THE DEVELOPMENT OF AN ANALYZING ABILITY TEST INSTRUMENT FOR HIGH SCHOOL STUDENTS ON THERMODYNAMICS

Febrina Rosa Winda[1], Fadillah Rahmayani[2], Rahmi Putri Z[3], Suharli AJ[4]

[1,2,3,4]*Universitas Jambi*
E-mail : [1]febrinarosawinda@unja.ac.id, [2]fadillahrahmayani@unja.ac.id, [3]rahmiputri.z@unja.ac.id, [4]suharliaj@unja.ac.id

ABSTRACT

The ability to analyze is a cognitive process that involves decomposing materials into their constituent parts and determining the relationships among those parts and the material as a whole. As a foundational component of Higher Order Thinking Skills (HOTS), the ability to analyze must be adequately assessed using valid and reliable instruments. This study aimed to develop a test instrument for measuring students' analyzing ability on thermodynamics material. Instrument development followed the model by Oriondo and Antonio (1998), which consists of three stages: (1) initial test development, (2) test trials, and (3) interpretation of test results. Instrument validation included both expert validation and empirical validation. Expert validation was reviewed from the aspects of content, construction, and language. Empirical validation involved the analysis of reliability, item validity, level of difficulty, and discrimination index. Based on expert validation, an overall Aiken's V score of 0.93 was obtained, which falls in the very valid category. Empirical validation yielded a Cronbach's Alpha value of 0.721, indicating a good and acceptable level of reliability. The final instrument consists of 12 valid items that are suitable for measuring students' analyzing ability (C4) on thermodynamics material.

Keywords : *Analyzing Ability; Higher Order Thinking Skills; Thermodynamics*

INTRODUCTION

The development of modernization and globalization in the 21st century has a tremendous impact. One of the concerning impacts is the inability of children independently to know, understand and overcome the problems that exist around them. Anticipating these demands, education is designed to improve high-quality performance through the learning process (Fransiska et al., 2021). Education is required to be able to produce quality students who are able to compete in the 21st century and industrialization 4.0 in the era of globalization. This condition needs to be solved immediately, one of which is learning and assessing Higher Order Thinking Skills (HOTS) (Kristiyono, 2018; Wartono et al., 2018). HOTS learning and assessment produce children who are able to remember, understand and apply their knowledge to solve problems.

Thinking ability is a reflective, critical, and creative reasoning activity oriented towards an intellectual process that involves concept formation, application, analysis, assessing information collected or generated through observation, experience, reflection, communication as a basis for beliefs and actions (Iskandar, 2009). Thinking skills are needed in solving problems, especially in physics (Winarti, 2015). Karthworl and Anderson stated that in the revised Bloom's Taxonomy there are three cognitive levels that measure HOTS, namely C4 (ability to analyze), C5 (ability to evaluate), and C6 (ability to create) (Liana et al., 2018). One form of high-level thinking skills that must be supplied is the skill of analyzing problems.

Analyzing is a process that involves parts and structures of thought as a whole to solve problems. Analyzing ability is a basic ability for students to think High Order Thinking and one of the main abilities that must be developed (Irawati & Mahmudah, 2018). Analyzing is an attempt to break down a material into its constituent parts and determine the relationship between these parts and the material as a whole. In this category there are three sub-categories, namely distinguishing, organizing and attributing (Nurhayati & Angraeni, 2017). Analyzing is a process that involves parts and structures of thought as a whole to solve problems (Rochman & Hartoyo, 2018). Analyzing includes cognitive processes of attributing, distinguishing, and organizing. Another name for analyzing is classifying. In classifying activities, students will be more detailed in explaining certain concepts, characteristics, and descriptions that are material or real. This skill is very important for students considering that they now live in an increasingly complex world (Lestari et al., 2019). Based on the description above, the ability to analyze is needed in physics learning.

The ability to analyze is one of the important skills in learning physics. This is because physics is an analytical science and requires a deep understanding of related concepts and principles. Some of the roles of analyzing skills in physics learning include the following: (1) Understand physics concepts more deeply; (2) Improve problem solving skills; and (3) Become the basis for conducting experiments (Docktor et al., 2015; Gök & Sýlay, 2010; Kurniawan & Taqwa, 2018). Therefore, the ability to analyze can be said to be a very important ability to have during the physics learning process. This is related to the character of physics learning which is closely related to physics learning ability. One of the physics materials that has many applications found in everyday life is thermodynamics.

Physics learning at school aims to develop students' ability to analyze problems and develop physics concepts. Educational success goes well if it is supported by good assessment and has an impact on the next learning process (Desiriah & Setyarsih, 2021). The success of an education can be known by evaluation. One way to find out whether students already have higher-

level thinking skills is by conducting an assessment (Liana et al., 2018). To conduct an assessment, a valid and reliable instrument is needed to measure students' abilities. Therefore, it is important to develop an instrument that can measure students' ability to analyze physics problems, which in this case is specific to thermodynamic material.

The development of this instrument is important to assist teachers in evaluating students' abilities. Thus, teachers can more easily evaluate student progress and overcome weaknesses that may appear in physics learning. In general, the development of instruments for measuring analytical skills in physics learning aims to provide reliable and valid evaluation tools to teachers, so that they can evaluate students' abilities as effectively as possible. Instruments in the form of tests can be used to hone students' thinking abilities, and are influential in determining students' thinking skills. If students are expected to think at a high level, then the test assessment used must represent students' higher-level thinking skills, where the assessment instrument can be in the form of higher-level thinking test questions, meaning that these types of questions are an instrument that can measure students' higher-level thinking skills. This is reinforced by the statement that questions used to test higher-order thinking skills can encourage students to think deeply about a material (Barnett & Francis, 2012).

Based on the description above, research was conducted to develop an analyzing test instrument on thermodynamic material for high school students. The purpose of this research is to develop a valid and reliable test instrument on thermodynamic material. In this development research, an analysis skill test instrument was made with indicators consisting of the ability to distinguish, attribute, and organize.

METHOD

This research aims to develop an instrument for measuring the ability to analyze students on thermodynamic. Instrument development was carried out based on the development model by Oriondo and Antonio (1998) which consists of three stages, namely (1) initial test development, (2) test trials, and (3) interpretation of test results.

The test subjects in this study were students of class XI MIPA SMA Negeri 1 Kota Sungai Penuh. One class of students was selected by cluster sampling to become a test subject totaling 35 students. In this trial, students were asked to work on thermodynamic material in the form of multiple-choice questions through google form. As for the interpretation stage of the test results, an analysis is carried out to determine the quality of the instrument empirically based on the results of the trial including reliability and item characteristics consisting of difficulty level, differentiating power, and item validity.

The data obtained through expert validation, empirical validation, and implementation consist of two types, namely qualitative and quantitative data. Qualitative data is in the form of suggestions from experts collected as consideration in improving the analyzing ability test instrument. Quantitative data in the form of test instrument assessment scores obtained from expert validation and test item characteristics and reliability obtained from empirical validation.

1. Expert Validation

Aiken's V formula is used to calculate the content-validity coefficient which is based on the results of the assessment of a panel of n experts on an item in terms of the extent to which the item represents the construct being measured. The V index value ranges from 0-1. From the calculation of the V index, an item or device can be categorized based on its index.

Table 1. Interpretation of Aiken's V Index

Aiken V index	Interpretation
V < 0,4	Invalid
0,4 < V < 0,8	Moderately Valid
V > 0,8	Very Valid

2. Empirical Validation

Empirical data from test trials were analyzed to determine item validity, reliability, difficulty level, and discrimination index. The validity test of multiple-choice items was carried out with the Point Biserial validity test using SPSS. Based on the results of SPSS output, information related to valid items and invalid items was obtained. An item is declared valid if $R_{count} > R_{table}$. Testing the reliability of test instruments that have been developed using Alpa Cronbach. Furthermore, the reliability results obtained were analyzed quantitatively and the reliability criteria were determined based on the interpretation of the value. Reliability interpretation can be seen in table below.

Table 2. Reliability Interpretation Criteria

Aiken V index	Interpretation
0,91 – 1,00	Excellent
0,81 – 0,90	Good
0,71 – 0,80	Good and Acceptable
0,61 – 0,70	Acceptable
0,01 – 0,60	Not Acceptable

To measure the level of difficulty is done by calculating the average score for each item and then proceed with calculating the level of difficulty. Analyzing the level of difficulty of the question is to determine the quality of a good question, knowing the classification of easy, medium and difficult questions. The formula used for objective questions is as follows:

$$P = \frac{R}{T}$$

Where P is the item difficulty index, R is the number of correct responses and T is the total number of responses (which includes both correct and incorrect responses). Interpretation of the results of the calculation of the question difficulty index number using the following criteria.

Table 3. Interpretation of Difficulty Level

Aiken V index	Interpretation
0,20 – 0,30	Very Difficult
0,30 – 0,40	Difficult
0,40 – 0,60	Moderate
0,60 – 0,70	Easy
0,70 – 0,80	Very Easy

The item discrimination is the ability of a question item to distinguish between learners/students who have mastered the material in question and learners/students who do not/do not master the material in question. Index of discrimination is that ability of an item on basis of which the discrimination is made between superiors and inferiors. To find out the item discrimination of multiple-choice tests is to use the following formula:

$$D = \left(\frac{RH}{NH}\right) - \left(\frac{RL}{NL}\right)$$

Where D is iscrimination index, RH is rightly answered in highest group, RL is rightly answered in lowest group, NH is number of examinees in highest group, and NL number of examinees in highest group. The criteria for the differentiating power index of the questions were classified as shown in the table below.

Table 4. Interpretation of Item Discrimination Index

Aiken V index	Interpretation
0,40 - 1,00	Very Good
0,30 - 0,39	Reasonably Good (needs minor improvement)
0,20 - 0,29	Marginal (needs improvement)
0,19 - 0,00	Poor (rejected or revised)

RESULTS AND DISCUSSIONS

Instrument development was carried out based on the development model by Oriondo and Antonio (1998) which consists of three stages, namely (1) initial test development, (2) test trials, and (3) interpretation of test results. The results of each development stage are described in the following description.

In the initial test development stage, the instrument was designed to measure the analyzing ability (C4) as defined in Bloom's revised taxonomy, specifically targeting the topic of thermodynamics at the Grade XI senior high school level.

The test consisted of multiple-choice items with five answer options. A score of 1 was assigned for a correct answer and 0 for an incorrect answer. Items were developed to address three dimensions of analyzing ability: differentiating, attributing, and organizing. The blueprint of the instrument is presented in Table 5.

Table 5. Blueprint of the Analyzing Ability Test Instrument

Cognitive process dimension	Indicator	Item Numbers
Differentiating	Able to identify relevant pieces of information	1,5,7,10,13
Attributing	Able to determine the purpose and relationships among presented information	2,3,6,8,14,15
Organizing	Able to organize relevant pieces of information into a coherent structure	4,9,11,12

The 15 completed items were then submitted for expert validation by four raters. Validation covered three aspects: (1) content, (2) construction, and (3) language. According to Aiken's V table (1985), for a four-point rating scale with four raters, the minimum acceptable validity coefficient is 0.75.

Table 6. Expert Validation Results

Item	Rater				Aiken's V	Category
	1	2	3	4		
Item 1	4	4	4	4	1,00	Very valid

Item 2	3	3	4	3	0,75	Moderate valid
Item 3	4	4	4	4	1,00	Very valid
Item 4	4	4	4	4	1,00	Very valid
Item 5	4	3	4	3	0,83	Moderate valid
Item 6	4	4	4	4	1,00	Very valid
Item 7	4	4	4	4	1,00	Very valid
Item 8	4	4	4	4	1,00	Very valid
Item 9	4	4	4	4	1,00	Very valid
Item 10	4	4	4	4	1,00	Very valid
Item 11	4	4	4	4	1,00	Very valid
Item 12	3	3	3	3	0,67	Moderate valid
Item 13	4	4	4	4	1,00	Very valid
Item 14	4	2	4	3	0,75	Moderate valid
Item15	4	4	4	4	1,00	Very valid
				Overall Aiken's V Score	**0,93**	

Based on the results of expert validation presented in Table 6, it is obtained that there are four items that have a validity index smaller than 0.92, namely items 2, 5, 12, and 14. These items must first make revisions according to the validator's suggestions before being used to collect data. Overall, the instrument to measure the ability to analyze on thermodynamics has an index of 0.93 so that it is in the good category. The suggestions obtained from experts are to make improvements in terms of grammar and formula writing. The suggestions obtained were then used as a consideration in revising the analyzing ability test instrument.

Following expert validation and revision, the instrument was administered to 35 Grade XI MIPA students of SMA Negeri 1 Kota Sungai Penuh, selected through cluster sampling. Students completed the 15-item thermodynamics test via Google Form. The trial data were subsequently analyzed to assess item validity, reliability, difficulty level, and discrimination index.

Item validity was evaluated using the Point Biserial correlation test in SPSS. Based on the results of SPSS output, information related to valid items and invalid items was obtained. An item is declared valid if $R_{count} > R_{table}$. SPSS output for the Point Biserial validity test using 15 multiple choice items is presented in the following table.

Table 7. Point Biserial Validity Results

Item	R_{count}	R_{tabel} (N=35, Sig. 5%)	Interpretation
Item 1	0,464	0,334	Valid
Item 2	0,472		Valid
Item 3	0,575		Valid
Item 4	0,505		Valid
Item 5	0,618		Valid
Item 6	0,511		Valid
Item 7	0,077		Invalid
Item 8	0,224		Invalid
Item 9	0,504		Valid
Item 10	0,564		Valid

Item 11	0,232	Invalid
Item 12	0,422	Valid
Item 13	0,357	Valid
Item 14	0,710	Valid
Item15	0,699	Valid

Based on the results of Point Biserial validity, three items were declared invalid, namely item 7, item 8, and item 11. These items were excluded from further analysis. Reliability was then computed using Cronbach's Alpha based on the remaining 12 valid items, with results presented in Table 8.

Table 8. Reliability Analisis Output

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Item	N of items
0,721	0,783	12

The Cronbach's Alpha value of 0.721 falls within the "Good and Acceptable" category, indicating that the instrument has adequate internal consistency. The difficulty level and discrimination index for all 15 items are presented in Table 9.

Table 9. Item Difficulty Level and Discrimination Index

Item	Difficulty Level Index	Cat	Discrimination Index	Cat
Item_1	0,66	Difficult	0,60	Very good
Item_2	0,29	Most difficult	0,50	Very good
Item_3	0,77	Most easy	0,50	Very good
Item_4	0,86	Most easy	0,50	Very good
Item_5	0,23	Most difficult	0,50	Very good
Item_6	0,71	Most easy	0,50	Very good
Item_7	0,94	Most easy	0,10	Rejected
Item_8	0,97	Most easy	0,10	Rejected
Item_9	0,66	Easy	0,60	Very good
Item_10	0,66	Easy	0,70	Very good
Item_11	0,57	Moderate difficult	0,40	Very good
Item_12	0,26	Most difficult	0,40	Very good
Item_13	0,26	Most difficult	0,20	Revise
Item_14	0,40	Moderate difficult	0,90	Very good
Item_15	0,43	Moderate difficult	0,70	Very good

The item difficulty analysis shows a distribution across categories. Items 3, 4, 6, 7, and 8 are classified as very easy; Items 1, 9, and 10 as easy; Items 11, 14, and 15 as moderate; and Items 2, 5, 12, and 13 as very difficult. This variety in difficulty levels indicates that the instrument is capable of differentiating student performance across different levels of ability. With regard to discrimination, most items fall in the very good category ($D \geq 0.40$). Item 13 is categorized as marginal and requires revision, while Items 7 and 8 had poor discrimination indices ($D = 0.10$)

and are recommended for rejection, consistent with their item validity results. The final instrument, after removing invalid items (7, 8, and 11), consists of 12 items that are declared valid, reliable, appropriately distributed in difficulty, and sufficiently discriminating for measuring students' analyzing ability (C4) on thermodynamics material. In order to measure the ability to analyze (C4), the items developed consist of the dimensions of the cognitive process of differentiating, organizing, and attributing. The resulting test instrument is in the form of multiple-choice questions with five alternative answers. Based on the results of content validity and empirical validity, it is known that the instrument can be used to measure the ability to analyze (C4) on thermodynamic material.

CONCLUSIONS

This study developed a test instrument for measuring the analyzing ability (C4) of Grade XI senior high school students on thermodynamics material. The instrument was developed using the Oriondo and Antonio (1998) development model. Expert validation yielded an overall Aiken's V score of 0.93, indicating that the instrument is very valid in terms of content, construction, and language. Empirical validation with 35 students produced a Cronbach's Alpha of 0.721, categorized as good and acceptable. Item analysis identified 12 valid items with appropriate difficulty distributions and good discriminating power. These items are declared suitable for use in measuring students' analyzing ability on thermodynamics.

It is recommended that future studies explore the use of this instrument with larger and more diverse sample groups, and that additional item formats (e.g., scenario-based or context-rich items) be developed to further assess higher-order thinking in physicss.

REFERENCES

Barnett, J. E., & Francis, A. L. (2012). Educational Psychology : An International Journal of Experimental Using higher order thinking questions to foster critical thinking : a classroom study. *An International Journal of Experimental Educational Psychology*, *32*(2), 201–2011. http://www.tandfonline.com

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

Docktor, J. L., Strand, N. E., Mestre, J. P., & Ross, B. H. (2015). Conceptual problem solving in high school physics. *Physical Review Special Topics - Physics Education Research*, *11*(2), 1–13. https://doi.org/10.1103/PhysRevSTPER.11.020106

Fransiska, A., Prasetyo, E., & Jufriansah, A. (2021). Desain LKPD Fisika Terintegrasi HOTS Untuk Meningkatkan Kemampuan Berpikir Kritis Peserta Didik. *Jurnal Pendidikan Fisika Dan Teknologi*, *7*(2), 153–158. https://doi.org/10.29303/jpft.v7i2.3098

Gök, T., & Sýlay, I. (2010). The Effects of Problem Solving Strategies on Students' Achievement, Attitude and Motivation. *Latin-American Journal of Physics Education*, *4*(1), 7–21. http://dialnet.unirioja.es/servlet/articulo?codigo=3694877

Irawati, T. N., & Mahmudah, M. (2018). Pengembangan Instrument Kemampuan Berpikir Analisis Siswa SMP dalam Menyelesaikan Soal Pemecahan Masalah Matematika.

Kadikma, *9*(2), 1–11.
Iskandar. 2009. *Psikologi Pendidikan.* Jakarta: gaung Persada Pers
Kristiyono, A. (2018). Urgensi dan Penerapan Higher Order Thingking Skills. *Jurnal Pendidikan Penabur*, *17*(31), 36–46.
Kurniawan, B. R., & Taqwa, M. R. A. (2018). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Fisika pada Materi Listrik Dinamis. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *3*(11), 1451–1457.
Lestari, P. E., Purwanto, A., & Sakti, I. (2019). Pengembangan Instrumen Tes Keterampilan Pemecahan Masalah Pada Konsep Usaha Dan Energi Di Sma. *Jurnal Kumparan Fisika*, *2*(3), 161–168. https://doi.org/10.33369/jkf.2.3.161-168
Liana, N., Suana, W., & Sesunan, F. (2018). Pengembangan Soal Tes Berpikir Tingkat Tinggi Materi Fluida Untuk Sma. *Journal of Komodo Science Education*, *01*(01), 66–78. http:ejournal.stkipsantupaulus.ac.id/index.php/jkse
Nurhayati, N., & Angraeni, L. (2017). Analisis Kemampuan Berpikir Tingkat Tinggi Mahasiswa (Higher Order Thinking) dalam Menyelesaikan Soal Konsep Optika melalui Model Problem Based Learning. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *3*(2), 119–126. https://doi.org/10.21009/1.03201
Rochman, S., & Hartoyo, Z. (2018). Analisis High Order Thinking Skills (HOTS) Taksonomi Menganalisis Permasalahan Fisika. *Science and Physics Education Journal (SPEJ)*, *1*(2), 78–88. https://doi.org/10.31539/spej.v1i2.268
Wartono, W., Hudha, M. N., & Batlolona, J. R. (2018). How are the physics critical thinking skills of the students taught by using inquiry-discovery through empirical and theorethical overview? *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(2), 691–697. https://doi.org/10.12973/ejmste/80632
Winarti. (2015). Profil Kemampuan Berpikir Analisis dan Evaluasi Mahasiswa dalam Mengerjakan Soal Konsep Kalor. *Jurnal Inovasi Dan Pembelajaran Fisika*, *2*(1), 19–24. https://ejournal.unsri.ac.id/index.php/jipf/article/view/2350/1214