Jurnal Gantang *10*(2) (2025): 291-301

JURNAL GANTANG

e-ISSN: 2548-5547
p-ISSN: 2503-0671

http://ojs.umrah.ac.id/index.php/gantang/index

Development of instruments of higher order thinking skills (HOTS) for phase D students

Kurnila Mildariah, Maimunah*, Rini Dian Anggraini

Universitas Riau, Pekanbaru, Riau, 28761, Indonesia
*Corresponding Author: maimunah@lecturer.unri.ac.id

Submission: June 29th 2025; Accepted: November 23rd 2025; Published: November 30th 2025
DOI: https://doi.org/10.31629/jg.v10i2.7387

Abstract

This research aims to develop test instruments for Higher Order Thinking Skills (HOTS) that are validated, reliable, and possess excellent discrimination power, as well as appropriate levels of difficulty. These tools are intended to boost the advanced thinking abilities of students at phase D. According to the Organization for Economic Cooperation and Development (OECD) report on December 5, 2023, which analyzed the 2022 Program for International Student Assessment (PISA) scores in mathematics, including HOTS questions, there has been a notable worldwide decline. Specifically, Indonesia ranked 68th, with a math score of 379. The focus of this study is on developing HOTS test tools that address topics such as relations, functions, linear equations, and systems of two-variable linear equations, drawing from the AKM (Minimum Competency Assessment) and HOTS mathematics question books. The HOTS questions are designed to encompass cognitive levels C4 (analyzing), C5 (evaluating), and C6 (creating), following the revised Bloom's Taxonomy by Krathwohl and Anderson. As highlighted by the Ministry of Education, Culture, Research, and Technology (2022), the development of HOTS questions must include elements such as (1) incorporating stimuli and (2) presenting new contexts within the material or question formulation. The questions are constructed in essay format, adhering to standards for validity and reliability, as well as optimal difficulty and discrimination levels. This study employs a development model comprising two phases: the preliminary phase, which involves analysis and design, followed by the formative evaluation phase, encompassing self-evaluation, expert review, one-on-one sessions, small group interactions, and field tests. The outcome of this study produced 17 valid questions with a reliability coefficient of 0.848; 14 questions exhibited strong discrimination power, and 3 demonstrated adequate discrimination power, with all 17 HOTS questions classified in the medium difficulty range.

Keywords: higher order thinking skills, validity, reliability, instruments

How to cite: Mildariah, K., Maimunah, M., Anggraini, R.D. (2025). Development of instruments of higher order thinking skills (HOTS) for phase D students. *Jurnal Gantang, 10*(2), 291 – 301. https://doi.org/10.31629/jg.v10i2.7387

I. Introduction

Advanced cognitive abilities are crucial for every person to possess. According to Tiwery (2019) and Wandari, Kamid & Maison (2018), in terms of farming character, particularly in shaping the character of 21st-century students, fostering Higher Order Thinking Skills (HOTS) is essential for their independence and ability to think

efficiently. The support needed to train students' high-level thinking skills is one of the key areas that educators focus on, because teachers are the spearhead for carrying out this training (Izzati & Febrian, 2021). According to Ariani, Candiyasa & Sriyasa (2023), teachers should apply assessment based on *Higher Order Thinking Skills* (HOTS) to face the competitive developments of this century. The implementation of HOTS-based assessments aims to enable students to compete globally and improve the quality of students from both social and educational aspects (Anita, Jumroh & Reta, 2023). Therefore, the government is trying to ensure that students can master high-level thinking skills by renewing the curriculum and implementing HOTS-based assessments (Hadi & Novaliyosi, 2019).

The efforts made by the government have not yet produced maximum results. Based on reports from *Organization for Economic Cooperation and Development* (OECD) on December 5th 2023 on students' mathematics scores on assessment *Programme for International Student Assessment* (PISA) 2022 which contains HOTS questions experienced a sharp decline in student performance (*steep learning loss*). Indonesia is ranked 68th globally, with a mathematics score of 379 (Kemendikbudristek, 2023). (Rodiah & Supardi, 2023) In their study on the development of mathematics HOTS test items, they developed five essay-type questions that were evaluated by validators. The results showed that four items (80%) were valid, while one item (20%) was invalid. The reliability of the test items was still low, with a reliability coefficient of 0.37. This is due to students' lack of understanding in problem-solving when attempting HOTS questions (Zulfah & Ulfa, 2020). Findings from Budiman & Jaliani (2014) in a private junior high school in Lampung City indicate that students' performance on higher-order thinking skills (HOTS) mathematics items, assessed through critical thinking indicators, remains relatively low, with an average score of 36%. The indicators related to analyzing and comprehending the problems reached 39%, the evaluation indicator yielded a score of 30%, and the indicator for designing appropriate solutions attained a score of 40% among students in Phase D.

The low high-level thinking skills of students are influenced by various factors (Karim, 2015). One of the factors that causes low high-level thinking skills or HOTS in students is the lack of habituation and training in working on contextual problems that require analytical, argumentative, and creative thinking skills in the solution process (Budiman & Jaliani, in Ramadhani, 2021). According to Rahayu, Suryana & Pranata (2020), the lack of habituation to working on HOTS-based questions is due to teachers still predominantly providing HOTS-based questions. *Lower Order Thinking Skills* (LOTS) with cognitive levels C1, C2 and C3. One of the reasons why HOTS questions are rarely given is that teachers' abilities to develop instruments with high-level abilities are still lacking, and the available references related to HOTS questions are still limited (Dinni, 2018; Rahayu et al., 2020). Therefore, questions based on high-level thinking skills are still very much needed (Hidayat, 2023).

Research from (Saputri, Hariyadi & Kamid, 2021) and (Khoriyah & Oktiningrum, 2021), and (Azmi, 2019) explained that by developing HOTS questions, students are motivated to work on questions because learning becomes more meaningful because it is presented with stimuli and question contexts related to everyday life, so that students are expected to understand more easily and be able to solve the problems given. One application of mathematics learning that can be presented in a way that is relevant to students' daily lives is the material on relations, functions, linear equations, and linear equations of two variables in phase D. (Kemendikbudristek, 2024)

Due to the conditions that occur and the importance of high-level thinking skills (HOTS) for students to have, researchers are encouraged to research the development of HOTS questions with the material of relations, functions, linear

equations, and systems of linear equations of two variables, phase D (Verina, Hanifah, Irsal & Stiadi, 2022). This study aims to produce Higher Order Thinking Skills (HOTS) test instruments that are valid, reliable, and have good discrimination power and difficulty levels so that they can enhance high-level thinking skills of students at phase D (Purbonugroho, Wibowo & Kurniawan, 2020) and (Khaerunnisa & Pamungkas, 2019). This research focuses on developing HOTS test instruments that cover the topics of relations, functions, linear equations, and systems of two-variable linear equations, based on the AKM (Minimum Competency Assessment) and HOTS mathematics question books. The HOTS questions include cognitive levels C4 (analyzing), C5 (evaluating), and C6 (creating) in accordance with Bloom's Taxonomy, revised by Krathwohl and Anderson. (Kemendikbud, 2019) As stated by the Ministry of Education, Culture, Research, and Technology (2022), the development of HOTS questions must consider the following: (1) using stimuli in the form of images, texts, passages, tables, graphs, or educational problems; (2) using new contexts either in the material or in the formulation of the questions. The questions are written in essay form and meet standards of validity, reliability, as well as optimal levels of difficulty and discrimination.

This research utilizes a development studies framework made up of two phases: the initial phase covering analysis and design, followed by the formative evaluation phase, which encompasses self-assessment, expert evaluation, individual interviews, small group sessions, and field testing (Saraswati & Agustika, 2020). The results of this study produced 17 questions categorized as valid, with a reliability coefficient of 0.848; 14 questions with good discrimination power and 3 questions with adequate discrimination power; and all 17 HOTS questions in this study fall into the medium difficulty category.

The HOTS questions developed are in the form of descriptions that can serve as a reference for educators in creating assessment instruments for formative and summative tests (Febiana, 2019).

II. Research Method

In this research, type development studies are employed, which are divided into two phases: the preliminary and formative evaluation stages (Ananda & Retnawati, 2023) The preliminary stage is further split into analysis and design phases, while the formative evaluation stage follows Tessmer's framework, involving five steps: self-evaluation, expert review, one-to-one, small group, and field testing (Rully in Suhady, Roza & Mimunah, 2020).

Data sources were obtained using validation sheets, student response questionnaires, and HOTS test questions on the material of relations, functions, linear equations, and two-variable linear equations in phase D. The validation sheet was used to obtain external validity values from the questions developed and then validated by the validator at the stage *of expert review*. The assessment carried out by the validator on the HOTS questions developed includes 3 aspects, namely the material/content, construction, and language aspects. The student response questionnaire is used to obtain input and comments from students regarding the questions developed, which are carried out during the s*mall group* and *field test stages.* HOTS test questions on the material on relations, functions, linear equations, and linear equations with two variables are carried out at the stage *field test* in order to obtain the validity value of the questions, reliability, discriminatory power, and level of difficulty of the questions developed.

At the *small group level,* Internal validity assessment of HOTS questions on material related to relations, functions, linear equations, and two-variable linear equations was carried out by 3 validators using a scale linked to a score of 1 to 5. The validation sheet grid is presented in Table 1 below.

Table 1. Validation assessment criteria

No	Aspect	Number of Indicators
1	Material Validation Sheet	3
2	Construction Validation Sheet	4
3	Language Validation Sheet	3
	Number of Indicators	10

The scores obtained from the validator assessment are tabulated and analyzed to determine the internal validity value of the HOTS questions on the material of relations, functions, linear equations, and two-variable linear equations (Azwar, 2019). The values in the range$0\% - 20\%$ categorized as very invalid, $21\% - 40\%$ categorized as weak/less valid, $41\% - 60\%$ categorized as quite valid, $61\% - 80\%$ categorized as strong/valid, and range $81\% - 100\%$ categorized *as* very strong/valid (Saadah in Ramadhani, 2021).

The validity analysis of HOTS question items on the material of relations, functions, linear equations, and linear equations of two variables was obtained by testing the developed question items on the research trial subjects of 33 class VIII students of SMPN 21 Pekanbaru were selected by the teacher based on their mathematics learning outcomes and were grouped into students with high, medium, and low abilities. The results of the students' answers were tabulated and analyzed to calculate the correlation for each question item using SPSS version 25. The steps used are to enter the tabulated data into SPSS, change the name to the adjusted data and change the decimal to the number 0, select *analyze,* click *correlate*, choose *bivariate*, move all variables from the left column to the right column, select Pearson, *two-tailed*, and click *flag significant correlation*, then click ok. After getting output in the form of value/person *correlation,* pay attention to the total score section. If the obtained person *correlation* $> r_{table}$, with $r_{table} = 0.3440$ for one star (*) and $and r_{table} = 0.4421$ for two stars (**), then the item is said to be valid. However, if Pearson *correlation* $\leq r_{table}$, then the question item is said to be invalid.

The reliability and discriminatory power analysis of HOTS questions were also calculated using SPSS version 25 software (Fatimah & Alfath, 2019). Techniques *internal consistency is used* to measure reliability by testing HOTS instruments or questions only once on students who are the subjects of the research trial, and the results of the trial data are analyzed (Rusdi, 2019). The steps used are using student answer data that has been tabulated, and entered into SPSS for validity calculations, select *analyze,* click *scale,* choose *reliability analyze*, move all the variables in the left column except the total score, click *statistics,* in the column *descriptive for*, click *item, scale*, And *scale if item deleted,* then click *continue*, the output for reliability will be obtained, namely *Cronbach's Alpha.* HOTS questions are said to be reliable or consistent if they are obtained. Cronbach's *Alpha* $> 0{,}60$. The output obtained for the discriminatory power is corrected, which will be accepted in this study if the *corrected* value is ≥ 0.30 in the sufficient category.

Bagiyono (2017) The difficulty level analysis is calculated using *software* SPSS version 25 using the data listed from the validity calculation, click analyze, choose *descriptive statistics*, then *frequencies,* move all variables in the left column to the right except the total score. On the menu *statistics,* click *mean, continue* and click OK. The output obtained is the Mean value that will be accepted in this study, provided it falls within the range $0.30 < M \leq 0.70$.

III. Results and Discussion

Results

This research employs a development model, specifically a type of *development study*, consisting of two stages: *preliminary* and *formative evaluation.* The *preliminary level consists* of analysis and design, while the *formative evaluation* stage comprises *self-evaluation, expert review, one-on-one meetings, small group* discussions, and *field tests*. The following is a description of the research results at each phase that has been carried out:

Level *Preliminary*

At the implementation level, *preliminary* work, specifically designing questions, is carried out using *Higher Order Thinking Skills* (HOTS) on material relations, functions, linear equations, and systems of linear equations in two variables, in phase D. *This preliminary* work is carried out in two stages: product analysis and design.

Analysis at the *preliminary level is* divided into three, namely needs analysis, student analysis, and curriculum analysis. A needs analysis is carried out through a review of relevant literature, observation, and structured interviews with junior high school/Islamic junior high school mathematics teachers in Pekanbaru regarding the assessment instruments used in mathematics learning. Based on the results of the literature review, observation of assessment instruments, and interviews, it was found that the questions given to students were rarely given a high-level thinking ability level, namely at the level of analyzing (C4), evaluating (C5) and creating (C6).

The analysis of students was conducted by reviewing various relevant literature. The students in question are those aged 12-15 years. According to *Piaget,* children in the 11-18 years age range have already developed formal operational thinking and can think abstractly, reason logically, and conclude (Meranti in Ramadhani, 2021), making HOTS questions suitable for phase D students. Students who are used as test subjects are those who have studied material relations, functions, linear equations, and systems of linear equations in the two-variable phase D. Therefore, you should be able to work on HOTS questions.

Curriculum analysis is carried out to form a match between the development of HOTS questions and the demands of the independent curriculum. The implementation of curriculum analysis is carried out to examine the learning objectives of material relations, functions, linear equations, and systems of linear equations in Phase D, which requires students to have them.

The results of the analysis obtained are the basis for designing research products to be developed. The products obtained at the design stage include grids, HOTS question materials, relations, functions, linear equations, and systems of linear equations of two variables, alternative solutions, scoring guidelines, validation sheets, and student response sheets. The HOTS questions produced totalled 17 descriptive questions, with 7 questions at cognitive level C4 (analyzing), 7 questions at cognitive level C5 (evaluating), and 3 questions at cognitive level C6 (creating), which were developed by adjusting the formulated question grids. The scoring guidelines have been developed holistically, where each step of the work contains categories that align with the HOTS question indicators of the revised Bloom's Taxonomy by Krathwohl and Anderson. Each category has a maximum score of 4 and a minimum score of 0, based on the student's answer.

Level *Formative Evaluation*

Level *formative evaluations in* this study include:

Self-Evaluation

Implementation-level *self-evaluation further* assessment is carried out by the researcher, together with the supervising lecturer, on the products produced at the preliminary stage. Improvements made at this stage include grammatical errors, the use of tables in designing products and assessment instruments. After improvements are made, the developed product is allowed to continue to the next stage.

Expert Review

Implementation level *expert review*: A validation assessment of the research product was carried out by three validator lecturers covering aspects of material/content, construction and language in order to obtain an internal validity score for the HOTS questions, material relations, functions, linear equations, and systems of linear equations of two variables that were developed. The results of the assessment, as obtained by the validator, included both qualitative and quantitative data.

The qualitative data obtained were in the form of validator suggestions and input, which were used as considerations in revising the research product. The decision to improve the question grid focused on refining the wording of the question indicators to match the cognitive level and the HOTS questions that had been developed. The following is an example of improving the wording of HOTS question indicators, material relations, functions, linear equations, and systems of linear equations in two variables, No. 1,which are adjusted to the questions developed according to the Table 2 below.

Table 2. Improvements to the question grid

Before Revision	After Revision
Given a problem related to sets related to SPLDV. Students are asked to design an arrow diagram that connects two sets of data.	Given 2 sets, set A, which contains the system of linear equations in 2 variables and set B, which contains the solutions of the system of linear equations in 2 variables in set A. Students are asked to describe the relation from set A to set B using an arrow diagram.

Quantitative data are obtained from the numbers provided by the validator regarding the instrument assessment, enabling the calculation of a score for internal validity. The recap of the average score of the three validators' assessments is 84.84% for the material aspect, 82.25% for the construction aspect and 82% for the language aspect. Thus, the average score of the validator's assessment for the three aspects of the HOTS questions is obtained material relations, functions, linear equations, and systems of linear equations with two variables amounted to 83.03% in the very valid category, resulting in HOTS questions material relations, functions, linear equations, and systems of linear equations in two variables that are internally valid.

One-to-one

The one-to-one level was conducted simultaneously with the expert stage review. During its implementation, the HOTS (Higher Order Thinking Skills) questions created during the design phase were administered to three students from MTs Muhammadiyah 02 Pekanbaru. This group included one student with high ability, one with medium ability, and one with low ability. Once the students finished the HOTS questions, a discussion session took place to gather feedback and evaluations from the students about the readability of the developed HOTS questions.

Students' suggestions and input at this stage, *one-to-one,* including: (1) improving the quality of the chart in question No. 2 and the image in No. 17; (2) improving the wording of question No. 13; and (3) improving the table to be colored to make it look more attractive. There are some suggestions and input from students were received by researchers and used as a basis for revising the research product. After revising the product and being declared valid, the researcher was permitted to proceed to the following research stage, namely the small *group*.

Small Group

Level *small group* conducted by testing 12 students, consisting of 4 students with high abilities, 4 students with medium abilities, and 4 students with low abilities. Students worked on HOTS questions, material relations, functions, linear equations, and systems of linear equations in two variables and after completion, students were asked to fill out a student response questionnaire containing suggestions and input related to the HOTS questions that had been worked on.

Students' suggestions and input at this stage, focused on *small group*s, were primarily concerned with typos in questions and the use of colour in tables and images, which researchers used to improve the HOTS questions that were developed. From the results of the student response questionnaire, the average percentage of all statements was 85.33% and it can be concluded that students gave a positive response to HOTS questions on material relations, functions, linear equations, and systems of linear equations in two variables that are developed, so it is justified to

continue the next stage *field test*.

Field Test

A level *field test* was implemented by revising HOTS questions with a *small group* of 33 students from class VIII.4 SMPN 21 Pekanbaru. The results of students' answers to HOTS questions were analyzed and then the validity value of the question items, reliability, level of difficulty and discriminatory power were determined. The validity value of the question items was analysed using SPSS version 25 to obtain output in the form of a Pearson *correlation*. Then, the value *of the Pearson correlation* for each question item was compared with *the* r_{table} $(df = n - 2)$. If obtained *pearson correlation* $> r_{table}$, with $r_{table} = 0{,}3440$ for one star (*) and $r_{table} = 0{,}4421$ for two stars (**), then the item is said to be valid. However, if Pearson *correlation* $\leq r_{table}$, then the question item is said to be invalid. The following are the results of the validity test calculation for the question items obtained.

Table 3. Results of the calculation of the validity test of question items

No. Question	*Pearson Correlation*	Category
1	0,439*	Valid
2	0,719**	Valid
3	0,518**	Valid
4	0,530**	Valid
5	0,429*	Valid
6	0,573**	Valid
7	0,493*	Valid
8	0,550**	Valid
9	0,542**	Valid
10	0,419*	Valid
11	0,650**	Valid
12	0,559**	Valid
13	0,636**	Valid
14	0,601**	Valid
15	0,589**	Valid
16	0,597**	Valid
17	0,490**	Valid

The outcomes from the calculation of the validity assessment for the question items shown in Table 3 indicated that all 17 HOTS questions developed were categorised as valid. Moving forward, these HOTS questions underwent tests for reliability, difficulty level, and discriminatory power. The reliability test was executed using SPSS version 25, yielding a reliability coefficient of 0.848, which falls under the very high category. Consequently, 17 HOTS questions, focusing on the topics of relations, functions, linear equations, and systems of linear equations with two variables, in phase D, were deemed both valid and reliable. The difficulty level analysis of each question item was conducted to ascertain the difficulty level of the HOTS questions. In this investigation, Table 4 below presents the findings regarding the difficulty level of each question item.

Table 4. Results of the calculation of the level of difficulty of question items

No. Question	Mean	Category
1	0,68	Currently
2	0,50	Currently
3	0,62	Currently
4	0,38	Currently
5	0,54	Currently
6	0,48	Currently
7	0,32	Currently
8	0,31	Currently
9	0,42	Currently
10	0,45	Currently
11	0,36	Currently
12	0,49	Currently
13	0,36	Currently
14	0,46	Currently
15	0,43	Currently
16	0,52	Currently
17	0,50	Currently

Table 4 shows that the level of difficulty obtained from 17 questions is in the medium category. In accordance with the criteria for the level of difficulty, as outlined by Rahman and Narsyah (2019), the difficulty level falls within the range of 0.30-0.70, indicating that the HOTS question has a good level of difficulty.

Meanwhile, the results of the analysis of the discriminating power for each HOTS question item on the material of relations, functions, linear

equations, and systems of linear equations with two variables in Phase D are presented in Table 5.

Table 5. Results of the calculation of the distinguishing power of question items

No. Question	Distinguishing Power	Category
1	0,36	Enough
2	0,69	Good
3	0,42	Good
4	0,43	Good
5	0,31	Enough
6	0,48	Good
7	0,42	Good
8	0,48	Good
9	0,44	Good
10	0,32	Enough
11	0,59	Good
12	0,47	Good
13	0,58	Good
14	0,55	Good
15	0,50	Good
16	0,52	Good
17	0,42	Good

Table 5 shows the results of the discriminatory power of each of the 17 questions received. In accordance with the discriminatory power criteria proposed by Rahman and Narsyah (2019), questions are categorised as having sufficient discriminating power, and 14 questions are categorised as having good discriminating power. All questions developed have acceptable discriminating power values, ensuring that HOTS questions on material related to relations, functions, linear equations, and systems of linear equations of two variables in Phase D of the final are valid, reliable, and have a good level of difficulty and discriminating power.

Discussion

Based on the description of the research results, the final product obtained in the form of HOTS questions on the material of relations, functions, linear equations, and linear equation systems of two variables, phase D, which are valid, reliable, have differentiating power and difficulty that are in accordance with the criteria that have been set. At the *preliminary level,* divided into analysis and design, researchers analyzed a starting point for implementing research aimed at establishing the primary foundation in the scope of developing HOTS questions on the material of relations, functions, linear equations, and systems of linear equations of two variables. The analysis carried out at this stage includes needs analysis, student analysis, and curriculum analysis. Implementing the needs analysis, researchers focused on the main problems in the research so that it was necessary to develop HOTS questions on the material of relations, functions, linear equations, and systems of linear equations of two variables phase D. The problem of this research is due to the lack of availability and ability of educators in developing HOTS questions to be used as assessment instruments in supporting efforts to train and improve students' high-level thinking skills.

After obtaining the primary basis of the research problem, the researcher then analyzed the characteristics of students through related literature so that the HOTS questions developed were in accordance with the stages of student thinking. The analysis of the mathematics learning curriculum includes an examination of Learning Outcomes (CP) related to the material on relations, functions, linear equations, and systems of linear equations of two variables in the D phase of algebraic elements, to describe several Learning Objective Achievement Criteria (KKTP). The HOTS questions developed are also accompanied by new stimuli and contexts. The development of HOTS questions on the material of relations, functions, linear equations, and systems of linear equations of two variables is also an effort by researchers to make it easier for students because they are related to everyday life.

In light of the analysis outcomes, the researcher crafted and designed a research product comprising a grid, HOTS questions, alternative solutions, scoring guidelines, validation sheets, and student response questionnaires. The HOTS questions, developed during the design phase, included 17 descriptive inquiries that adhered to the revisions of Bloom's Taxonomy by Krathwohl and Anderson. These

consisted of 7 questions at the C4 cognitive level (analyzing), 7 questions at the C5 level (evaluating), and 3 questions at the C6 level (creating), aligned with the developed question grid. The questions have been supplemented with alternative solutions and scoring guidelines tailored to high-level thinking indicators. The HOTS questions developed will serve as a question bank that educators can use as a reference for developing HOTS questions or as a learning resource for students in training and improving their high-level thinking skills. A total of 17 HOTS questions were obtained, consisting of 7 questions at level C4 (analyzing), 7 questions at level C5 (evaluating), and 3 questions at level C6 (creating), which were approved for further validation by three validator lecturers.

At the implementation level, the *expert review* validated the HOTS questions developed by three mathematics education lecturers. Three aspects were validated: the material/content aspect, the construction aspect, and the language aspect, to ensure internal validity. The assessment components of each aspect are assessed by the validator have different assessments. The differences are outlined in the validation sheet rubric, which presents the assessment criteria for each scoring component in each assessment aspect. The validators conducted validation in different periods, namely from April 28 to May 31, 2025.

The validator's assessment results for all aspects of each question developed yielded an average score of 83.07%, categorizing them as very valid. The average scores varied across different aspects. For the material/content aspect, the average was 85%, which was very valid. In the construction aspect, the average score was 82.2%, which is also very valid. For the language aspect, the average score was 82%, maintaining the very valid category.

Implementation of validator evaluation at the *expert review* level included 17 HOTS questions that were developed during the self-*evaluation* stage. These questions were also tested on 3 students at the one-to-one stage. Implementation level *one-to-one conducted* on May 16, 2025, with a HOTS question completion time of 2 x 40 minutes. After completing the HOTS questions, the researcher and students conducted unstructured interviews to gather suggestions and feedback on the HOTS questions that had been developed. Students' suggestions and input included improvements to writing, ambiguous question formulations, colour quality in tables and images, and the time given to complete the HOTS questions. After the revision was carried out according to the validator's suggestions and input, the researcher was allowed to continue the research at the next stage *small group.*

Implementation level *small group* HOTS test questions were conducted on 12 students. The implementation time given was 3 x 40 minutes on June 4, 2025, in a face-to-face setting. The addition of work time was based on suggestions and input from students at the one-to-one *stage.*

The final stage is the implementation stage, which includes the *field test.* HOTS questions on the material of relations, functions, linear equations, and systems of linear equations of two variables were tested on 33 students of class VIII.4 of SMPN 21 Pekanbaru to obtain the validity value of the questions, reliability, discriminating power, and level of difficulty of each question that had been developed. The trial was carried out on June 5, 2025, at SMPN 21 Pekanbaru with a working time of 3 x 40 minutes. After the HOTS questions were finished, students filled out the response questionnaire that had been provided. The average result obtained from student responses was 87.26%, which met very good criteria. This means that the questions developed are practical for students to use, allowing for the final HOTS questions on the material of relations, functions, linear equations, and systems of linear equations with two variables to be obtained.

The results of the validity analysis obtained 17 HOTS questions categorized as valid, so that it can be continued to calculate the reliability, level of difficulty and differentiating power using SPSS

version 25. The results of the reliability analysis with SPSS obtained a value of 0,848with a very high category. In the results of the analysis of the level of difficulty obtained all questions developed were categorized in the medium category. Finally, for the analysis of the discriminating power obtained, the study accepted the following values: 3 questions with a sufficient discriminating power category and 14 questions with a good discriminating power category. From the results of the validity of the questions, reliability, discriminating power and level of difficulty obtained, it can be concluded that the HOTS questions on the material of relations, functions, linear equations, and systems of linear equations of two variables that were developed are in accordance with having acceptable values.

IV. Conclusion

Based on the results of the research and discussion, it was concluded that 17 HOTS questions (*Higher Order Thinking Skills*) material on relations, functions, linear equations, and systems of linear equations in two variables for Phase D students developed using the model development *studies* which consists of levels *preliminary* (analysis and design) and level *formative evaluation* (*self-evaluation, expert review, one-to-one, small group dan field test*) has met the criteria of being valid, reliable and has good difficulty and differentiating power.

References

Ananda, A. S., & Retnawati, H. (2023). Analisis muatan higher order thinking skills buku teks matematika SMP kelas VIII. *JNPM (Jurnal Nasional Pendidikan Matematika)*, *7*(2), 367. Https://Doi.Org/10.33603/Jnpm.V7i2.8084

Anita, Jumroh, & Retta, A. M. (2023). Pengembangan soal higher order thinking skills (HOTS) pada materi statistika SMP kelas VIII. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(2), 2123. Https://Doi.Org/10.24127/Ajpm.V12i2.7133

Ariani, K. D., Candiasa, I. M., & Sariyasa. (2023). Pengembangan instrumen penilaian berbasis berbasis HOTS. *Jurnal Ilmiah Global Education*, *4*(2), 640–649.

Azmi, M. P. (2019). Analisis pengembangan tes kemampuan analogi matematis pada materi segi empat. *Journal For Research In Mathematics Learning*, *2*(2), 99–110.

Azwar, S. (2019). *Reliabilitas dan validitas* (4th Ed.). Pustaka Belajar.

Bagiyono. (2017). Analisis tingkat kesukaran dan daya pembeda butir soal ujian pelatihan radiografi tingkat 1. *Pusdiklat*, *16*(1), 1–12.

Budiman, A., & Jaliani. (2014). Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1. *Jurnal Riset Pendidikan Matematika*, *1*(2), 139–151.

Dinni. (2018). HOTS (High order thinking skills) dan kaitannya dengan kemampuan literasi matematika. *PRISMA*.

Fatimah, L. U., & Alfath, K. (2019). Analisis kesukaran soal, daya pembeda dan fungsi distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, *8*(2), 37–64.

Febiana, M. V. (2019). *Pengembangan soal HOTS materi luas bangun datar dan volume bangun ruang untuk siswa kelas V SD*. Universitas Sanata Dharma.

Hadi, S., & Novaliyosi. (2019). *TIMSS Indonesia (Trends in international mathematics and science study)*.

Hidayat, S. (2023). *Analisis kemampuan siswa dalam menyelesaikan soal-soal HOTS pada pembelajaran matematika kelas V di SDN 12 Rejang Lebong*. IAIN Curup.

Izzati, N., & Febrian, F. (2021). Kemampuan mahasiswa calon guru dalam mengembangkan instrumen tes hasil belajar kategori higher order thinking skill. *Jurnal Gantang*, *6*(1), 75–82. Https://Doi.Org/10.31629/Jg.V6i1.2672

Karim, N. (2015). Kemampuan berpikir kritis siswa dalam pembelajaran matematika dengan menggunakan model jucama di sekolah menengah pertama. *EDU-MAT:*

Jurnal Pendidikan Matematika, *3*(1), 92–104.

Kemendikbud. (2019). *Panduan penulisan soal HOTS* (P. 2).

Kemendikbudristek. (2023). *Laporan PISA*

Kemendikbudristek. (2024). *Keputusan Kepala Badan Standar, Kurikulum, Dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi Nomor 032/H/KR/2024*.

Khaerunnisa, E., & Pamungkas, A. S. (2019). Pengembangan instrumen kecakapan matematis dalam konteks kearifan lokal budaya banten pada materi bangun ruang sisi datar. *KREANO: Jurnal Matematika Kreatif-Inovatif*, *9*(1), 17–27.

Khoriyah, M., & Oktiningrum, W. (2021). Pengembangan soal higher order thinking skills (HOTS) berbasis budaya lokal blitar untuk mengukur dimensi pengetahuan matematika siswa kelas V sekolah dasar. *Bina Gogik*, *8*(1), 93–100.

Purbonugroho, H., Wibowo, T., & Kurniawan, H. (2020). Analisis berpikir kritis siswa dalam menyelesaikan masalah open ended matematika. *Jurnal Ilmiah Pendidikan Matematika*, *7*(2), 53–62.

Rahayu, S., Suryana, Y., & Pranata, O. H. (2020). Pengembangan soal high order thinking skill untuk meningkatkan kemampuan berpikir tingkat tinggi matematika siswa sekolah dasar, *7*(2). Http://Ejournal.Upi.Edu/Index.Php/Pedadidaktika/Index

Rahman, A. A., & Nasryah, C. E. (2019). Evaluasi pembelajaran. *Evaluasi Pembelajaran*.

Ramadhani, A. S. (2021). *Pengembangan soal higher order thinking skills (HOTS) berkonteks budaya melayu riau pada materi segiempat dan segitiga kelas VII SMP/Mts*. Universitas Riau.

Rodiah, D., & Supardi. (2023). Pengembangan butir soal matematika berdasarkan standar hots pada materi sistem persamaan linear dua variabel. *Jurnal Pendidikan Indonesia*, *4*(6), 548–562. Https://Doi.Org/10.59141/Japendi.V4i6.1878

Rusdi, M. (2019). *Penelitian desain dan pengembangan kependidikan* (1st Ed.). Rajawali Pers.

Saputri, R., Hariyadi, B., & Kamid. (2021). *Pengembangan soal higher order thinking skills berbasis budaya Jambi*. *05*(02), 1793–1806.

Saraswati, P. M. S., & Agustika, G. N. S. (2020). *Kemampuan berpikir tingkat tinggi dalam menyelesaikan soal HOTS mata pelajaran matematika*.

Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan soal untuk mengukur higher order thinking skill (HOTS) siswa. *Jurnal Gantang*, *5*(2), 143–150. Https://Doi.Org/10.31629/Jg.V5i2.2518

Tiwery, B. (2019). *Kekuatan dan kelemahan metode pembelajaran dalam penerapan pembelajaran HOTS (higher order thinking skills)*. Media Nusa Creative.

Verina, I., Hanifah, H., Irsal, N. A., & Stiadi, E. (2022). Analisis tingkat kognitif soal uji kompetensi buku teks matematika terbitan erlangga kelas VIII semester 1 berdasarkan taksonomi Bloom revisi. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *6*(2), 177–190. Https://Doi.Org/10.33369/Jp2ms.6.2.177-190

Wandari, A., Kamid, & Maison. (2018). Pengembangan lembar kerja peserta didik (LKPD) pada materi geometri berbasis budaya Jambi untuk meningkatkan kreativitas siswa. *Edumatika Jurnal Riset Pendidikan Matematika*, *1*(2), 47–55.

Zulfah, & Ulfa Insani, S. (2020). Pengembangan soal matematika berbasis kearifan lokal dan daya tarik wisata Riau pada tahap preliminary research. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *04*(02), 787–799.

JURNAL GANTANG. November 2025; *10*(2): 291 – 301
p-ISSN. 2503-0671
e-ISSN. 2548-5547