in Table 2.

$$= \frac{\sum s}{n\,(c-1)} \tag{2}$$

Table 2. Validity Categories (Azwar, 2015)

Value	Criteria
≥ 0.6	Valid
≤ 0.6	Invalid

One to One and Small Group Practicality

The practicality of HOTS assessment on renewable energy material to assess the problem-solving abilities of grade X high school students is seen from the results of the student questionnaire. The practicality test assessment questionnaire is arranged based on a Likert scale as shown in Table 3.

Table 3. Likert Scale (Retnawati, 2016)

Likert scale	Criteria
1	strongly disagree
2	Don't agree
3	Neutral
4	Agree
5	Strongly agree

The data obtained were analyzed using the following percentage values:

$$\text{Value} = \frac{Number\ of\ scores\ obtained}{Maximum\ score}\ x\ 100\% \quad (3)$$

After the percentage of practicality values is obtained, grouping is carried out according to the criteria contained in Table 4.

Table 4. Criteria for Awarding Practicality Scores (Riduwan, 2010)

Percentage (%)	Criteria
0 – 20	Not practical
21 – 40	Less practical
41 – 60	Quite practical
61 – 80	Practical
81 – 100	Very practical

Assessment Stage

Construct Validity Test

To calculate the validity of the test in this study, the following formula was used (Sukardi, 2012):

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2(N\sum Y^2 - (\sum Y^2))}} \quad (4)$$

Information: r_{xy} = Empirical validity of questions; N = Number of subjects; X = The total score for each question item for each student; Y = The total score of each student. The r_{xy} value will be compared with the r_{xy} table value with the following conditions:

Table 5. Validity Test Provisions (Sukardi, 2012)

r_{xy}	Criteria
r_{xy} Count > r_{xy} table	Valid
r_{xy} Count < r_{xy} table	Invalid

Reliability Test

$$\text{Cronbach Alpha} = \frac{k}{k-1}\left(\frac{1-\sigma_i{}^2}{\sigma_T{}^2}\right) \quad (5)$$

Information: k = Number of test items; $\sum \sigma_i{}^2$ = Number of grain variants; $\sigma_T{}^2$ = Total score variance.

Table 6. Reliability Criteria (Arikunto, 2009)

Reliability Index	Reliability Criteria
Cronbach Alpha ≥ 0.70	High reliability
Cronbach Alpha < 0.70	Low reliability

Difficulty Level Test

The difficulty index is a number that indicates the degree of difficulty of a question item. To determine the difficulty index for descriptive questions, the following formula is used:

$$TK = \frac{\bar{X}}{\text{SMI}} \quad (6)$$

Information: TK = Difficulty level index; $\bar{X}$ = Average value for each question item; SMI = Ideal maximum score.

Table 7. Difficulty Level Criteria

Difficulty Level Value	Difficulty Level Criteria
0.00 ≤ IK < 0.20	Very Difficult
0.20 ≤ IK < 0.40	Difficult
0.40 ≤ IK < 0.60	Currently
0.60 ≤ IK < 0.80	Easy
0.80 ≤ IK < 1.00	Very easy

Distinguishing Power Test

To determine discriminatory power, test takers' scores are first ranked from highest to lowest. The top 50% of scores are then taken as the top group and the bottom 50% of scores as the bottom group. To calculate the discriminatory power for essay-type questions, the following formula is used:

$$D_p = \frac{S_A - S_B}{J_A} \quad (7)$$

Information: S_A = Total group scores on an item; S_B = The sum of the scores of the lower group of an item; J_A = The ideal score for an item

Table 8. Differential Power Criteria

Distinguishing Power Value	Differential Power Criteria
0.70 < DP ≤ 1.00	Very good
0.40 < DP ≤ 0.70	Good
0.20 < DP ≤ 0.40	Enough
0.00 < DP ≤ 0.20	Bad
DP ≤0.00	Very bad

Result and Discussion

Preliminary Research Results

Results of Needs Analysis

The results of the needs analysis indicate that students' physics problem-solving abilities are still low. Students tend to be able to solve only routine and procedural problems, but experience difficulties when faced with problems that require higher-order thinking skills. Interviews with teachers and test questions also show that students often fail to understand problems, struggle to plan solution strategies, and are less thorough in double-checking their answers. This condition has implications for low levels of completion, especially on problems requiring analytical, evaluation, and creativity skills.

Furthermore, there is currently no valid, practical, and reliable assessment instrument specifically designed to assess problem-solving skills in physics learning. The instruments used are still limited to mastery of basic concepts or routine calculations, thus failing to provide a comprehensive picture of students' higher-order thinking skills. Therefore, it is necessary to develop a Higher Order Thinking Skills (HOTS)-based assessment that can both assess and encourage improved physics problem-solving abilities.

Prototyping Phase Results

Prototype Design Results

The prototype was developed based on the analysis results from the preliminary research phase, which included needs analysis. In this study, the prototype design phase focused on producing a HOTS assessment to assess physics problem-solving skills. The series of activities carried out during the prototype design phase included:

Make a Question Grid

The grid of HOTS assessment questions on renewable energy material that is designed, contains learning outcomes (CP), learning objectives (TP), learning objective achievement indicators (IKTP), material scope, HOTS indicators, problem-solving ability indicators, question formats and question numbers.

Development of Question Items

In the essay question development stage, the question formulation focuses on real-life contexts that are close to everyday life so that students are encouraged to connect physics concepts with phenomena they experience. Each question is designed to demand higher-order thinking skills, such as analyzing problems, evaluating alternative solutions, and creating appropriate solutions. Furthermore, to maintain objectivity in assessment, each question is equipped with a detailed and clear assessment rubric, so that teachers can assess students' answers based on the level of conceptual accuracy, logical reasoning, and creativity in presenting solutions.

Formative Evaluation Results and Prototype Revision

Self-Evaluation Results

The results of the self-evaluation of the HOTS assessment indicate that the developed instrument has fulfilled several important aspects. (1) In terms of suitability to objectives, all questions are relevant to physics learning outcomes and are designed to emphasize problem-solving and higher-order thinking skills. (2) In terms of HOTS indicators, the questions have included cognitive levels C4 (analyzing), C5 (evaluating), and C6 (creating) through the presentation of real, contextual phenomena. (3) The context of the questions is based on everyday problems. (4) The essay-based questions chosen allow students to provide answers with various solution strategies. (5) This instrument has also been equipped with a detailed assessment rubric so that the assessment is more objective and focused. (6) In terms of readability, the language and instructions of the questions are generally easy for students to understand.

Results of Content Validity Test by Experts (Expert Review)

The product validation instrument covers three assessment aspects, namely: (1) HOTS assessment construction aspect, (2) content/material aspect, (3) language aspect. Each aspect consists of several assessment indicators. The validation results provided by the validators in the form of suggestions and input are used as considerations in the product revision process.

Data from Content Validity Test Results Based on Question Items

Validity test data was systematically analyzed using Aiken's V formula, where each item was assessed individually according to predetermined assessment aspects. This assessment involved three main aspects, which were then broken down into 13 assessment indicators, providing a more detailed picture of the instrument's quality.

Aspects of Construction Validity

The average value of each indicator for each construct validity question can be seen in Table 9. Based on Table 9 regarding the construct validity of the HOTS assessment, it can be concluded that all test items have Aiken's V values in the range of 0.89 to 0.97. This value is above the minimum limit generally used as a reference for instrument feasibility, so each test item is declared valid. The overall average Aiken's V value is 0.93, which is included in the very high category, thus indicating that the developed assessment instrument has very good construct validity. This means that the test items in the instrument are able to consistently measure the intended construct, namely higher-order thinking skills (HOTS).

Table 9. Construct Validity of HOTS Assessment

No Question	Aiken's V value	Criteria
1	0.94	Valid
2	0.92	Valid
3	0.89	Valid
4	0.92	Valid
5	0.94	Valid
6	0.97	Valid
7	0.92	Valid
8	0.94	Valid
9	0.89	Valid
10	0.94	Valid
Average	0.93	Valid

Content Validity Aspect

The average value of each indicator for each construct validity question can be seen in Table 10.

Table 10. Content Validity of HOTS Assessment

No Question	Aiken's V value	Criteria
1	0.88	Valid
2	0.87	Valid
3	0.87	Valid
4	0.89	Valid
5	0.91	Valid
6	0.89	Valid
7	0.91	Valid
8	0.90	Valid
9	0.92	Valid
10	0.91	Valid
Average	0.89	Valid

Based on Table 10 regarding the content validity of the HOTS assessment, it can be concluded that all test items have Aiken's V values in the range of 0.87 to 0.92. This value indicates that all questions are declared valid according to the applicable criteria. The overall average Aiken's V value is 0.89, which is included in the high category, thus confirming that this assessment instrument generally has good content validity. This means that the test items are in accordance with the established indicators and measurement objectives, and are able to represent the content measured in the HOTS assessment. Thus, this instrument can be declared suitable for use because it has met the content validity criteria, so it can be trusted to accurately assess students' higher-order thinking skills.

Language Validity Aspects

The average value of each indicator for each construct validity question can be seen in Table 11. Based on Table 11 regarding the validity of the HOTS assessment language, it can be concluded that all test items obtained Aiken's V values in the range of 0.89 to 0.96. This value indicates that all questions are included in the valid category. The average Aiken's V value is 0.93, which is included in the very high category, thus proving that the overall language used in the HOTS assessment instrument is clear, precise, and easy to understand. This indicates that the wording of the test items does not cause ambiguity and is in accordance with the rules of good instrument writing.

Table 11. Validity of HOTS Assessment Language

No Question	Aiken's V value	Criteria
1	0.96	Valid
2	0.93	Valid
3	0.93	Valid
4	0.89	Valid
5	0.93	Valid
6	0.96	Valid
7	0.89	Valid
8	0.89	Valid
9	0.96	Valid
10	0.93	Valid
Average	0.93	Valid

Based on the validity analysis data for the three aspects—construct validity, content validity, and language validity—these are presented in the graph below. This makes it easier to compare the validity levels of each aspect and assess the overall feasibility of the instrument.

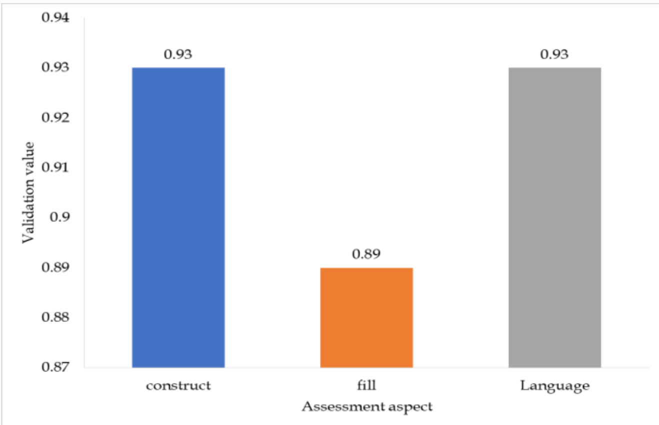

Figure 2. Results of product validity analysis

Based on the results of the HOTS assessment product validity analysis displayed in the graph, the average value obtained for the construct aspect was 0.93, the language aspect was 0.93, and the content aspect was 0.89. All of these values are in the high category, so it can be concluded that the developed assessment instrument has met the validity criteria and is suitable for use. These results indicate that, in general, the instrument is able to measure high-level thinking skills with good quality.

Practicality Test Results

One to One Practicality

The instrument used was a practicality questionnaire compiled based on five assessment aspects: clarity and ease, time, attractiveness, relevance of the material, and usefulness. The questionnaire used a Likert scale so that student responses could be expressed in the form of scores. The data obtained were then analyzed using a percentage formula to determine the practicality category of each aspect. The results of the One-to-One practicality test analysis can be seen in Table 12 below.

Table 12. Practicality of One to One HOTS Assessment

Aspect	Value	Criteria
Clarity and Ease	88.89	Very Practical
Time	86.67	Very Practical
Attractiveness	75.56	Practical
Related Materials	90.00	Very Practical
Benefit	86.67	Very Practical
Average	85.56	Very Practical

Based on Table 12 regarding the practicality of one-to-one HOTS assessment, it can be concluded that the assessment instrument has a high level of practicality with an average score of 85.56 and is in the very practical category. The assessment results show that the aspects of clarity and ease (88.89), time (86.67), relevance of material (90.00), and benefits (86.67) are all in the very practical category, which means the instrument is easy to understand, efficient to use, relevant to the material, and beneficial for students. Meanwhile, the attractiveness aspect obtained a score of 75.56 with a practical category.

Small Group Practicality

This stage was conducted at SMAN 4 Kota Sungai Penuh, involving two high-ability students, two medium-ability students, and two low-ability students. The questionnaire used a Likert scale so that student responses could be expressed in the form of scores. The data obtained were then analyzed using a percentage formula to determine the practicality category of each aspect. The results of the small group practicality test analysis can be seen in Table 13 below.

Table 13. Small Group Practicalities

Aspect	Value	Criteria
Clarity and Ease	90.00	Very Practical
Time	88.33	Very Practical
Attractiveness	85.56	Practical
Related Materials	91.67	Very Practical
Benefit	88.33	Very Practical
Average	88.78	Very Practical

Based on Table 13 regarding the practicality of the HOTS small group assessment, the average score was 88.78, categorized as very practical. The analysis results show that most aspects, namely clarity and ease (90.00), time (88.33), relevance of material (91.67), and benefits (88.33) are in the very practical category, meaning the instrument is easy to use, efficient in implementation, relevant to the material, and beneficial for learning. Meanwhile, the attractiveness aspect obtained a score of 85.56 in the practical category.

Assessment Phase Results

Field Test Results

A field test was conducted to verify the feasibility of the Higher Order Thinking Skills (HOTS) assessment instrument developed to assess high school students' physics problem-solving abilities. The trial was conducted at SMAN 4 Kota Sungai Penuh, involving students as respondents to obtain empirical data on the quality of the instrument.

Validity Test

The test results for item validity were conducted by comparing the calculated R value with the R table value at a certain significance level. An item is declared valid if the calculated R value is greater than the R table value, indicating that the item is capable of measuring the aspect being measured. Furthermore, the interpretation of validity is also determined based on the validity level category according to the applicable criteria. The results of the HOTS assessment item validity analysis on the renewable energy material that has been developed can be seen in Table 14 below.

Table 14. Results of HOTS Assessment Validity Test

No Question	R_{count}	R_{table}	Criteria
1	12.85439	1.688298	Valid
2	13.45331	1.688298	Valid
3	13.67075	1.688298	Valid
4	16.27349	1.688298	Valid
5	15.78778	1.688298	Valid
6	16.35256	1.688298	Valid
7	17.67379	1.688298	Valid
8	14.30155	1.688298	Valid
9	12.48131	1.688298	Valid
10	15.03031	1.688298	Valid

Based on the validity test results shown in Table 19, it was found that all HOTS assessment questions had a calculated R value that was much greater than the R table value (1.688298). The calculated R values of the questions ranged from 12.48131 to 17.67379 so that all questions were declared valid. Validation aims to determine the feasibility of the product developed before being implemented to students (Harahap et al., 2022). Research by Rohmatika et al. (2025) showed that the creative thinking skills test instrument on renewable energy material was declared quite valid and reliable based on the Rasch model analysis, although there were several questions that needed to be revised due to extreme logit and suboptimal indicators. Similar results were found by Rahmatillah et al. (2025) who evaluated the HOTS instrument on the topic of renewable energy, where most of the questions met good psychometric criteria with a normal distribution of student abilities, although the level of difficulty of the questions was still uneven. Furthermore, Diansah et al. (2023) and Bakhti et al. (2025) through the development of STEM-based multimedia on the topic of renewable energy also reported that expert-validated products were declared valid, practical, and effective in stimulating students' higher-order thinking skills, especially the aspects of analysis, evaluation, and creation.

Reliability Test

The reliability test in this study used the Cronbach's Alpha coefficient, where an instrument is considered reliable if the Cronbach's Alpha value is greater than 0.70. The results of the HOTS assessment reliability test with 10 questions are presented in Table 15 below.

Table 15. Results of HOTS Assessment Reliability Test

Cronbach's Alpha	N of item	Criteria
0.98108244	10	Realiable

Based on the reliability test results presented in Table 20, a Cronbach's Alpha value of 0.98 was obtained for 10 items. This value far exceeds the minimum required reliability limit of 0.70 (Menon et al., 2025). Thus, it can be concluded that the developed HOTS assessment instrument has very strong internal consistency.

Difficulty Level Test

The difficulty level test was conducted to determine the extent to which HOTS assessment items were easy or difficult for students to complete. The difficulty level was calculated based on the proportion of students who were able to answer the questions correctly. The results of the HOTS assessment item difficulty analysis for renewable energy can be seen in Table 16 below.

Table 16. HOTS Assessment Difficulty Level Test Results

No Question	TK	Criteria
1	0.68	Easy
2	0.65	Easy
3	0.50	Currently
4	0.63	Easy
5	0.52	Currently
6	0.51	Currently
7	0.52	Currently
8	0.62	Easy
9	0.53	Currently
10	0.68	Easy

Based on the analysis of question difficulty levels, it was found that of the 10 HOTS assessment items, 5 were categorized as easy (numbers 1, 2, 4, 8, and 10) with TK scores between 0.62 and 0.68, and 5 other questions were categorized as medium (numbers 3, 5, 6, 7, and 9) with TK scores between 0.50 and 0.53.

Research by Saepuzaman et al. (2022) found that most items in the HOTS basic physics test had a difficulty index in the medium category. Therefore, the distribution of difficulty levels in this instrument is consistent with the findings of other relevant studies and is adequate for measuring high school students' physics problem-solving abilities in renewable energy.

Differential Power Test

The item discrimination test was conducted to determine the ability of an item to differentiate between students with high ability and students with low ability. The results of the HOTS assessment item discrimination test can be seen in Table 17 below.

Table 17. Results of the HOTS Assessment Distinction Power Test

No Question	D	Criteria
1	0.43	Good
2	0.47	Good
3	0.41	Good
4	0.52	Good
5	0.41	Good
6	0.41	Good
7	0.41	Good
8	0.47	Good
9	0.42	Good
10	0.49	Good

Based on the results of the item discrimination analysis in Table 17, it is known that all HOTS assessment items have discrimination values ranging from 0.41 to 0.52, which is considered good. This indicates that each item is able to effectively differentiate between high-ability students and low-ability students.

A test item is considered to meet quality criteria if it falls into the adequate, good, or excellent categories (Warju et al., 2020). Conversely, a test item with low discriminating power is unsuitable for use because it fails to accurately differentiate student abilities. Factors contributing to low discriminating power include bias in the test item, excessive difficulty, or poorly functioning distractors. Other research has shown similar results, suggesting that poor discriminating power can arise from questions that are too easy or too difficult, leading to bias. This condition results in the inability to accurately identify students who have understood the material and those who have not (Mulyani et al., 2022).

Conclusion

Based on the research results, the following conclusions can be drawn: The HOTS assessment development procedure for assessing high school students' physics problem-solving abilities was conducted using the Plomp development model, which includes an initial investigation stage (needs analysis and problem identification), a design stage (instrument design), a realization or construction stage (HOTS question development), a testing, evaluation, and revision stage (expert validation and limited trials), and an implementation stage (wider trials). This procedure resulted in a systematic and contextual assessment instrument. The characteristics of the HOTS assessment instrument developed include real-world context-based stimuli, varying levels of difficulty (easy to medium), and higher-order thinking indicators at the analysis (C4), evaluation (C5), and creation (C6) levels. The instrument's validity and reliability demonstrated good results. All test items were declared valid based on expert testing and empirical analysis, and were reliable, with Cronbach's Alpha values far exceeding the minimum required reliability limit. In addition, the discrimination power of the questions is included in the good category, so that the instrument is able to differentiate students with high and low abilities effectively.

Acknowledgments
We would like to thank the school leaders and teachers of SMA Negeri 4 Kota Sungai Penuh were willing to be informants in our research. We also thank Padang State University for supporting this research.

Author Contributions
The authors in this research are divided into executor and advisor.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest in this research.

References

Abdullah, A. H., Abidin, N. L. Z., Ali, M. (2015). Analysis of Students' Errors in Solving Higher Order Thinking Skills (HOTS) Problems for Topic of Fraction. *Asian Social Science*, *11*(21), 133-142. https://doi.org/10.5539/ass.v11n21p133

Aisyah, P. N., Nilam, N., Akbar, P., & Anik, Y. (2018). Analisis Hubungan Kemampuan Pemecahan Masalah Matematis dan Self Confidence Siswa SMP. *Journal On Education P*, *1*(1), 58–65. https://doi.org/10.31004/joe.v1i1.11

Aina, Q., & Hariyono, E. (2023). Penerapa PhET Simulations pada Pembelajaran Fisika untuk Meningkatkan Kemampuan Literasi Sains Peserta Didik SMA Kelas X. *Jurnal Ilmu Pendidikan dan Pembelajaran*, *1*(2), 56-65. https://doi.org/10.58706/jipp.v1n2.56-65

Azis, A., Rahim, I., & Herdiani, R. (2025). Comparison of HOTS Integrity Quality in Indonesian Language Textbooks for Senior High Schools: Evaluation Based on the 2013 Curriculum and the Independent Curriculum. *Indonesian Language Education and Literature*, *10*(2), 483. https://doi.org/10.24235/ileal.v10i2.20244

Sukardi. (2012). *Metodologi Penelitian*. Jakarta: PT.Bumi Aksara.

Arikunto, S. (2009). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Aprilia, D. A., Miftahul, Nuraini, L., & Sedayu, A. (2022). Efektivitas Model Pembelajaran Problem Based Learning dengan Pendekatan Scientific Berbantuan PhET Simulations terhadap Hasil Belajar Siswa. *Jurnal Pendidkan Fisika Undhiksa*, *12*(2), 176–180. https://doi.org/10.20527/jipf.v7i1.7473

Azwar, S. (2015). *Metode Penelitian*. Yogyakarta: Pustaka Pelajar

Bhakti, Y. B., Arthur, R. & Supriyanti, Y. (2025). Development of Physics HOTS Asswssment Intruments in High School: A Comprehensive Literature Review. *Jurnal Ilmu Pendidikan Fisika*, *10*(2), 194-207. https://doi.org/10.26737/jipf.v10i2.5962

Diansah, I., Suyatna, A., & Viyanti, V. (2021). STEM-Based Multimedia Design for Stimulating HOTS on Water and Wind Energy Topic Physics Teacher Perception. *Journal of Physics Conference Series*, *1796*(1), 012002. https://doi.org/10.1088/1742-6596/1796/1/012002

Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan Instrumen Penilaian Berbasis HOTS Menggunakan Aplikasi QUIZIZZ. *Efisiensi: Kajian Ilmu Administrasi*, *18*(1), 105-124. https://doi.org/10.21831/efisiensi.v18i136997

Harahap, H. T., Mushlihuddin, R., & Nurafifah. (2022). Pengembangan Bahan Ajar Berbasis Masalah Terhadap Kemampuan Berpikir Kreatif Matematis. *Jurnal EduTech*, *8*(1), 1–9. Retrieved from https://media.neliti.com/media/publications/377003-none-4e494300.pdf

Hartiani, B. S., Rokhmat, J., & Taufik M. (2022). Validitas Perangkat Pembelajran Model Kausalitik Berbantuan Google Classroom untuk Meningkatkan Kemampuan Pemecahan Masalah dan Berpikir Kreatif Fisika Peserta Didik. *Jurnal Ilmiah Pendidikan Fisika*, *6*(2), 393. https://doi.org/10.20527/jipf.v6i2.4992

Hermaini, J., & Nurdin, E. (2020). Bagaimana Kemampuan Pemecahan Masalah Matematis Siswa dari Perspektif Minat Belajar? *JURING (Journal for Research in Mathematics Learning)*, *3*(2), 141. https://doi.org/10.24014/juring.v3i2.9597

Hussin, W. N. T. W., Harun, J., & Shukor, N. A. (2018). Problem Based Learning to Enhance Students Critical Thinking Skill via Online Tools. *Asian Social Science*, *15*(1), 14. https://doi.org/10.5539/ass.v15n1p14

Intan, F. M., Kuntarto, E., Alirmansyah, A. (2020). Kemampuan Siswa dalam Mengerjakan Soal HOTS (Higher Order Thingking Skills) pada Pembelajaran Matematika di Kelas V Sekolah Dasar. *Jurnal Pendidikan Dasar Indonesia*, *5*(1), 6-10. https://doi.org/10.26737/jpdi.v5i1.1666

Kusumaningtyas, D. A., Manyunu, M., Kurniasari, E., Awalin, A. N., Rahmaniati, R., & Febriyanti, A. (2023). Enhancing Learning Outcomes: A Study on Development of Higer Order Thingking Skills Based Evaluation Instrument for Work and Energy in High School Physics. *Indonesian Journal on Learning and Advanced Education*, *6*(1), 14-31. https://doi.org/10.23917/ijolae.v6i1.23125

Litna, K.O., Mertasari, N.M.S., & Sudirtha, G. (2021). Pengembangan Instrumen Tes Higher Order Thinking Skills (Hots) Matematika Sma Kelas X. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *11*(1), 10–20. https://doi.org/10.23887/jpepi.v11i1.278

Mahardika, I. K., Handono, S., Ernasari, Rofida, H. A., Zahro, F., & Seftiyani, M. A. (2023). Hakikat Fisika sebagai Pilar Kehidupan. *Jurnal Pendidikan Ilmu Transformatif*, *7*(12), 30–34. Retrieved from https://edu.ojs.co.id/index.php/jpit/article/view/103/243

Masitoh, L. F., & Aedi, W. G. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) Matematika di SMP Kelas VII. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *4*(2), 886–897. https://doi.org/10.31004/cendekia.v4i2.328

Menon, V., Grover, S., Gupta, S., Indu, P. V., Chacko, D., & Vidhukumar, K. (2025). A primer on reliability testing of a rating scale. *Indian Journal of Psychiatry*, 67(7), 725–729. https://doi.org/10.4103/indianjpsychiatry_584_25

Mulyani, S., Krismonita, M., & Yamtinah, S. (2022). Analisis Butir Soal dan Kecukupan HOTS Soal Ujian Akhir Semester Mata Pelajaran Kimia SMK Kelas X. *Paedagogia*, *25*(2), 162. https://doi.org/10.20961/paedagogia.v25i2.60913

Nurmaliati, N., Ilwandri, I., & Santosa, T. A. (2024). Development of Instrumental for Assessing Higher-Order Thinking Skills (HOTS) Physics in High School: Literature Review. *IndonesiaJournal of Engineering and Education Technology*, 2(2), 378-384. https://doi.org/10.61991/ijeet.v2i2.75

Nuryana, D., & Rosyana, T. (2019). Analisis Kesalahan Siswa SMK dalam Menyelesaikan Soal Pemecahan Masalah Matematik pada Materi Program Linear. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *3*(1), 11–20. https://doi.org/10.31004/cendekia.v3i1.74

Oktavianty, E., Haratua, T. M.S., & Supriadi, Y. (2021). Physics Higher Order Thingking Skills Analysis: Motion at Secondary School. *Budapest International Research and Critics Institute-Journal*, *4*(4), 13103-13110. https://doi.org/10.33258/birci.v4i4.3374

Plomp, T., & Nieveen, N. (2013). Educational Design Research Educational Design Research. *Netherlands Institute for Curriculum Development: SLO*, 1–206. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ815766

Rahmatillah, N., & Zainuddin, A. (2025). Evaluation of HOTS Test on Renewable Energy Materials through Rasch Model Analysis. *Jurnal Pijar MIPA*, *20*(5), 933–939. https://doi.org/10.29303/jpm.v20i5.9458

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. The Development of an Instrument to Measure the Higher Order Thingking Skill in Physics. *European Journal of Education Research*, *8*(3), 743-751. https://doi.org/10.12973/eu-jer.8.3.743

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing.

Riduwan. (2010). *Skala Pengukuran Variabel-Variabel Penelitian*. Bandung: Alfabeta.

Rohman, N., & Hendra, S. H. (2023). Peran Pendidikan dalam Pengembangan Keterampilan Abad Ke-21: Tinjauan Literatur Tentang Kurikulum dan Metode Pengajaran. *ABDAU: Jurnal Pendidikan Madrasah Ibtidaiyah*, *6*(2), 133–149. Retrieved from https://jurnal.stpi-bim.ac.id/index.php/abdau/article/view/267/151

Rohmatika, R., Fatmawati, S., & Suhartono, S. (2025). Analisis butir soal instrumen tes keterampilan berpikir kreatif pada materi energi terbarukan menggunakan rasch model. *Jurnal EDUCATIO: Jurnal Pendidikan Indonesia*, *11*(1), 92. https://doi.org/10.29210/1202525556

Ropi'i, N., Nugroho, S. E., & Ellianawati, E. (2025). Analysis of Physics Problem Solving Skills of Junior High School Students Through PBL-HOTS on Magnitism Material. *Physics Communication*, *9*(1), 27-39. https://doi.org/10.15294/pc.v9i1.1821

Rusmin, L., Misrahayu, Y., Pongpalilu, F., & Dwiyanto, D. (2024). Critical Thinking and Problem- Solving Skills in the 21st Century. *Journal of Social Science*, *1*(5), 144-162. https://doi.org/10.59613/svhy3576

Saepuzaman, D., Istiyono, E., & Haryanto. (2022). Characteristics of Fundamental Physics Higher-order Thinking Skills Test Using Item Response Theory Analysis. *Pegem Egitim ve Ogretim Dergisi*, *12*(4), 269–279. https://doi.org/10.47750/pegegog.12.04.28

Sari, Y., Rokhmat, J., & Hikmawati. (2020). Pengaruh Model Pembelajaran Kausalitik Terhadap Kemampuan Pemecahan Masalah Fisika Peserta Didik. *GeoScienceEdu Journal*, *1*(1), 11–16. https://doi.org/10.29303/goescienceedu.v1i1.37

Sukardi. (2012). *Metodologi Penelitian*. Jakarta: PT.Bumi Aksara.

Sulistyorini, Y., Napfiah, S., Yazidah, N. I. Argarini, D. F., Listiani, W. (2020). Profil Higher Order Thingking Skills Mahasiswa dalam Memecahkan Masalah Geometri. *Jurnal Review Pembelajaran Matematika*, *5*(2), 87-97. https://doi.org/10.15642/jrpm.2020.5.2.87-97

Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *17*(1), 95. https://doi.org/10.23887/jptk-undiksha.v17i1.22914

Yee, H. M., Othman, W., Yunos, J. M., Tee, T. K. (2011). The Level of Marzano Higher Order Thinking Skills Among Technical Education Students. *International Journal of Social Science and Humanity*, *1*(2), 121-125. https://doi.org/10.7763/IJSSH.2011.V1.20

Zaeni, A., Kartono, K., & Pujiastuti, E. (2025). HOTS Math Analysis in Tpack-Based Flip Classroom Learning Using Dynamic Assessment. *Cypriot Journal of Educational Sciences*, *20*(3), 163-174. https://doi.org/10.18844/cjes.v20i3.8525