Proceedings of the 3rd International Conference on Social and Islamic Studies
Online ISSN: 2809-6339
Volume: 3, 2023
Publisher: Alauddin University Press

DEVELOPMENT OF HOTS TYPE TEST INSTRUMENTS TO MEASURE STUDENTS' CRITICAL THINKING ABILITY

Ramdani Sari[1], Fitriani Nur[2], Andi Dian Angriani[3], Ahmad Farham Majid[4], & Nurul Ainun Fajriah[5]

[1,2,3,4]Universitas Islam Negeri Alauddin Makassar
[5]Universitas Negeri Yogyakarta
Correspondence Email: fitrianinur@uin-alauddin.ac.id

ABSTRACT

This study aims to determine the procedure and to analyze the quality of the HOTS-type test instrument development for number theory courses, as well as to measure students' critical thinking skills. This type of research is research and development (Research and Development) with a formative research model (Tessemer). The test subjects in this study were students of mathematics education class PMAT A semester 2 of the Tarbiyah and Teacher Training Faculty, UIN Alauddin Makassar with a total of 23 people. Based on the results of this study, it is known that 1) the procedure for developing the Formative Research (Tessemeer) model goes through 4 stages, namely the preliminary stage, the self-evaluation stage, the prototyping stage, and the field test stage. 2) the quality of test development is stated to be valid with CVR (Content Validity Ratio) and CVI (Content Validity Index) values, the average total score is 1 with the very appropriate category, practically based on the results of the student response questionnaire analysis is 85.03% in the very positive response category, the results of the lecturer's response were 90% with a very positive response category. reliable with an average total score of 0.798 in the high category (reliable), the average difficulty level of the total score is 0.48 in the medium category, and the differential power of the average total score is 0.35 in the good category. 3) students' critical thinking skills are 47.67 which is in the sufficient category. Thus, it can be concluded that the HOTS-type test instrument for number theory courses to measure students' critical thinking skills is stated to be of good quality.

Keywords: Critical thinking; test instrument; HOTS

1. INTRODUCTION

Every activity that is carried out requires an evaluation process to assess the condition of an object using instruments (Thoha, 2003). The evaluation of learning in education aims to determine the abilities of students (Ahmad, 2015). Evaluation of learning outcomes in tertiary institutions is used to measure students' abilities and determine the success of lecturers in teaching. The form of testing to assess a person's ability is called a test (Arifah & Yustisianisa, 2012) in the form of questions prepared with learning indicators.

One of the lessons that take place in the Mathematics Education Study Program is number theory which involves higher-order thinking processes. In addition, in studying number theory it is hoped that students will have critical, logical, and systematic thinking patterns, as well as creativity in problem-solving (Afriansyah et al., 2020). Based on interviews conducted with several students, it was shown that the instruments often given by lecturers in learning number theory were in the form of descriptions, but based on existing textbooks or modules. In addition, the questions given by the lecturer are questions that have been used before so that students only write answers that are memorized without thinking critically, students need test questions that can train thinking skills, especially student teacher candidates. Not only that, but the researcher also identified the Number Theory Semester Program Plan (RPS) obtained from the powerful lecturer, the results obtained showed that the cognitive level that also supports the achievement of learning objectives includes bloom taxonomy, namely C4 and C5. This means that administratively the learning plan has been packaged well in constructing students' critical thinking abilities, it's just that in conducting evaluations in the field the use of instruments should be adjusted and redeveloped by looking at the characteristics and learning objectives to be achieved based on applicable guidelines. This is because the ability to think can affect the level of mastery of one's material. The low thinking ability of students triggers difficulties in mastering the material (Andriyani & Yenni, 2019).

Therefore, a test instrument is needed that is following the characteristics of the course. The number theory course test instrument is needed as a stimulus to trigger students to think critically. A good test instrument for measuring critical thinking skills is in the form of a description. Essay tests are good to use because they can measure students' high-level thinking skills and can reduce the possibility of students guessing answers. Apart from that, it can be used to determine the ability profile of students, it can also be used as a means to train students' abilities to think at a higher level. The questions used as exercises can contain questions that test students in terms of problem-solving, critical thinking, and creative thinking.

In evaluating cognitive level adjustments between the HOTS type test instrument and the ability to be measured, in this case, critical thinking when viewed based on Bloom's taxonomy indicates the compatibility of the term critical thinking ability with higher order thinking ability or HOTS (Jiwandono, 2019). That is, HOTS questions are measurement instruments used to measure higher-order thinking skills, one of which is critical thinking (Andriyani & Yenni, 2019).

Higher-order thinking skills or HOTS (Higher Order Thinking Skills) are the ability to process information by thinking critically, evaluating, and solving a problem (Kenedi, 2018). HOTS fields are analysis, evaluation, and creativity (Umami et al., 2021). Mathematics is a subject that requires individuals to think at a more important level or HOTS. As revealed by Hadi and Faradillah (2019) that higher-order thinking skills are basic abilities that must be developed in the process of learning mathematics.

This research is in line with research conducted by Arifin (2016) which shows that the HOTS test instrument can be used as an example or reference to be able to measure students' HOTS thinking skills in learning mathematics. In line with research conducted by Bakhri and Rosnawati (2018) shows that HOTS-based test instruments can trigger analytical thought processes by checking the correctness of the information and evaluating by assessing or selecting the right information from a problem. This instrument also creates thoughts that build, formulate, and summarize information from a problem. Thus, the HOTS-type evaluation tool has the potential to support and construct students' critical thinking.

Based on this, the researcher intends to develop a HOTS-type test instrument to measure critical thinking skills. This can train students to think critically. Readiness to face challenges in the 21st century can be started by practicing and getting used to working on questions that trigger critical thinking. Because this habit can be a provision for students to take on roles in the future. The test instrument developed is in the form of descriptive questions that can be used as a training evaluation tool or a reference in making student worksheets.

2. METHODS

The type of research used in this study is research and development (Research and Development). Research and Development with the Tessmer development model. Tessmer's formative research type development model consists of the preliminary stage, the self-evaluation stage, the formative evaluation (prototyping) stage which includes expert reviews, one-to-one, and small groups, as well as the field test stage (field trials) (Jurnaidi & Zulkardi, 2013).

This research was conducted at the Tarbiyah and Teacher Training Faculty of UIN Alauddin Makassar with the test subjects being 23 students of mathematics education class A. The data collection techniques in this study were: (1) Tests, (2) questionnaires consisting of student response questionnaires and lecturer response questionnaires, and (3) validation sheets. The data analysis technique used was content validity analysis, student and lecturer response analysis, student Islamic character activity observation sheet analysis, test instrument difficulty level analysis, test instrument differentiability analysis, and test instrument data analysis.

The test used is a HOTS-type instrument to measure students' critical thinking skills. The test that will be given to students is a description test derived from number theory course material. The questionnaire contains questions to obtain information, opinions, responses, and others. In this study, the questionnaire used was a validation sheet and student and lecturer response questionnaires. The research instrument is a measuring tool in the research process to collect data in a study. The instruments in this study were test instruments in the form of description questions, test instrument validation sheets, and response questionnaires.

The data obtained using the instrument will then be analyzed qualitatively and quantitatively. Qualitative data were obtained from validator input. While quantitative data is directed to determine the validity, and reliability of the items and student responses to the HOTS-type test instrument. Analysis of the content validity of the questions was carried out to determine the suitability of the questions with the material used as the basis for making the questions. The content validity test used is the content validity ratio (CVR). According to Lawshe, CVR is a content validity analysis approach to measure the degree of agreement of experts regarding the suitability of the question items with the material (Hendryadi, 2017).

$$CVR = \frac{2ne}{N} - 1$$

Description

ne = The number of expert reviews that stated valid

N = Number of expert reviewers who conducted the assessment

Furthermore, the overall validity value of the question can be determined using the Content Validity Index (CVI) with the formula (Bashooir & Supahar, 2018).

$$CVI = \frac{\sum CVR}{Total\ questions}$$

The results of the CVR and CVI calculations are in the form of a ratio of 0-1. Analysis of the responses of students and powerful lecturers can be analyzed using the formula (Sugiyono, 2016).

$$P = \frac{f}{N} \times 100\%$$

Description:

P = Percentage

f = Total score of data collection results

N = Sum of criterion scores

Then the results of the analysis obtained next are to conclude the results of calculations based on aspects (Khairiyah, 2019).

Table 1. Student and Lecturer Response Criteria

Assessment	**Criteria**
$85\% \leq x < 100\%$	Very positive
$70\% \leq x < 85\%$	Positive
$50\% \leq x < 70\%$	Less positive
$x < 50\%$	Not positive

Analysis of data on the results of instrument trials by calculating reliability, level of difficulty, and discriminating power and test results. Reliability indicates that the data is reliable and consistent. To measure the reliability of this instrument using the Chronbach-alpha formula,

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right)$$

Description :

r_{11} : reliability

σ_t^2 : total variance

σ_i^2 : variance of item i

n : the number of items

Assessment with reliability criteria (Cahyanti et al., 2019).

Table 2. Reliability Criteria

Score	**Criteria**
$0{,}90 < r \leq 1{,}00$	Very high
$0{,}70 < r \leq 0{,}90$	High
$0{,}40 < r \leq 0{,}70$	Medium
$0{,}20 < r \leq 0{,}40$	Low
$0{,}00 < r \leq 0{,}20$	Very low

The level of difficulty shows the degree of difficulty and ease of a test. The instrument difficulty index number can be obtained by the formula (Salmina & Adyansyah, 2017).

$$TK = \frac{Mean}{Maximum\ Score}$$

Table 3. Difficulty Criteria

Assessment	**Criteria**
$0{,}71 - 1{,}00$	Easy
$0{,}31 - 0{,}70$	Medium
$0{,}00 - 030$	Difficult

The discriminating power of the items shows the ability of the items to differentiate the abilities of the test takers. Discriminating power is calculated using the formula (Salmina & Adyansyah, 2017).

$$DB = \frac{X_t - X_r}{X_{maks}}$$

DB = Power Difference

X_{maks} = maximum score

X_t = average score for the high group

X_r = average score of the low group

Table 4. Criteria for Different Power

Assessment	Criteria
$0{,}50 < x \leq 1{,}00$	Very good
$0{,}30 < x \leq 0{,}49$	Good
$0{,}20 < x \leq 0{,}29$	Enough
$0{,}00 < x \leq 0{,}10$	Less enough
$x \leq 0{,}00$	Not good

Analysis of student test data was carried out to determine students' higher-order thinking skills by correcting answers based on predetermined scoring indicators. Furthermore, it can be determined the category of student abilities with the category table (Sagala, 2019).

Table 5. Criteria for Students' Critical Thinking Ability

Assessment	Criteria
$80 < value \leq 100$	Very good
$60 < value \leq 80$	Good
$40 < value \leq 60$	Enough
$20 < value \leq 40$	Less
$0 < value \leq 20$	Very less

3. RESULTS AND DISCUSSION

The process of developing HOTS-type test instruments for number theory courses to measure students' critical thinking skills uses the formative research (Tessemer) development model. The following are the stages of developing a test instrument.

Preliminary Stage

This stage begins with the collection of references related to this research, namely the development of a HOTS-type test instrument for number theory courses to measure students' critical thinking skills. Furthermore, activities are carried out to determine the place and subject of the trial. The test site for this study was the Tarbiyah and Teacher Training Faculty of UIN Alauddin Makassar. While the test subjects in this study were 23 students of mathematics education class A semester 2.

Self-Evaluation Stage

This stage consists of two parts, namely the analysis stage and the design stage. The analysis phase consists of curriculum analysis, student analysis, and material analysis.

Curriculum analysis is carried out to find out how lesson plans are used as a reference in the development process. The learning model used by lecturers is the cooperative AIR type. This learning model is very supportive and has a major influence on students' critical thinking skills. Student analysis focused on students of mathematics education class A semester 2 at UIN Alauddin Makassar who were used as test subjects. The number of students in the mathematics education class A semester 2 is 23 students who have different cognitive abilities based on their respective study results cards. Furthermore, material analysis is focused on identifying, compiling, and detailing the materials studied by students which will be used in tests in number theory courses. Based on the results of the material analysis, it was found that the questions to be developed were questions that were following the 5 subject matter and learning indicators in the textbooks or number theory modules used in class. The five main topics of discussion include prime numbers, single factorization, congruence, applications of congruence and Linear congruence, and the Diophantus linear equation. After the researcher has gone through the analysis phase, the next stage is design. At this stage, the researcher designed and designed the test instruments, including grids, test questions in the form of descriptions, answer keys, and scoring guidelines.

Prototyping Stage

At this stage, all research instruments that have been designed by researchers will be evaluated. Test instruments that have been made based on self-evaluation of a prototype I are given to the expert review group, one-to-one, and the small group. The prototype I was given to expert review for instrument validation (grids, test questions, and answer keys). The one-to-one stage was carried out to 3 4th semester mathematics education students who had passed or studied number theory courses in the previous semester. The three students consist of students with high, medium, and low abilities based on data obtained from student study results cards. The selection criteria for the three students intended to determine the legibility of the instrument based on their level of ability. When the test instrument has been tried out, students are then asked to fill out a questionnaire and write comments about the HOTS-type test instrument for the number theory course. The revised results of the expert review and one-to-one stages are called prototype II. Then in the small group stage, prototype II was tested on 6 4th semester mathematics education students as test subjects, each of which was two high-ability students, two medium-ability students, and two low-ability students. After working on the questions given, the six students filled out a questionnaire to provide an overview of the HOTS-type test instrument. The results of the revision of this stage are called prototype III.

Field Test Stage

The results of prototype III were tested on test subjects, namely students in mathematics education class A semester 2 at UIN Alauddin Makassar, totaling 23 people. The trial was conducted in one meeting, namely Thursday, June 23, 2022, with a total of 15 item numbers for HOTS-type description questions with an allocation of 100 minutes of processing time. The trial systematics at this stage, namely before the date of the trial, students are directed to study the modules given by the lecturer. After that, on the trial day, students were given HOTS-type test questions to work on.

Development Results

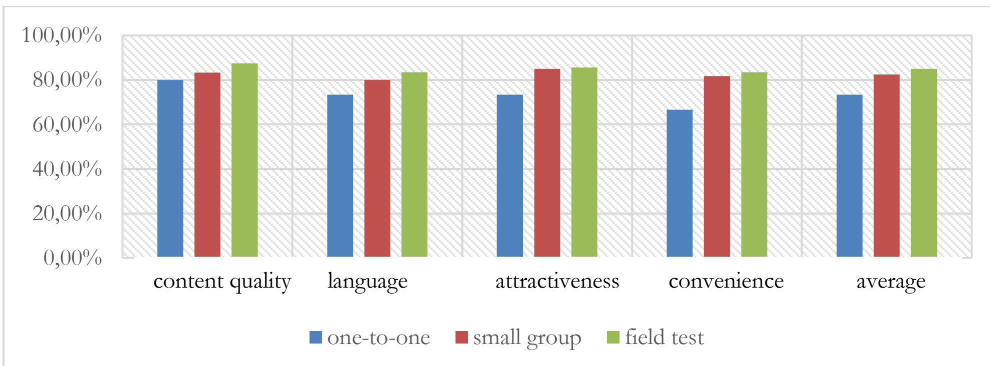
The validation analysis method uses the CVR (Content Validity Ratio) and CVI (Content Validity Index) methods. The results of the validation analysis are as follows.

Table 6. CVR and CVI Validation Results After Revision

Question items	Expert 1	Expert 2	CVR	CVI	Description
1	Yes	Yes	1		Item supports
2	Yes	Yes	1		Item supports
3	Yes	Yes	1		Item supports
4	Yes	Yes	1		Item supports
5	Yes	Yes	1		Item supports
6	Yes	Yes	1		Item supports
7	Yes	Yes	1		Item supports
8	Yes	Yes	1		Item supports
9	Yes	Yes	1	1	Item supports
10	Yes	Yes	1		Item supports
11	Yes	Yes	1		Item supports
12	Yes	Yes	1		Item supports
13	Yes	Yes	1		Item supports
14	Yes	Yes	1		Item supports
15	Yes	Yes	1		Item supports

The results of content validation in table 1 show that of the 15 items assessed by the validators (experts), it shows that these items support the validity of the test. Then from the results of the Content Validity Ratio (CVR), the value of the Content Validity Index (CVI) is obtained, which is the average of the Content Validity Index (CVI) for all items of 1, meaning "very appropriate" about the topic to be analyzed. This means that the HOTS-type test instrument for number theory courses to measure students' critical thinking abilities is very appropriate (valid) with the material or topic being measured.

The results of the analysis of student responses in the one-to-one trial obtained an average score of 73.33%. This score is included in the interpretation criteria of $70\% \leq x < 85\%$ with the "Positive" category. Meanwhile, the small group trial obtained an average score of 82.78%. This score is included in the interpretation criteria of $85\% \leq x < 100\%$ with the attractiveness category "Very Positive". It can be concluded that the legibility of the questions is good and can be continued to the field trial stage. Based on the analysis of student responses at the field test stage, an average score of 85.03 was obtained. This score is included in the interpretation criteria $of\ 85\% \leq x < 100\%$ with the "Very Positive" category. From the three stages of the trials that have been carried out, the results of student responses to the HOTS-type test instrument in number theory courses to measure students' critical thinking abilities are shown in graph 1:



Graph 1. Student Response Analysis

Based on the student response analysis graph above, it can be seen that on average several assessment indicators (content quality, language, attractiveness, and convenience) show a response value above 70%. This means that the student's response to the number theory course test instrument is 'positive' so that the product being developed meets the practical criteria.

The results of the analysis of the lecturers' responses obtained an average score of the percentage of lecturers' responses of 90%. This percentage is included in the interpretation criteria of $85\% \leq x < 100\%$ with the category "Very Positive". This shows that the HOTS-type test instrument for number theory courses meets practical criteria.

Based on the results of the reliability analysis of the HOTS-type test instrument for number theory courses to measure students' critical thinking skills using SPSS Statistics 22, it was found that the test instrument was classified as reliable with an average Croncbach's alpha value of 0.798 with a high interpretation.

Then an analysis of the difficulty level of the HOTS-type test instrument for number theory courses was carried out:

Table 7. Analysis of the difficulty level of the test instrument

Question Items	Average Score	Maximum Score	Difficulty Level	Category
1	3.70	6	0.62	Medium
2	4.40	8	0.55	Medium
3	4.83	8	0.60	Medium
4	3.80	6	0.63	Medium
5	2.7	6	0.36	Medium
6	2.21	6	0.37	Medium
7	2.57	4	0.64	Medium
8	2.09	6	0.35	Medium
9	4.70	8	0.59	Medium
10	2.69	6	0.45	Medium
11	1.65	4	0.41	Medium
12	2.39	6	0.40	Medium
13	2.74	6	0.46	Medium
14	3.09	6	0.51	Medium
15	1.35	4	0.34	Medium
	Average		0.48	Medium

Based on table 2, it was found that the category of difficulty level at the trial stage with a total of 15 items was included in the medium category with an interval of 0.31-0.70. According to the criteria for the quality of the test instrument, the test instrument items can be said to be good if the test items have a level of difficulty at intervals of 0.31-0.70 in the moderate category.

The results of the analysis of the differentiating power of the test instrument items can be shown in table 3 below:

Table 8. Analysis of Different Power Tests

Question Items	Upper Band Average	Lower Group Average	Score Max	Discrimination	Category
1	5.45	2.08	6	0.56	Good
2	5.91	3.0	8	0.36	Good
3	6.36	3.42	8	0.37	Good
4	4.90	2.75	6	0.36	Good
5	3.45	1.00	6	0.41	Good
6	3.18	1.33	6	0.31	Good

Question Items	Upper Band Average	Lower Group Average	Score Max	Discrimination	Category
7	3.60	1.60	4	0.50	Good
8	3.45	0.83	6	0.44	Good
9	5.73	3.75	8	0.25	Enough
10	4.0	1.5	6	0.42	Good
11	2.18	1.16	4	0.25	Enough
12	3.64	1.25	6	0.40	Good
13	3.55	2.00	6	0.26	Enough
14	3.82	2.42	6	0.23	Enough
15	1.8	0.9	4	0.23	Enough
Average				0.35	Good

Based on table 3, it was found that the average power of differential HOTS-type test instruments for number theory courses to measure students' critical thinking abilities was 0.35. this means that the test instrument is in the criteria of 0.30-0.49 with a good category.

Table 9. Analysis of Student Critical Thinking Ability Test Results

Number of Questions	Student Grades	Frequency	Percentage (%)	Category
15	$80 < value \leq 100$	0	0	Very good
	$60 < value \leq 80$	11	47.82	Good
	$40 < value \leq 60$	0	0	Enough
	$20 < value \leq 40$	12	52.17	Less
	$0 < value \leq 20$	0	0	Very poor
	Number of subjects	23	100	
	Average value	47.67		Enough

Based on data analysis in table 4, information was obtained that from the results of testing the test instrument on 23 students, there were 11 students (47.82) who had good critical thinking skills and 12 students (52.17%) were in the very poor category of critical thinking skills. Thus, the average value of critical thinking ability of 23 students as research trial subjects was 47.67. The average value is in the critical thinking ability interval with the sufficient category, namely $40 < value \leq 60$. This shows that the critical thinking skills of students in mathematics education class A semester 2 at UIN Alauddin Makassar with a total of 23 students are in the sufficient category, which means they have sufficient critical thinking skills.

The development of the HOTS tie test instrument for the number theory course to measure students' critical thinking skills uses the formative research (Tessemer) development model. As found in instrument development research that is relevant to this research procedure, the Tessemer (1993) model used consists of several flow stages which include the preliminary stage, the self-evaluation stage, the formative evaluation (prototyping) stage which includes expert reviews, one to one, and small groups, as well as the field test stage (field trials) (Makmur, 2020). The same thing was also obtained in the test instrument development research conducted (Sagala & Andriani, 2019) using the Formative Research Tessemer development procedure. The research stages for developing HOTS-based tests with Islamic characters were carried out according to the research stages of the Tassemer model (Dwidelia, 2021). Several relevant studies above apply Tassemer's Formative Research development model which shows the existence of this model in producing good products, chosen because the steps used are clearer and more systematic (Makmur, 2020). In addition, there are differences in the procedure of this study with previous studies, including; 1) instrument development research was carried out (Tanujaya, 2016), using standard instrument development procedures, starting with developing conceptual definitions, developing operational definitions, determining constructs, dimensions, compiling items, expert validation, and trials. 2). The instrument development research carried out (Mas'ula & Rokhis, 2020), used standard ADDIE model development research procedures. Differences in previous research procedures with this research are based on the characteristics of the instruments developed and the initial motives for development. Development of appropriate models, approaches, and methods to support the effectiveness of student competence achievement.

The initial stage of the Formative Research model is the preliminary stage, where the researcher seeks and collects references that are relevant to the research to be carried out. Through the interview process, researchers dig up information related to the problems that occur. Determine the location and research subject based on the problems found. Based on preliminary information and assessment according to several relevant references, the research to be carried out is to produce a product in the form of a test instrument that supports students' critical thinking skills. This stage is in line with the research conducted Putra and Vebrian (2019) and Ningsih and Annajmi (2020) that at the preliminary stage, the research location was determined, the subject determined, and the research schedule. Research development conducted Sagala and Andriani (2019) and Hervanda et al. (2020) that at the preliminary stage researchers immediately prepare the materials needed for development and determine the location and research subjects. In line with research Jannatasari et al. (2017) that at the preliminary stage, schedule arrangements and collaboration with supporting lecturers are carried out where the research will be carried out.

The next stage is self-evaluation where at this stage the researcher conducts curriculum analysis, student analysis, material analysis, designs, and develops a HOTS-type test instrument for number theory courses to measure students' critical thinking skills in the form of grids, question designs, answer keys, and scoring guidelines. Curriculum analysis aims to find out how lesson plans and learning models apply in the target class. This is because the learning model determines the selection of strategies and methods for a lesson (Jannatasari et al., 2017). It is known that the learning model applied is the AIR-type model. This model is highly believed to have a good influence on critical thinking skills (Kamarullah, 2017). Meanwhile, student analysis is intended to gather information about the number and different cognitive abilities, and material analysis is carried out to determine the material and its suitability with the indicators to be used in the test instrument to be developed. This stage is in line with research conducted by As'ari et al. (2019) and Nursalam et al. (2017) that the activities in the analysis stage are analyzing the applicable curriculum, analyzing students, and analyzing material to identify the main concepts that will be used in the test instrument to be developed. Furthermore, at the design stage, the researcher designs a grid of questions on the test instrument, then writes questions based on the grid and compiles scoring guidelines. This is in line with research conducted Nursalam et al. (2017) and Suhady et al. (2020) that after the analysis stage is carried out, then designing test instruments which include compiling test grids, test questions, answer sheets, and assessment guidelines. The development research conducted by Dwidelia (2021), described the same design and development process but differed in the selection of courses and the context of the questions, namely the linear algebra course with essay questions in the form of stories filled with Islamic characters. Whereas in this study the test instruments were in the form of number theory course description questions to measure critical thinking skills. The results of the development of the test instruments were consulted with the supervisor to produce prototype I.

Next, the prototyping stage includes assessments from expert reviews, one-to-one, and small groups. The results of the prototype I was given to the validator as an expert review to be assessed. Along with that, the prototype I was also tested on three semester 4 mathematics education students. The results of the assessment from the validator and student comments in the one-to-one stage were used to revise prototype I which resulted in prototype II. Then, prototype II was tested on six semester 4 mathematics education students. The results from the small group stage were used to revise prototype II and produce prototype III which was then tested at the field test stage. the characteristics of the test subjects at the one-to-one and small group stages are relevant to the research conducted by Agustina and Adesti (2019), one to one evaluation was carried out with three students selected representing high, medium, and low cognitive abilities, small group evaluation carried out with 9 students who were selected based on abilities that represent the character of students who will be the object of research. The prototyping process applied to this study is inseparable from three processes namely validation, evaluation, and revision as in the research conducted by Putra et al. (2017) that at the product prototype stage, the evaluation passed was expert review, one to one in parallel, as well as small groups.

The quality of the HOTS-type test instrument for number theory courses to measure students' critical thinking skills can be seen based on predetermined criteria. The results of the validation of the test content obtained a CVR value for each item, namely 1, and a CVI value of 1. This means that the items on the test instrument can be used to measure students' critical thinking skills. Furthermore, student responses to field trials obtained an average score of 85.03% with a very positive practicality category. Analysis of the support lecturer's response was 90%. The presentation is included in the interpretation criteria of $85\% \leq x < 100\%$ with the "very positive" attractiveness category. This shows that the test instrument meets practical criteria. The results of the reliability analysis of the HOTS-type test instrument for a number theory course to measure students' critical thinking skills using SPSS Statistics 22 show that the test instrument is classified as reliable with an average Cronbach's alpha value of 0.798 with high interpretation. Results analysis of the level of the HOTS type test instrument in number theory courses to measure students' critical thinking skills obtained an average of 0.48 in the medium category. This means that the question is neither easy nor difficult. The results of the analysis of differentiating power for each item of the test instrument that was developed obtained an average result of discriminating power of 0.35 in a good category so that the HOTS type test instrument for number theory courses to measure students' critical thinking skills has good discriminating quality.

Based on the test results, when viewed from the cognitive aspect, it shows that the presentation of the achievement of critical thinking skills in mathematics education students A semester 2 at UIN Alauddin Makassar is in the sufficient category with an average score of 47.67%. It was found that the acquisition of scores of students who were in the category of critical thinking was less due to factors of unpreparedness, lack of focus, and lack of practice in applying the steps to solving HOTS-type questions which began with making known and being asked until answered. This is in line with the research conducted by Saraswati (2019) with research results showing students' HOTS thinking skills are in the sufficient category, and students' abilities are found to be in a low category due to the inability of students to make step designs problem-solving steps to answer HOTS questions.

REFERENCES

Afriansyah, E. A., Herman, T., Turmudi, T., & Dahlan, J. A. (2020). Mendesain Soal Berbasis Masalah untuk Kemampuan Berpikir Kritis Matematis Calon Guru. *Mosharafa: Jurnal Pendidikan Matematika*, *9*(2), 239–250. https://doi.org/10.31980/mosharafa.v9i2.649

Agustina, N., & Adesti, A. (2019). Pengembangan Modul Mata Kuliah Strategi Belajar dan Pembelajaran Pada FKIP-Universitas Baturaja. *Syntax Literate: Jurnal Ilmiah Indonesia*, *4*(9). https://doi.org/10.36418/syntax-literate.v4i9.713

Ahmad, N. (2015). *Buku Ajar Evaluasi Pembelajaran*. Interpena.

Andriyani, R., & Yenni. (2019). Analisis High Order Thinking Skills (HOTS) Mahasiswa Pada Mata Kuliah Trigonometri. *SNP2M*, *1*(2), 175–179. https://doi.org/https://doi.org/10.31539/spej.v1i2.268

Arifah, F., & Yustisianisa. (2012). *Evaluasi Pembelajaran*. Mentari Pustaka.

Arifin, Z. (2016). Pengembangan Instrumen Pengukur Berpikir Kritis Matematika Siswa Sma Kelas X. *Jurnal THEOREMS (The Original Research of Mathematics)*, *1*(1), 58–74. https://media.neliti.com/media/publications/301751-pengembangan-instrumen-pengukur-berpikir-71a6c8eb.pdf

As'ari, A. R., Ali, M., Basri, H., Kurniati, D., & Maharani, S. (2019). *Mengembangkan HOTS Melalui Matematika*. Universitas Negeri Malang.

Bakhri, S., & Rosnawati, R. (2018). Developing an Assessment Instrument of Junior High School Students' Higher Order Thinking Skills in Mathematics. *5th ICRIEMS Proceedings*, 311–318. http://seminar.uny.ac.id/icriems/sites/seminar.uny.ac.id.icriems/files/proceeding 2018/ME43_Syaifulloh Bakhri.pdf

Bashooir, K., & Supahar. (2018). Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.19590

Cahyanti, A. D., Farida, F., & Rakhmawati, R. (2019). Pengembangan Alat Evaluasi Berupa Tes Online/Offline Matematika dengan Ispring Suite 8. *Indonesian Journal of Science and Mathematics Education*, *2*(3), 363–371. https://doi.org/10.24042/ijsme.v2i3.4362

Dwidelia, N. (2021). *Pengembangan Instrumen Tes Mata Kuliah Aljabar Linear Elementer Berbasis HOTS Bermuatan Karakter Islami*. 9–25.

Hadi, W., & Faradillah, A. (2019). The Algebraic Thinking Process in Solving Hots Questions Reviewed from Student Achievement Motivation. *Al-Jabar: Jurnal Pendidikan Matematika*, *10*(2), 327–337. https://doi.org/10.24042/ajpm.v10i2.5331

Hendryadi, H. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen Dan Bisnis (JRMB) Fakultas Ekonomi UNIAT*, *2*(2), 169–178. https://doi.org/10.36226/jrmb.v2i2.47

Hervanda, Y., Fajriah, N., & Suryaningsih, Y. (2020). Soal Model PISA dengan Konteks Etnomatematika untuk Mengukur Kemampuan Pemecahan Masalah Siswa. *Theta: Jurnal Pendidikan Matematika*, *2*(2), 76–82. https://repo-dosen.ulm.ac.id//handle/123456789/24867

Jannatasari, N. Z., Trapsilasiwi, D., & Kurniati, D. (2017). Pengembangan Soal Terbuka (Open-Ended Problem) pada Pembelajaran Kolaboratif untuk Mengukur Kemampuan Berfikir Tingkat Tinggi Siswa SMP Kelas VIII. *Kadikma*, *8*(3), 123–133. https://doi.org/10.19184/kdma.v8i3.6964

Jiwandono, N. R. (2019). Kemampuan Berpikir Kritis (Critical Thinking) Mahasiswa Semester 4 (Empat) Pada Mata Kuliah Psikolinguistik. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, *4*(1). https://doi.org/10.33752/ed-humanistics.v4i1.351

Jurnaidi, J., & Zulkardi, Z. (2013). Pengembangan Soal Model Pisa Pada Konten Change and Relationship Untuk Mengetahui Kemampuan Penalaran Matematis Siswa Sekolah Menengah Pertama. *Jurnal Pendidikan Matematika*, *8*(1). https://doi.org/10.22342/jpm.8.1.1860.25-42

Kamarullah, K. (2017). Pendidikan Matematika Di Sekolah Kita. *Al Khawarizmi: Jurnal Pendidikan Dan Pembelajaran Matematika*, *1*(1), 21. https://doi.org/10.22373/jppm.v1i1.1729

Kenedi, A. K. (2018). Desain Instrument Higher Order Thingking Pada Mata Kuliah Dasar-Dasar Matematika Di Jurusan PGSD. *AR-RIAYAH : Jurnal Pendidikan Dasar*, *2*(1), 67. https://doi.org/10.29240/jpd.v2i1.440

Khairiyah, U. (2019). Respon Siswa Terhadap Media Dakon Matika Materi KPK dan FPB pada Siswa Kelas IV di SD/MI Lamongan. *AL-MURABBI: Jurnal Studi Kependidikan Dan Keislaman*, *5*(2), 190–204. https://doi.org/10.53627/jam.v5i2.3476

Makmur, F. (2020). Pengembangan Instrumen Tes Berbantuan Aplikasi Kahoot Pada Mata Pelajaran Matematika Semester I Kelas VIII MTs Negeri Gowa. *Skripsi*.

Mas'ula, N., & Rokhis, T. A. (2020). Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Mahasiswa Pada Pokok Bahasan Kinematika. *SAP (Susunan Artikel Pendidikan)*, *4*(3), 177–185. https://doi.org/10.30998/sap.v4i3.6279

Ningsih, R., & Annajmi. (2020). *Pengembangan Soal Higher Order Thinking Skills (HOTS) pada Materi Sistem Persamaan Linear Tiga variabel (SPLTV) Kelas X SMA*. *30*(1), 212–219. https://doi.org/10.30606/absis.v3i1.492

Nursalam, Angriani, A. D., & Usman, H. (2017). Pengembangan Tes Kemampuan Penalaran Matematis Peserta Didik Madrasah Tsanawiyah di Makassar. *Jurnal Ilmu Tarbiyah Dan Keguruan*, *20*(1), 85–97. https://doi.org/10.24252/lp.2017v20n1a7

Putra, M., Novita, R., & Dazrullisa. (2017). Pengembangan Prototype Pertama LKS Berbasis Tahapan Pemecahan Masalah Polya untuk Meningkatkan Kemampuan Pemecahan Masalah Siswa SMP. *Numeracy*, *4*(2), 178–187. https://doi.org/10.46244/numeracy.v4i2.285

Putra, Y. Y., & Vebrian, R. (2019). Pengembangan Soal Matematika Model PISA Konteks

Kain Cual Bangka Belitung. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *3*(2), 333–340. https://doi.org/10.31004/cendekia.v3i2.114

Sagala, P. N., & Andriani, A. (2019). Development of Higher-Order Thinking Skills (HOTS) Questions of Probability Theory Subject Based on Bloom's Taxonomy. *Journal of Physics: Conference Series*, *1188*(1). https://doi.org/10.1088/1742-6596/1188/1/012025

Salmina, M., & Adyansyah, F. (2017). Analisis Kualitas Soal Ujian Metematika Semester Genap Kelas XI SMA Inshafuddin Kota Banda Aceh. *Numeracy*, *4*(1), 37–47. https://doi.org/10.46244/numeracy.v4i1.250

Saraswati, P. M. S., & Agustika, G. N. S. (2020). Kemampuan Berpikir Tingkat Tinggi Dalam Menyelesaikan Soal HOTS Mata Pelajaran Matematika. *Jurnal Ilmiah Sekolah Dasar*, *4*(2), 257–269. https://doi.org/10.23887/jisd.v4i2.25336

Sugiyono. (2016). *Metode Penelitian Pendidikan: Pendekatan Kualitatif, Kuantitatif, dan R&D*. Alfabeta.

Suhady, W., Roza, Y., & Maimunah. (2020). Pengembangan Soal untuk Mengukur Higher Order Thinking Skill (HOTS) Siswa. *Jurnal Gantang*, *2*, 143–150. https://doi.org/10.31629/jg.v5i2.2518

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. https://files.eric.ed.gov/fulltext/EJ1109390.pdf

Thoha, M. C. (2003). *Teknik Evaluasi Pendidikan*. RajaGrafindo Persada.

Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student asessment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, *7*(1), 57–68. https://doi.org/10.37058/jp3m.v7i1.2069