Psychometric Analysis of High School Mathematics Ability Test

Aminah Ekawati
Department of Mathematics Education, Faculty of Mathematics and Natural Science
State University of Surabaya
Jl. Ketintang Selatan Gedung D1, Surabaya, Jawa Timur, 60231, Indonesia
&
Department of Mathematics Education, Faculty of Social and Humanities
University PGRI of Kalimantan
Jl. Sultan Adam Kompleks H. Iyus No.18, Banjarmasin, 70122, Indonesia
E-mail address: aminah.21017@mhs.unesa.ac.id
ORCID: https://orcid.org/0009-0008-5596-9579

Tatag Yuli Eko Siswono
Department of Mathematics Education, Faculty of Mathematics and Natural Science
State University of Surabaya
Jl. Ketintang Selatan Gedung D1, Surabaya, Jawa Timur, 60231, Indonesia
E-mail address: tatagsiswono@unesa.ac.id
ORCID: https://orcid.org/0000-0002-7108-8279

Agung Lukito
Department of Mathematics Education, Faculty of Mathematics and Natural Science
University State of Surabaya
Jl. Ketintang Selatan Gedung D1, Surabaya, Jawa Timur, 60231, Indonesia
E-mail address: agunglukito@unesa.ac.id
ORCID: https://orcid.org/0000-0003-1277-5834

ABSTRACT

Aim. Mathematics is essential for developing higher-order thinking skills. However, Indonesian students' performance in PISA 2022 and TIMSS 2015 shows a lack of proficiency in these skills. One contributing factor is school assessments that do not adequately foster higher-order thinking. This study evaluates the quality of a high school mathematics proficiency test designed to assess higher cognitive levels, focusing on validity, reliability, discrimination, difficulty, and variation in student responses.

Methods. This quantitative study used the MEASURE approach with five essay problems on exponents, logarithms, geometric sequences, arithmetic series, and trigonometry, involving three validators and 102 high school students from three district schools in Banjarmasin.

Results. The I-CVI values for questions 1, 2, 3, and 5 met the required criteria, requiring no revision. Question 4, with an I-CVI of 0.90, needed revision. All kappa values (κ) were ≥ 0.9, confirming reliability. Empirical data showed all questions had significance values below 0.05, and Cronbach's alpha was 0.72. Discrimination analysis categorized 60% of questions as excellent and 40% as good. Difficulty levels indicated that 80% were moderate and 20% difficult. The instrument generated diverse student responses, reflecting varying ability levels.

Conclusion. The high school mathematics proficiency test, designed for cognitive levels C4 and C5, met validity and reliability criteria, demonstrated good discrimination, and had varied difficulty levels. While emphasizing higher-order thinking, a more comprehensive assessment could integrate multiple-choice questions for C1-C3 and essays for C4-C5.

Keywords: difficulty level, discrimination level, MEASURE, reliability, validity

Introduction

Mathematics has become the basis and fundamental subject taught from elementary to the college level because it plays a role in developing critical thinking abilities (ArIsoy & Aybek, 2021; Sari & Juandi, 2023), logical thinking abilities (Sarnoko et al., 2024), and creative thinking abilities (Azaryahu et al., 2023; Grégoire, 2016). National Council of Teachers of Mathematics (NCTM) has determined that mathematics learning should focus on five essential abilities that enable students to solve problems effectively (NCTM, 2000). In line with this, through the Indonesian Curriculum 2021, the Indonesian Government has established the objective of mathematics learning to develop independence, critical reasoning abilities, and creativity (Kemendikbud, 2022).

Cognitive processes are divided into low-level and high-level thinking (Anderson & Krathwohl, 2001; Brame, 2019). Low-level thinking consists of three levels: remembering (C1), understanding (C2), and applying (C3), while high-level thinking includes analysing (C4), evaluating (C5), and creating (C6). There are three levels of C4, namely differentiating, organising, and attributing. At this level, students divide the problem into smaller sub-problems. They can determine how the parts are related or not, both to each other and the structure or purpose, so there is a logical relationship between them (Lewy et al., 2009; Zulkifli et al., 2021). There are two levels of C5, namely checking and critiquing. At this level, students can evaluate or assess a solution based on criteria and standards, such as understanding the material, critical thinking, and communication skills. Examples of this evaluation include exploring and determining the truth of the solution that has been given (Radmehr & Drake, 2018). Next, at level C6, there are three: generating, planning, and producing. At this level, students combine various components into one logically interrelated unit to form a new structure from before. Thus, mastering this cognitive level can help students distinguish between ideas and opinions, put forward arguments, and understand and interpret complex ideas (Nababan & Tanjung, 2020).

Students' mathematical abilities have gained significant attention in education (Etang & Regidor, 2022). This ability includes skills that enable a person to accomplish mathematical tasks and solve problems effectively (Karsenty, 2020; Wieczerkowski et al., 2000). Mathematical ability is an essential cognitive skill to develop, closely related to problem-solving, using number symbols, and logical reasoning (Abed & Hassan, 2021; Muhammad & Angraini, 2023). In addition, this ability includes understanding mathematical concepts and applying strategies to solve problems in academic and everyday contexts (Krutetskii, 1976). Hence, mathematical ability includes two things, namely cognitive aspects and practical skills needed to solve various problems.

Numerous international studies indicate that Indonesian students' mathematics performance requires enhancement. According to Organization for Economic Cooperation and Development (OECD) PISA 2022 report, Indonesia got a score of 366, significantly lower than the global average of 472; the percentage of students with low abilities increased by 13% compared to the PISA 2018 results (OECD, 2023). Similarly, the TIMSS 2015 results revealed that Indonesian students achieved an average score of 397, ranking them fifth from the bottom (Mullis et al., 2015). This data has demonstrated that many Indonesian students still have yet to master mathematics skills that involve high-level thinking abilities, such as applying, analysing, synthesising, and evaluating (Rahayu et al., 2019). This suggests that Indonesian students' mathematics abilities at a high level of thinking are still low.

One of the causes of low students' high-level thinking abilities is believed to be that the problems given by teachers at school cannot facilitate and encourage students to think at a higher level (Gradini et al., 2022; Ramdhani et al., 2024), whereas problems with high-level thinking abilities can provide stimulation to develop students' thinking skills (Kusuma et al., 2017). Well-designed problems will push students to engage in higher-level thinking, aiding them in developing the skills necessary to tackle complex problems. In this context, psychometric theory is essential in providing a set of high-quality, tested measurement instruments (Furr, 2021; Raykov & Marcoulides, 2011). Measurement instruments can be developed using psychometrics to ensure high reliability and validity and accurately measure students' high-level thinking abilities (Yang et al., 2022). Therefore, this study aims to determine the quality of high school mathematics ability test instruments at a high cognitive level.

Methods

This research used a quantitative research design with the MEASURE approach. The MEASURE approach, designed by Michael T. Kalkbrenner et al., consists of seven steps: (a) Making the Aim and Rationale Obvious; (b) Establishing an Empirical Framework; (c) Articulating the Theoretical Blueprint; (d) Synthesizing Content and Expanding Development; (e) Using Expert Reviewers; (f) Recruiting Participants; and (g) Evaluate Validity and Reliability (Kalkbrenner, 2021). This approach helps evaluate and develop instruments based on leading psychometric principles.

Making the Aim and Rationale Obvious

This study aims to determine the quality of high school mathematics ability test instruments at a high cognitive level. Qualities include validity, reliability, discrimination level, level of difficulty, and variation of answers.

Establishing an Empirical Framework

At this stage, the researcher identified the theory of mathematics ability that would be used. The researcher determined the cognitive dimensions of the revised Bloom's taxonomy to determine the level of ability on the mathematics test. There are six cognitive levels in the revised Bloom's taxonomy, and the researcher determined the ability test at levels C4 (analyse) and C5 (evaluate). These two levels were chosen because they could measure high-level thinking abilities. In this research, level C4 was related to the description that students must have been able to analyse the relationship between concepts and break down the given problem into smaller components in order to solve the given problem. By contrast, level C5 was linked to the description that students were asked to evaluate the problems given, involving assessing the solutions, strategies, or methods used to solve the problem. Both levels, C4 and C5, were considered more challenging and relevant to preparing students to face real-world problems.

Articulating Theoretical Blueprint

According to Kalkbrenner (2021), the blueprint consists of two main components: the content and domain areas. The content area refers to the specific aspects of the subject being measured. In this case, the subjects to be measured are high school students who have implemented the independent curriculum and studied the topics of exponents and logarithms, sequences and series, and trigonometry. The five selected concepts were selected based on several considerations:

- studied in grade X odd semester;
- to measure higher-order thinking abilities because mathematical problems can be designed to involve understanding and applying complex mathematical concepts;
- the basics for more advanced mathematical abilities and applications at higher education levels. Measuring students' abilities in these areas provides insight into their readiness for more complex material.

The domain area, on the other hand, describes the construction and the type of mathematical problems used. There are two types of mathematical problems, namely applied mathematical and pure mathematical problems (Obeng-Denteh & Amoah-Mensah, 2011). Both types of mathematical problems are used in this study.

Table 1
The Blueprint of the High School Mathematics Ability Test Instrument

Indicator	Domain area	Cognitive level	Number of items
Analysing logarithmic problems	pure mathematical	C4	1
Evaluating exponent problems	pure mathematical	C5	1
Evaluating geometric sequence problems	applied mathematical	C5	1
Analysing trigonometric problems	applied mathematical	C4	1
Evaluating arithmetic series problems	applied mathematical	C5	1

Source. Own research.

Synthesising Content and Expanding Development

Two main activities are carried out at this stage: synthesising content and determining the scale. Synthesising content means integrating exponent, logarithmic, arithmetic series, geometric series, and trigonometry materials to create problems that will be used as instruments. This process consists of selecting, arranging, and simplifying problems to make them easier for students to understand, using good and correct Indonesian, as well as time for students to work on problems. The time available was only 75 minutes, so we decided that students would only work on five essay problems, because they are considered effective in assessing students' reasoning and thinking skills (Reiner et al., 2002).

We have a scoring guideline with a maximum score for each problem. If students answer according to the guidelines' steps, they will receive a score of one. However, there is a possibility that students' answers need to comply with the guidelines made entirely. If the steps follow mathematical rules, the answer will still be considered correct, although the assessment will remain capped according to the maximum score for each problem. Thus, the scoring guidelines are flexible, while we strive to ensure that the alternative answers are aligned with typical students' answers because these alternatives are based on the material taught in high school.

Using Expert Reviewers

Involving different validators — individuals not involved in developing the initial items — is crucial to obtaining new and objective feedback (Davis, 1992; Kalkbrenner, 2021). We use three validators with over 20 years of teaching experience, doctoral education, and research focus on student cognition. We provided assessment sheets and instruments to each Validator and ensured no discussion between validators when assessing the instruments. The validators returned the instruments to us two weeks after they were received. After all the instruments were returned, we began analysing the data obtained.

Recruiting Participants

Participants are students in in three high schools in Banjarmasin City, chosen based on the following criteria: (a) the schools hold A accreditation; (b) they are located in different sub-districts; (c) the distribution of students' academic abilities in one class is even; (d) the schools have implemented the national curriculum since 2021; and (e) both the schools and students were willing to participate in the trial. Based on these characteristics, the following three schools were selected.

Table 2
Schools for Testing the Instrument

No.	School Name	District
1	SMAn 1	Central Banjarmasin
2	SMAn 7	Eastern Banjarmasin
3	SMAn 4	Western Banjarmasin

Source. Own research.

One class from each of these schools was selected, comprising students who had already studied the material being tested. Students were given 75 minutes to complete the test. Below is the distribution of students in each school.

Table 3
The Distribution of Students at Each Trial School

School Name	Male	Female	Sum
SMAn 1	17	19	36
SMAn 7	11	18	29
SMAn 4	16	21	37
Sum	102		

Source. Own research.

Evaluate Validity and Reliability

After the instrument was created, it was first validated by experts before being empirically tested. The validators give scores on the validation sheet, which consists of the main aspects: problem construction, language clarity, and suitability of the material to the 2021 National Curriculum. The validators assess each aspect on a scale of one to four, where one means not appropriate, two means less appropriate, three means appropriate, and four means very appropriate (Davis, 1992). Based on this assessment, statements scoring three or four were assigned a value of one, while those scoring one or two were assigned a value of zero (Almanasreh et al., 2019; Davis, 1992; Lynn, 1986; Polit et al., 2007; Yusoff, 2019). CVI can be calculated for each item on the instrument (item-level CVI or I-CVI) or for the overall instrument (instrument-level CVI) (Almanasreh et al., 2019; Yusoff, 2019). The formula for I-CVI is: . With three validators, an I-CVI 1 indicates acceptance, while less than 1 requires revision (Almanasreh et al., 2019). Reliability, based on validators' assessments, was calculated using the kappa coefficient, , with (Almanasreh et al., 2019; Zamanzadeh et al., 2015). In this formula, N represents the total number of validators, and A is the number of validators rated the item as 3 or 4. The expected κ value is ≥ 0.6.

We also calculated empirical data related to validity, reliability, discrimination, and difficulty level. Validity and reliability are empirically calculated using SPSS 25. Validity was measured using the Pearson correlation test (Wijaya & Kloping, 2021) at a 5% significance level; if the significance value was < 5%, the item was deemed valid. Value of Cronbach's alpha ≥ 0.7, the instrument is reliable (Lacave et al., 2018). Discriminant analysis was used to measure whether the problem can distinguish between high-ability and low-ability students, with the formula (Boopathiraj & Chellamani, 2013). A value between 0.3 and 0.39 is considered good, while a value of ≥ 0.4 is very good (Diki & Yuliastuti, 2018; Finch & French, 2019). Finally, the difficulty value was calculated by the formula of the average score of all students divided by the maximum score (Finch & French, 2019). Items scoring above 0.90 are considered too easy and may not be suitable for testing. Conversely, items scoring below 0.20 are considered too difficult and should be reviewed for possible issues with language or re-instructions on the content. The expected difficulty values are 0.30 and 0.70 for moderate difficulty, between 0.2 and 0.30 for difficult, and between 0.7 and 0.90 for easy problems (Finch & French, 2019).

Results

Table 4 shows the results of the I-CVI and Kaffa coefficient based on the aspects assessed by the Validator. As shown in Table 4, the I-CVI values for problems 1, 2, 3, and 5 meet the required criteria, indicating that these problems can be accepted without revision. However, problem 4 obtained an I-CVI of 0.90, indicating the need for revision. Revisions were made primarily in the construction and language aspects. Below are the details of the revisions made.

Table 5

Question 4 Before and After Receiving Validator Input

Before revision	After revision
Sisno was asked to measure the height of the flagpole using a clinometer. The elevation angle indicated by the clinometer was 60°. The height of Sisno's eyes from the ground was 1.6 meters. Sisno then moved 10 meters from the starting position, and the elevation angle on the clinometer was 45 °. Determine the height of the flagpole	Sisno was in a place close to the flagpole. Sisno was asked to measure the height of the flagpole using a clinometer. The elevation angle indicated by the clinometer was 60°. The height of Sisno's eyes from the ground was 1.6 meters. Sisno then moved 10 meters from the starting position, and the elevation angle was 45° on the clinometer. Determine the height of the flagpole.

Source. Own research.

Based on the I-CVI value and the revisions made, the test instrument can be considered valid. This means the problems are in agreement with the curriculum national Indonesia 2021, the scoring guidelines for the alternative answers given are appropriate, the number of problems fits the allotted time, and the language used does not give rise to multiple interpretations and follows Indonesian spelling language.

Table 4 shows that the instrument is reliable because it met the kappa coefficient value (κ) criteria. Based on these two results, we conducted a trial to obtain empirical data regarding validity, reliability, discrimination level, and difficulty index. The following are the validity results.

Tabel 4

The Results of I-CVI and Kappa Coefficient

Aspects assessed		Item									
		1		2		3		4		5	
		I-CVI	*Kappa*	*I-CVI*	*Kappa*	*I-CVI*	*Kappa*	*I-CVI*	*Kappa*	*I-CVI*	*Kappa*
construction	The information in the problems is clear and easy to understand, so it is sufficient to solve them.	1,00	1,00	1,00	1,00	1,00	1,00	0,67	0,78	1,00	1,00
	The problems are arranged using easy-to-understand questions or commands	1,0	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	The problems are arranged using correct mathematical sentences	1,00	1,00	1,00	1,00	1,00	1,00	0,67	0,78	1,00	1,00
material suitability	The problems can be presented and used to express high school mathematics abilities	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	The problems presented are by the question indicators	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	The expected boundaries of problems and answers are clear	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	The material is based on the learning outcomes in the independent curriculum	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
language	The problems presented use simple language and do not give rise to multiple interpretations	1,00	1,00	1,00	1,00	1,00	1,00	0,67	0,78	1,00	1,00
	The problems use terms that are easy to understand	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	The problems are formulated using words that in the Indonesian Language	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Average	1,00	1,00	1,00	1,00	1,00	1,00	0,90	0,93	1,00	1,00	

Source. Own research.

Table 6

The Results of Empirical Validation

Problem	Value Pearson Correlation	Significance	Category
1	0.80	0.00	Valid
2	0.69	0.00	Valid
3	0.66	0.00	Valid
4	0.60	0.00	Valid
5	0.76	0.00	Valid

Source. Data were analysed using SPSS 25.

From Table 6, all significance values <0.05, so it can be concluded that all problems on the mathematics ability test instrument are valid. After the validity test, it was continued with the Reliability test. The instrument obtained is reliable because Cronbach's alpha value = 0.72, which meets the requirements ≥ 0.7. Next, we calculate the level of discrimination and the level of difficulty.

Furthermore, we calculated the discrimination level and difficulty level. The results of the discriminatory level are presented in Table 7.

Table 7

Discrimination Level

Problem	D_i	Category
1	0.82	very good
2	0.50	very good
3	0.31	good
4	0.31	good
5	0.65	very good

Source. Own research.

Table 7. shows three (60%) problems in a very good and two (40%) in a good category. This shows that the problems can distinguish students with high abilities and students with low abilities, meaning that the problems can be answered by students with high abilities but not by students with low abilities.

The level of difficulty of the problems is shown in Table 8. The test instrument has varying levels of difficulty, namely moderate and difficult. Although no problems are included in the easy category, these problems can still be used in general because 80% of the problems are in the moderate level, while only 20% are included in the difficult level.

Table 8

Difficulty Level

Problem	Difficulty level	Category
1	0.61	moderate
2	0.35	moderate
3	0.31	moderate
4	0.22	difficult
5	0.60	moderate

Source. Own research.

We present one problem each from C4 and C5 and some of the students' work. Here is the C4 problem.

Figure 1

C4 Problem

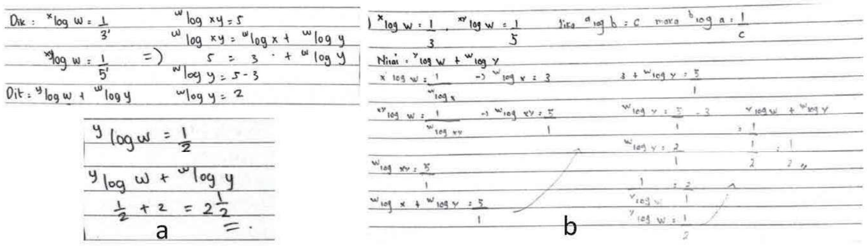
If $\log_x w = \frac{1}{3}$ and $\log_{xy} w = \frac{1}{5}$. Find the value of $\log_y w + \log_w y$.

Source. Own research.

In Figure 1, students are asked to determine the value of the logarithm by first breaking down the known logarithm and analysing the relationship between elements in the logarithmic statement. Next, students must be able to compile steps to solve the given problem using the properties of logarithms. The following is a display of various student answers. The following are some of the answers given by students.

Figure 2

Student Answers to C4 Problem



Dik : ${}^{x}\log w = \frac{1}{3'}$

${}^{xy}\log w = \frac{1}{5'}$ =)

${}^{w}\log xy = 5$

${}^{w}\log xy = {}^{w}\log x + {}^{w}\log y$

$5 = 3 + {}^{w}\log y$

${}^{w}\log y = 5 - 3$

${}^{w}\log y = 2$

Dit : ${}^{y}\log w + {}^{w}\log y$

${}^{y}\log w = \frac{1}{2}$

${}^{y}\log w + {}^{w}\log y$

$\frac{1}{2} + 2 = 2\frac{1}{2}$

=.

1) ${}^{x}\log w = \frac{1}{3}$, ${}^{xy}\log w = \frac{1}{5}$ Jika ${}^{a}\log b = c$ maka ${}^{b}\log a = \frac{1}{c}$

Nilai : ${}^{y}\log w + {}^{w}\log y$

${}^{x}\log w = \frac{1}{{}^{w}\log x}$ -> ${}^{w}\log x = 3$

${}^{xy}\log w = \frac{1}{{}^{w}\log xy}$ -> ${}^{w}\log xy = \frac{5}{1}$

${}^{w}\log xy = \frac{5}{1}$

${}^{w}\log x + {}^{w}\log y = \frac{5}{1}$

$3 + {}^{w}\log y = \frac{5}{1}$

${}^{w}\log y = \frac{5}{1} - 3$

${}^{w}\log y = \frac{2}{1}$

$\frac{1}{{}^{y}\log w} = \frac{2}{1}$

${}^{y}\log w = \frac{1}{2}$

${}^{y}\log w + {}^{w}\log y$

$= \frac{1}{\frac{1}{2}} : \frac{1}{2}$,,

Source. Own research.

This is an example of a C5 problem.

Figure 3

C5 Problem

Suppose x is a positive real number, $A^{2x} = 2$. Is it true that

$$\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}} = \frac{31}{17}?$$

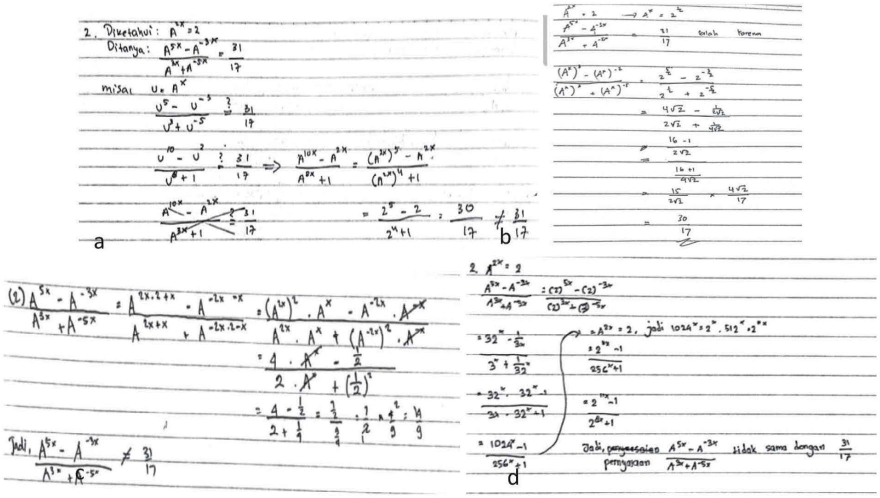
Explain the steps and mathematical concepts used!

Source. Own research.

In Figure 2, students are asked to evaluate the truth of a mathematical statement involving exponents and the properties of real numbers. Students must first understand the concept of exponents, including their properties of multiplication and division, changing their form, and manipulating problems algebraically to simplify a given expression. After that, students can assess the truth of the statement. The following are some of the answers given by students.

Figure 4

Student Answers to C5 Problem



2. Diketahui: $A^{2x} = 2$

Ditanya: $\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}} = \frac{31}{17}$

misal $u = A^x$

$\frac{u^5 - u^{-3}}{u^3 + u^{-5}} \overset{?}{=} \frac{31}{17}$

$\frac{u^{10} - u^2}{u^8 + 1} \overset{?}{=} \frac{31}{17} \Rightarrow \frac{A^{10x} - A^{2x}}{A^{8x} + 1} = \frac{(A^{2x})^5 - A^{2x}}{(A^{2x})^4 + 1}$

$\frac{A^{10x} - A^{2x}}{A^{3x} + 1} \overset{?}{=} \frac{31}{17}$ $= \frac{2^5 - 2}{2^4 + 1} = \frac{30}{17} \neq \frac{31}{17}$

$A^{2x} = 2 \longrightarrow A^x = 2^{\frac{1}{2}}$

$\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}} = \frac{31}{17}$ salah karena

$\frac{(A^x)^5 - (A^x)^{-3}}{(A^x)^3 + (A^x)^{-5}} = \frac{2^{\frac{5}{2}} - 2^{-\frac{3}{2}}}{2^{\frac{3}{2}} + 2^{-\frac{5}{2}}}$

$= \frac{4\sqrt{2} - \frac{1}{2\sqrt{2}}}{2\sqrt{2} + \frac{1}{4\sqrt{2}}}$

$= \frac{\frac{16-1}{2\sqrt{2}}}{\frac{16+1}{4\sqrt{2}}}$

$= \frac{15}{2\sqrt{2}} \times \frac{4\sqrt{2}}{17}$

$= \frac{30}{17}$

(2) $\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}} = \frac{A^{2x \cdot 2 + x} - A^{-2x - x}}{A^{2x + x} + A^{-2x \cdot 2 - x}} = \frac{(A^{2x})^2 \cdot A^x - A^{-2x} \cdot \cancel{A^{-x}}}{A^{2x} \cdot A^x + (A^{-2x})^2 \cdot \cancel{A^{-x}}}$

$= \frac{4 \cdot \cancel{A^x} - \frac{1}{2}}{2 \cdot \cancel{A^x} + (\frac{1}{2})^2}$

$= \frac{4 - \frac{1}{2}}{2 + \frac{1}{4}} = \frac{\frac{7}{2}}{\frac{9}{4}} = \frac{7}{2} \times \frac{4}{9} = \frac{14}{9}$

Jadi, $\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}} \neq \frac{31}{17}$

2. $A^{2x} = 2$

$\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-3x}} = \frac{(2)^{5x} - (2)^{-3x}}{(2)^{3x} + (2)^{-5x}}$

$= \frac{32^x - \frac{1}{3x}}{3^x + \frac{1}{32^x}}$

$= \frac{32^x \cdot 32^x - 1}{3x \cdot 32^x + 1}$

$= \frac{1024^x - 1}{256^x + 1}$

$= A^{2x} = 2$, jadi $1024^x = 2^x \cdot 512^x \cdot 2^{9x}$

$= \frac{2^{9x} - 1}{256^x + 1}$

$= \frac{2^{11x} - 1}{2^{8x} + 1}$

Jadi, ~~penyelesaian~~ pernyataan $\frac{A^{5x} - A^{-3x}}{A^{3x} + A^{-5x}}$ tidak sama dengan $\frac{31}{17}$

Source. Own research.

Discussion

The I-CVI and kappa coefficient results were obtained quantitatively based on the validation sheet filled out by the validator. We used I-CVI to measure content validity. I-CVI aims to ensure the suitability of the test items with the learning objectives, including cognitive level, indicators, materials, language, and test construction (Sürücü & Maslakçı, 2020). Meanwhile, the kappa coefficient is used to assess reliability. This reliability is based on the level of agreement between experts on each test item, considering the possibility of agreement by chance (Furlan et al., 2021; Polit et al., 2007; Zamanzadeh et al., 2015).

The Validator's opinion is important in instrument development (Tanujaya, 2016). The results of the study showed that the developed high school mathematics ability test instrument had met the I-CVI and kappa coefficient. Given the potential for subjectivity in validator assessments (Zamanzadeh et al., 2015), researchers also used construct validity. Construct validity was analysed using empirical data to ensure the effectiveness of the measuring instrument (Sürücü & Maslakçı, 2020; Tobón & Luna-Nemecio, 2021).

Empirical results show that the instruments produced are proven to be valid and reliable, where valid means that the measuring instrument measures what it should measure, and reliability means that the instrument produces consistent results when used at different times (Sürücü & Maslakçı, 2020; Tobón & Luna-Nemecio, 2021). From the results of the discriminant power analysis, it was obtained that most of the instruments developed were included in the very good category. This means that the instrument effectively distinguishes students with high and low-level thinking abilities. Good problems have a balanced difficulty level between difficult, medium, and easy problems and are not too difficult or too easy to answer (Dunn et al., 2003). In this study, an easy category was not found. This is likely due to the cognitive level determined by the researcher being at levels C4 and C5, while easy problems are generally at levels C1 and C2 (Giani et al., 2015).

Figures 2 and 4 show the diversity of ways students solve the given problems. In Figure 2, students solve the given problem by changing the logarithmic form using the properties of logarithms. In Figure 2a, students have started with the right steps using the logarithmic properties . Next, students use the properties of logarithmic multiplication, , and the steps students use to solve are correct. Meanwhile, in Figure 2b, students also start with the same steps as students in Figure 2a, but there is an error in manipulating the last part of the algebra. Students seem to have difficulty in simplifying algebra.

Figure 4a, the student assumes the variable u as and then replaces each with its equivalent. The results of the equation are then multiplied with . The student 4a successfully answers the given problem by applying the properties of exponents. On the other hand, in Figure 4b, students start by simplifying the known information using the properties of exponents and then apply the properties of rational root form to solve the problem. Although the methods of solving students 1a and 1b are different, they produce the same answer. In Figure 4c, students understand the purpose of the problem and can apply the rules of exponents but need help solving the problem due to errors in simplifying the form.

Meanwhile, in Figure 4d, the student still needs to understand the concept of exponents, so he needs help to solve the problem correctly. Based on the answers given, the student answers the task only by justifying or blaming the problem given without writing the reasons or concepts used in the answer. This condition is likely caused by the habit of students who are more focused on solving problems using certain strategies rather than explaining the reasons for choosing the strategy.

Two examples of student answers show the diversity in student responses to the problems given. The variation in students' answers reflects their thinking abilities (Tanujaya, 2016) and shows the quality of the instrument in evaluating and facilitating higher-order thinking abilities. This instrument challenges students to think at a higher level, apply mathematical concepts, and encourage interpretation and analysis of information, which are the core of higher-order thinking. Higher-order thinking ability is essential to teach (Heong et al., 2011), because students can solve problems in higher-order thinking. Students will solve complex problems by connecting previous knowledge with newly obtained information to achieve goals (Yee et al., 2015). Reasoning ability can be trained and developed as part of higher-order thinking (Tanujaya, 2016).

Conclusions and Limitations

The high school students' mathematics ability test instrument, developed with cognitive levels C4 and C5, meets the validity and reliability criteria, has a good minimal discrimination level, and has varying difficulty levels (moderate and complex). The difficulty level with the easy category was not found because this instrument focuses entirely on measuring cognitive levels C4 and C5. This instrument is a handy evaluation tool that can be used as a reference in developing future tests. In addition, this instrument can be used by teachers and researchers to evaluate high school students' mathematics ability at a higher cognitive level, thus supporting the enhancement of the quality of mathematics education.

The variation of students' answers in solving problems shows students' thinking abilities and the instrument's effectiveness in facilitating and evaluating high-level thinking abilities. This instrument encourages students to apply mathematical concepts, think critically and logically, and connect existing knowledge to solve problems. High-level thinking abilities can be trained and developed, one of which is through this instrument.

However, this instrument has areas for improvement because researchers focus on high cognitive levels, namely C4 and C5. The abilities of high school students are not only limited to high cognitive levels but also include lower cognitive abilities. Therefore, test problems should also accommodate cognitive levels C1 to C3. The problems used are descriptive so that the material tested cannot be too much. To overcome this limitation, a mathematics ability test using a combination of multiple-choice and essay problems is more suitable. Multiple-choice problems can measure cognitive levels C1 to C3, while essay problems can measure cognitive levels C4 and C5. Thus, the scope of the material becomes more expansive, and the variation of students' cognitive levels can be more comprehensive.

Acknowledgements

The authors (ID Number 202101120874) express their appreciation to the Indonesian Education Scholarship (Beasiswa Pendidikan Indonesia [BPI]), Center for Higher Education Funding and Assesment (Pusat Pelayanan Pembiayaan dan Assesmen Pendidikan Tinggi [PPAPT]) and the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan [LPDP]), under the Indonesian Ministry of Higer Education, Science, and Technology of Republic (Kementerian Pendidikan Tinggi, Sains, dan Teknologi Indonesia), for their support through scholarships and funding for this article.

REFERENCES

Abed, R. K., & Hassan, A. K. (2021). Multiple mathematical representations according to the (lesh) model of high school students and its relationship to their mathematical ability. *Journal of Contemporary Issues in Business and Government*, *27*(3), 2200–2211. https://cibgp.com/au/index.php/1323–6903/article/view/1830/1800.

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. a revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman, Inc.

ArIsoy, B., & Aybek, B. (2021). The effects of subject-based critical thinking education in mathematics on students' critical thinking skills and virtues. *Eurasian Journal of Educational Research*, *21*(92), 99–119. https://doi.org/10.14689/ejer.2021.92.6

Azaryahu, L., Broza, O., Cohen, S., Hershkovitz, S., & Adi-Japha, E. (2023). Development of creative thinking patterns via math and music. *Thinking Skills and Creativity*, *47*, Article 101196. https://doi.org/10.1016/j.tsc.2022.101196

Boopathiraj, C., & Chellamani, D. K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, *2*(2), 189–193.

Brame, C. J. (2019). Spotlight 1. Writing Learning Objectives Using Bloom's Taxonomy. In C. J. Brame (Ed.), *Science Teaching Essentials* (pp. 29–34). Elsevier. https://doi.org/10.1016/B978–0-12–814702–3.00025–1

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194–197. https://doi.org/10.1016/S0897–1897(05)80008–4

Diki, D., & Yuliastuti, E. (2018). Discrepancy of difficulty level based on item analysis and test developers' judgment: Department of Biology at Universitas Terbuka, Indonesia. In K. Persichitte, A., Suparman, M., Spector (Eds.), *Educational Technology to Improve Quality and Access on a Global Scale* (pp. 215–225). https://doi.org/10.1007/978–3-319–66227–5_17

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2003). *The Student Assessment Handbook*. Routledge. https://doi.org/10.4324/9780203416518

Etang, M. A. G., & Regidor, R. M. (2022). Students' mathematical beliefs and attitudes as predictors to students' mathematical ability. *International Journal of Education and Social Science Research*, *5*(3), 23–60. https://doi.org/10.37500/IJESSR.2022.5303

Finch, W. Holmes., & French, B. F. (2019). *Educational and Psychological Measurement*. Routledge.

Furlan, R. M. M. M., Santana, G. A., Motta, A. R., & Casas, E. B. de Las. (2021). An instrument for tongue performance assessment in activities associated with digital games: Content and construct validity. *Revista CEFAC*, *23*(5). https://doi.org/10.1590/1982–0216/20212359621

Furr, R. M. (2021). *Psychometrics: An introduction* (4th ed.). SAGE publications. (Original work published 2007)

Giani, Zulkardi, & Hiltrimartin, C. (2015). Analisis tingkat kognitif soal-soal buku teks matematika kelas VII berdasarkan taksonomi Bloom [Analysis of cognitive level of questions in grade VII mathematics textbooks based on Bloom's taxonomy]. *Jurnal Pendidikan Matematika*, *9*(2), 78–98.

Gradini, E., Khairunnisak, C., & Noviani, J. (2022). Development of higher-order thinking skill (hots) test on mathematics in secondary school. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(1), 319–330. https://doi.org/10.24127/ajpm.v11i1.4649

Grégoire, J. (2016). Understanding creativity in mathematics for improving mathematical education. *Journal of Cognitive Education and Psychology*, *15*(1), 24–36. https://doi.org/10.1891/1945–8959.15.1.24

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. B., & Mohamad, M. M. B. (2011). The level of Marzano Higher Order Thinking Skills among technical education students. *International Journal of Social Science and Humanity, 1*(2), 121–125.

Kalkbrenner, M. T. (2021). A practical guide to instrument development and score validation in the social sciences: The measure approach. *Practical Assesment, Research & Evaluation*, *26*(1), Article 1. https://doi.org/10.7275/svg4-e671

Karsenty, R. (2020). Mathematical ability. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 494–497, 2nd ed.).

Kemendikbud. (2022). *Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada Kurikulum Merdeka* [Learning Outcomes in Early Childhood Education, Elementary Education Level, and Secondary Education Level in the Independent Curriculum]. https://kurikulum.kemdikbud.go.id/wp-content/unduhan/CP_2022.pdf

Krutetskii, V. A. (1976). *The Psychology of Mathematical Abilities in School Children.* University of Chicago Press.

Kusuma, M. D., Rosidin, U., Abdurrahman, & Suyatna, A. (2017). The development of higher order thinking skill (hots) instrument assessment in physics study. *IOSR Journal of Research & Method in Education*, *7*(1), 26–32. https://doi.org/10.9790/7388–070103XXXX

Lacave, C., Molina, A. I., & Redondo, M. A. (2018). A preliminary instrument for measuring students' subjective perceptions of difficulties in learning recursion. *IEEE Transactions on Education*, *61*(2), 119–126. https://doi.org/10.1109/TE.2017.2758346

Lewy, L., Zulkardi, & Aisyah, N. (2009). Pengembangan soal untuk mengukur kemampuan berpikir tingkat tinggi pokok bahasan barisan dan deret bilangan di kelas IX akselerasi SMP Xaverius Maria Palembang [Development of questions to measure high-level thinking skills on the topic of number sequences and series in class IX acceleration at SMP Xaverius Maria Palembang]. *Jurnal Pendidikan Matematika*, *3*(2), 15–28.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385.

Muhammad, I., & Angraini, L. M. (2023). Research on students' mathematical ability in learning mathematics in the last decade: A bibliometric review. *JOHME: Journal of Holistic Mathematics Education*, *7*(1), 108–122. https://doi.org/10.19166/johme.v7i1.6867

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2015). *TIMSS 2015 International Result in Mathematics*. International Study Center, Lynch School of Education, Boston College.

Nababan, S. A., & Tanjung, H. S. (2020). Development of learning instrument based on realistic mathematics approach to improve mathematical disposition ability. *Advances in Mathematics: Scientific Journal*, *9*(12), 10325–10333. https://doi.org/10.37418/amsj.9.12.24

NCTM. (2000). *Principles and standards for school mathematics*. NCTM: Reston VA.

Obeng-Denteh, W., & Amoah-Mensah, J. (2011). Pure mathematicians' and applied mathematicians' saga: but one family! a mathematical panacea. *Continental Journal of Education Research*, *4*(2), 1–10. https://www.researchgate.net/publication/326232786

OECD. (2023). *PISA 2022 Results (Volume I)*. OECD. https://doi.org/10.1787/53f23881-en

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Radmehr, F., & Drake, M. (2018). An assessment-based model for exploring the solving of mathematical problems: Utilising revised Bloom's taxonomy and facets of metacognition. *Studies in Educational Evaluation*, *59*, 41–51. https://doi.org/10.1016/j.stueduc.2018.02.004

Rahayu, W., Sinaga, O., Oktaviani, M., & Zakiah, R. (2019). Analysis of mathematical ability of high school students based on item identification of national examination set. In Jufrizal, Y. Rozimela, Atmazaki, A. Fauzan, R. Syahrul, Hamzah, R. Refnaldi, & Yerizon (Eds.), *Proceedings of the 1st International Conference on Innovation in Education (ICoIE 2018)* (pp. 412–416). Atlantis Press. http://dx.doi.org/10.2991/icoie-18.2019.88

Ramdhani, S. S., Susanti, R., & Meilinda. (2024). Cognitive level of Program for International Student Assessment (PISA) questions based on the revised Bloom's taxonomy. *European Journal of Education and Pedagogy*, *5*(2), 104–112.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to Psychometric Theory.* Routledge.

Reiner, C. M., Bothell, T. W., & Sudweeks, R. R. (2002). *Preparing effective questions a Self-directed workbook for educators*. New Forums Pres.

Sari, R. N., & Juandi, D. (2023). Improving Student's Critical Thinking Skills in Mathematics Education: A Systematic Literature Review. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *7*(1), 845–861. https://doi.org/10.31004/cendekia.v7i1.2091

Sarnoko, S., Asrowi, A., Gunarhadi, G., & Usodo, B. (2024). An analysis of the application of problembased learning (PBL) model in mathematics for elementary school students. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(1), 188–202. https://doi.org/10.22437/jiituj.v8i1.32057

Sürücü, L., & Maslakçı, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726. https://doi.org/10.15295/bmij.v8i3.1540

Tanujaya, B. (2016). Development of an instrument to measure higher order thinking skills in senior high school mathematics instruction. *Journal of Education and Practice*, *7*(21), 144–148.

Tobón, S., & Luna-Nemecio, J. (2021). Complex thinking and sustainable social development: Validity and reliability of the complex-21 scale. *Sustainability (Switzerland)*, *13*(12), Article 6591. https://doi.org/10.3390/su13126591

Wieczerkowski, W., Cropley, A. J., & Prado, T. M. (2000). Nurturing talents/gifts in mathematics. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 413–425). http://dx.doi.org/10.1016/B978–008043796–5/50029–2

Wijaya, M. C., & Kloping, Y. P. (2021). Validity and reliability testing of the Indonesian version of the eHealth Literacy Scale during the COVID-19 pandemic. *Health Informatics Journal*, *27*(1). https://doi.org/10.1177/1460458220975466

Yang, E., Halpin, P., & Handy, D. (2022). *Using psychometric analysis to improve soft-skill assessments*. MDRC.

Yee, M. H., Yunos, J. Md., Othman, W., Hassan, R., Tee, T. K., & Mohamad, M. M. (2015). Disparity of learning styles and higher order thinking skills among technical students. *Procedia – Social and Behavioral Sciences*, *204*, 143–152. https://doi.org/10.1016/j.sbspro.2015.08.127

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, *11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zulkifli, I. Z., Mohamad Razi, N. F., & Mohammad, N. H. (2021). Cognitive levels towards performance of mathematics score in secondary school. *Mathematical Sciences and Informatics Journal*, *2*(1), 70–78. https://doi.org/10.24191/mij.v2i1.13042