Kreano 12 (1) (2021) : 26-38
KREANO
Jurnal Matematika Kreatif-Inovatif
http://journal.unnes.ac.id/nju/index.php/kreano

Development of Measurement Tool for Understanding, Application, and Reasoning Mathematics of Madrasah Ibtidaiyah Students

Ilyas Ismail[1], Nursalam[2], Andi Dian Angriani[3], and Andi Kusumayanti[4]

[1,2,3,4]Jurusan Pendidikan Matematika Fakultas Tarbiyah dan Keguruan
Universitas Islam Negeri Alauddin Makassar
Corresponding Author: ilyas.ismail@uin.alauddin.ac.id[1]
nursalam_ftk@uin-alauddin.ac.id[2], dian.angriani@uin-alauddin.ac.id[3], andi.kusumayanti@uin-alauddin.ac.id[4]

History Article
Received: November, 2020 *Accepted: April, 2021* *Published: June, 2021*

Abstract

This study aims to develop a measurement tool for students' understanding, application, and mathematical reasoning ability at Madrasah Ibtidaiyah in South Sulawesi. This research is a formative research type of Research and Development (R&D) research through 4 (four) stages, namely: the preliminary stage; the stage of self evaluation (curriculum analysis, material, students) and design; prototyping stage (validation, evaluation, and revision) which includes expert reviews, one-to-one, and small groups; and the field test stage. This research's product was an assessment instrument in the form of questions that measured the ability of understanding, application, and mathematical reasoning of Madrasah Ibtidaiyah students. The test subjects in this study were fourth grade MI students in Makassar City, Gowa regency, and Sinjai regency. The content validity results based on the rater indicated that the test items needed were declared valid after being assessed by the rater and analyzed using the CVR (Content Validity Ratio) and CVI (Content Validity Index). In the reliability test, the measurement tools for comprehension, application, and reasoning were indicated to be reliable.

Abstrak

Penelitian ini bertujuan untuk mengembangkan alat ukur kemampuan pemahaman, aplikasi, dan penalaran matematika siswa pada Madrasah Ibtidaiyah di Sulawesi Selatan. Penelitian ini merupakan jenis penelitian Research and Development (R & D) tipe formative research melalui 4 (empat) tahap yaitu: tahap preliminary; tahap self evaluation (analisis kurikulum, materi, siswa) dan design; tahap prototyping (validasi, evaluasi dan revisi) yang meliputi expert reviews, one-to-one dan small group; dan tahap feld test (uji coba lapangan). Produk penelitian ini adalah instrument penilaian berupa soal-soal yang akan mengukur kemampuan pemahaman, aplikasi, dan penalaran matematika siswa Madrasah Ibtidaiyah. Subjek uji coba pada penelitian ini adalah siswa kelas IV MI di Kota Makassar, Kab. Gowa, dan Kab. Sinjai. Hasil validitas isi (Content validity) berdasarkan rater menunjukkan bahwa butir-butir tes yang dibut dinyatakan valid setelah dinilai oleh rater dan dianalisis dengan menggunakan CVR (Content Validity Rasio) dan CVI (Content Validity Indeks). Pada uji reliabilitas, alat ukur kemampuan pemahaman, aplikasi, dan penalaran dikatakan reliabel.

***Keywords**: Measurement Tool; Mathematics; Comprehension; Application; Reasoning.*

p-ISSN: 2086-2334; e-ISSN: 2442-4218

BACKGROUND

Mathematics is one branch of science that has a very important role in science and technology. Mathematics is a subject that can train students in improving critical, logical, and creative thinking. Indonesia's education curriculum places mathematics as a compulsory subject that must be given to students from basic education to tertiary education. The development of students' critical, logical, and creative thinking skills in mathematics can be done by accustoming students to solving mathematical problems of understanding, applying concepts, and reasoning.

The ability to understand mathematics is essential in the principles of learning mathematics (NCTM, 2000: 35). Students' understanding ability is important because everything related to learning will require understanding and meaning of the material (Hikmah, 2017) . Understanding ability is a measure of a person's ability both in quality and quantity of new ideas with the previous ideas. A student is expected to understand the mathematical concepts he/she learned in his/her own way or through the relationship between new concepts with the student's concepts. The students are claimed to have the ability to understand mathematics if they can recognize, understand and apply concepts, principles, procedures, facts, and mathematical ideas (Maulana, 2011). The ability of mathematical understanding will occur if the teacher facilitates students by giving mathematical problems that lead to linking concepts in mathematics. The student's ability to apply mathematical concepts in problem solving is one form of the ability to apply mathematical concepts themselves.

Likewise, the student's ability to apply mathematical concepts in problem solving is one form of the ability to apply mathematical concepts themselves. At the same time, the ability to solve problems is also one of the abilities that the students must possess because many problems in mathematics involve several aspects of knowledge. This ability is also related to students' mathematical reasoning abilities. There has been much research related to the student's abilities to improve the understanding, application, and mathematical reasoning, including higher order thinking skills or HOTS. It recommends the importance of students being trained to think critically in mathematics learning, including solving mathematical problems using mathematical reasoning abilities (Tjalla, 2010 dan Anisah, Zuslkardi, & Darmawijoyo, 2011).

Teachers have made various attempts to foster student enthusiasm in learning mathematics, including learning mathematics with an open-ended approach to developing critical thinking skills (Novikasari, 2009). The use of manipulative objects to improve the understanding of the geometrical concepts of 5th grade elementary school students (Yeni, 2011), Math games using games (Bakker et al., 2015; Pareto, 2014). Jablansky, Alexander, Dumas, & Compton (2016) researched the differences in mathematical reasoning development between students in primary and secondary schools. Moss, Hawes, Naqvi, & Caswell (2015) researched the impact of applying lesson study in Japan to improve students' reasoning abilities for geometry material.

However, the various efforts that have been made have not provided maximum results because the acquisition of Indonesian students' cognitive mathematical abilities is still weak. It can be proved based on the survey finding of *Trend in International Mathematics and Science Study* (TIMSS) 2015 showed that Indonesian students are weak in all aspects, both content and cognitive mathematics and science for grade 4 (Yeni,

2011). It shows that students' cognitive mathematical abilities, especially in understanding, applying concepts, and reasoning abilities, still give weak results. This weakness is the consequence of the students are not accustomed to and are not trained to solve problems that require them to develop the ability to understand, apply concepts and reasoning abilities. It is in line with research by Nuraeni, Mulyati, dan Maya; (Yuliani et al., 2018), who said that the students' understanding, concept application, and reasoning abilities were still low. It can be proven when students forget about the studied material even though the material is related, and students have not been able to apply the concept of learning into their lives. Nursalam, Angriani, & Usman (2017) suggested that the mathematical reasoning ability of Madrasah Tsanawiyah students was in the lacking category, so the teacher needed to train even more students by using mathematical reasoning questions. Students who are given problems related to inquiry, experiment, compare, conclude, reason, stimulate and find many difficulties in making decisions, problem solving, error analysis, and classification. (Heong et al.,2011; Ramos, Dolipas, & Villamor, 2013; dan Saido, Siraj, Bin Nordin, & Al Amedy, 2015). Their results recommend the importance of developing students' higher-order thinking skills, especially in terms of mathematical reasoning. Other studies from Nursalam (2016) stated that students in grade V Elementary School/MI had difficulty learning the material on the rank and fractions operation caused by weakness in arithmetic operations.

Meanwhile, based on the results of observations made at several Madrasah Ibtidaiyah schools in South Sulawesi, it is known that the students' ability to understand, apply concepts, and mathematical reasoning abilities are still low. The low ability of students' mathematical reasoning can be seen in the indicator of the ability to propose initial assumptions of the value obtained by students of 45% of students, the ability to compile evidence of the correctness of the solution by 40%, the ability to draw conclusions by 35%. Meanwhile, the low comprehension ability with indicators of applying the concept obtained a value of 25%. Based on these observations' results, the researcher thinks that one of the factors causing the low ability of reasoning, understanding, and application of students' mathematical concepts is because the teacher still emphasizes the skill of working on questions (drill). So that it does not provide opportunities for students to build their knowledge. It is in line with the results of a survey study that shows around 12.18% of teachers in Indonesia do not understand the curriculum. Despite the curriculum's implementation, almost everything depends on the teacher's creativity and persistence because they know the field's situation and conditions. Teachers should select and carry out the learning and teaching process according to students 'abilities, especially those related to students' mathematical cognitive.

National Council of Teacher of Mathematics (NCTM, 2000) states that in the implementation of mathematics learning, there are 5 (five) things that must be a part of the attentions of an educator, namely connections, reasoning, communication, problem solving, and representation. Hence, it shows that an educator plays an essential role in growing students' mathematical reasoning in learning activities or evaluation activities, especially in making math problems.

In the learning process, students need to be trained to develop the ability of understanding, concepts application, and mathematical reasoning so that the learning process should be given real problems to be trained and solved by students. It is important for students who are not used to be-

ing given a level of understanding, application, and reasoning.

Therefore it is important to develop an instrument to measure students' understanding, application, and reasoning abilities at the basic education level, especially at the Madrasah Ibtidaiayah (MI) in South Sulawesi.

METHODOLOGY

This study is a research and development (R&D) formative research type (Tessmer, 1993). This research involved 4 (four) stages, namely, the preliminary stage, the self-evaluation stage (curri-culum analysis, material, students) and design, the prototyping stage (validation, evaluation, and revision), which included expert reviews, one-to-one and small groups, and the field test stage. The product developed in this study was a measurement tool for students' under-standing, application, and mathematical reasoning ability at Madrasah Ibtidaiyah (MI) in South Sulawesi. The test subjects in this study were grade IV MI Al-Abrar Makassar students, MI Madani Pao-Pao, and MIN 2 Sinjai. The data collection instruments used in this study were: (1) validation sheets which were used to measure the validity of students' mathematical understanding, application, and reasoning abilities and other research instruments, (2) student questionnaire sheets which were used to measure the practicality of teaching materials (3) Test students' understanding, application, and mathematical reasoning abilities. The design of the development is presented in Figure 1.

Each research method certainly has shortcomings, especially in this development research. The R & D approach can be said to be "here and now" research. R & D research cannot be fully generalized because basically R & D research is modeling the sample, not the population. Meanwhile, in this study, based on the development stages passed, there are several drawbacks, including this development method that requires a lot of money and takes a long time.

RESULT AND DISCUSSION

The process of developing measurement devices for students' understanding, application, and mathematical reasoning through 4 stages are broken down as follows:

The Preliminary Stage

This stage is begun with gathering several references related to this research, namely research on the development of measuring tools for students' understanding, application, and reasoning skills, also known as higher order thinking skills. Some

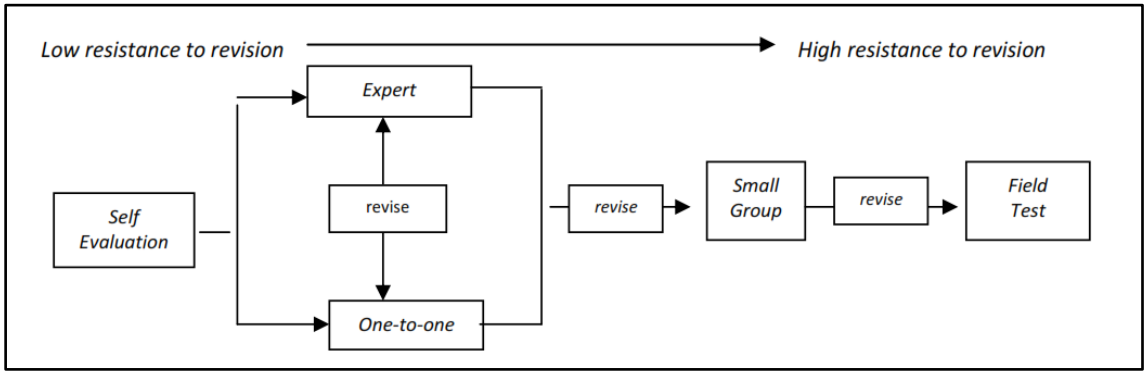

Figure 1. Tessmer Design Development (1999)

references on higher order thinking skills Susan M Brookhart (2010), who said that higher order thinking skills transfer knowledge, critical thinking skills, and problem solving. In Bloom's Taxonomic Revision, there are levels in the cognitive domain, namely: remembering, understanding, applying, analyzing, evaluating, and creating. Anita Lie et al. (2020) said that higher order thinking skills are the ability to use and cultivate thought processes above facts. Ridwan Abdullah Sani (2013)said that higher order thinking skills include critical, logical, reflective, metacognitive, reasoning, and creative thinking skills. Based on these opinions, it can be concluded that higher order thinking skills are a fact-based thinking process consisting of several indicators of remembering, understand, apply, analyze, and evaluate, and create, think critically, logically, reflectively, metacognitively, reasoning, and creatively. Meanwhile, in this study, the part of higher-order thinking skills that will be studied is the ability to understanding, applying, and reason.

The next step was to determine the location and subjects of research trials, namely Grade IV MI students in South Sulawesi consisting of 3 schools, namely MI Al Abrar Makassar, MI Madani Pao-Pao, and MIN 2 Sinjai. After determining the location and the trial subjects, the observation, which aimed at identifying learning activities and students' understanding, application, and reasoning abilities at the school was carried out. The method used in this observation was the interview method. Interviews were conducted with mathematics teachers at the three schools. The results of the interviews conducted showed that the teacher's creativity in choosing the question stimulus was not following the situation and condition of the area around the education unit. The questions made by the teacher were not much different from the examples given. It could be proven by the general questions created by the teacher is in the form of stuffing. The teacher does not determine the form of the stimulus and determines the description of the stimulus. The consideration that not all students will be able to answer questions in the HOTS form, the busy schedule makes the teacher not have enough time to develop HOTS-based questions. Extensive knowledge is needed to develop HOTS questions.

The Self Evaluation Stage

This stage aims to design a measuring instrument of comprehension, application, and reasoning abilities based on the preliminary stage results. The measuring instrument to be designed consists of a grid, test questions in multiple choice and gap filling, and assessment guidelines. It is in line with the opinion of Mardapi (2008), which states that in developing a test instrument, what needs to be done is compiling a test grid, choosing a test form, and making assessment guidelines. Several designs need to be done, including curriculum analysis, material analysis, student analysis, and design. The curriculum that has been reviewed at the curriculum analysis stage is the MI mathematics curriculum, which in the 2013 curriculum aims to improve understanding, application, and reasoning abilities. Student analysis activities were focused on grade IV students as test subjects. Based on observations and results of interviews from mathematics teachers, it can be seen that the mathematics knowledge of students in grade IV MIN 2 Sinjai, MI Al-Abrar Makassar, and MI Madani Pao-Pao varied. There were less, medium, and high ability. In general, students' understanding, application, and reasoning abilities have never been explored either by

the teacher or other researchers. The students also rarely get questions that can engage their understanding, application, and reasoning abilities. Material analysis is an identifying activity of the main concepts that will be used in tests on MI mathematics material. Regarding the curriculum analysis activities, it was found that the material to be used in the development of measuring instruments was following the material in the Curriculum 2013 for mathematics subjects. The materials are number, geometry, and data presentation.

After carrying out those three analyses, namely curriculum analysis, students, and material, the next stage is designing or designing measurement tools for comprehension, application, and reasoning, which include: test grids, test questions, and assessment guidelines. The researcher's initial stage was designing the test questions for comprehension, application, and reasoning abilities. The questions were designed based on the material analyzed and based on indicators of understanding ability, application, and reasoning called prototype I. Researchers designed 40 measuring instrument questions, such as 30 multiple choice questions and 10 numbers of filling gap questions. The test questions designed were problems related to daily life and bring up pictures relating to the students' daily lives. Verbal problems in mathematics are a type of mathematical problem designed to help students apply abstract mathematical concepts to real-world situations (Tasni & Susanti, 2017).

The researcher also made a test grid and consideration material for the validator to check the validity of the comprehension matter, application, and reasoning. The test grid was designed referring to the indicators of achievement and cognitive domain of each question. Besides, researchers also designed assessment guidelines to facilitate researchers, teachers, or other researchers in providing the test results assessment that the students had done. In learning evaluation, scoring guidelines are needed that can be used as a guide to assess student work (Charlotte Danielson, 2007). Scoring guidelines are required for both multiple-choice and essay form tests. Both forms of tests require clear guidelines on what and how the assessment is carried out (Sumaryanta, 2015). What is done at this stage can then be used as a reference in designing test instruments related to HOTS questions, especially for mathematics teachers in improving student mathematics learning outcomes in cognitive aspects. Moreover, so far, HOTS question-making training is rarely held for math teachers. It reinforces the opinion of Karim et al. (2019), who said that training specifically for making HOTS-oriented math problems is lacking, even there are still teachers who have never attended HOTS question training.

The Prototyping (Validation, Evaluation, dan Revision) Stage

The purpose of this prototyping phase is to produce prototyping II of the revised measuring instrument based on input from experts (reviewers) consisting of 6 lecturers of the Department of Mathematics Education at the Faculty of Tarbiyah and Teacher Training at the Alauddin Islamic State University Makassar in charge of checking and correcting the measuring instruments understanding, application, and reasoning abilities. In this validation stage, the validator assessed aspects related to the instrument that had been designed (Prototype I). Each aspect had a maximum value of 1 and a minimum of 0.1 means that the respondent considered the questions in accordance with the topic to be measured but needed improvement. 0 means that the respondent considers the problem was not

in accordance with the topic to be measured and needed improvement. Referring to the first validation measuring instrument results, the ability of understanding, application, and reasoning shows that there were 10 items to be revised, namely question No. 2, 4, 9, 11, 12, 17, 20, 25, 30, 35. After the items that were not well revised according to the expert team's input and rearranged the problem's composition, the content validation was re-done to the experts who previously given an assessment. Thus, the second validation results show that 40 items examined by 6 validators (experts) had shown that these items support the test's validity. Then from the CVR results, the CVI (Content Validity Index) value which is the average CVI of all items, is 1 meaning "very appropriate" questions with the topic to be analyzed.

After all the research instruments were considered valid, the next step was to conduct a one-to-one test, namely testing on a small scale to determine the questions' readability level by taking 3 fourth grade students of MI Madani Pao-Pao. These students are 3 non-research subjects on the same level, each with high, medium, and low abilities based on the learning outcomes known by the mathematics teacher who taught those three students. According to the findings of the percentage of the 8 statement items above with predetermined criteria and based on the results of the analysis on the student questionnaire responses to the measuring instrument of comprehension, application, and reasoning in one-to-one trials, the average positive response of students was 79.16%, and the average negative response of students was 20.83%. The analysis results in this one-to-one test will serve as a reference for revising the students' mathematical understanding, application, and mathematical reasoning abilities.

Measuring instruments that had been revised in the previous stage were retested, namely the small group trial, which was testing on a small scale to determine the level of readability of the questions by taking 6 fourth grade students of MI Madani Pao-Pao. These students are 6 non-research subjects on the same level, each with 2 high, medium, and low ability students based on the learning outcomes known by the mathematics teacher who taught those three students. In accordance with the results of the percentage of the 8 statement items above with predetermined criteria and the results of the analysis on the student questionnaire responses to the measuring instrument of comprehension, application, and reasoning in the one-to-one trial, the average positive student response was 79.16%, and the average negative response of students was 20.83%. The analysis results in this one-to-one test will serve as a reference for revising the students' mathematical understanding, application, and mathematical reasoning abilities. In small group trials, students' average positive response was 79.16%, and the average negative response of students was 20.83%. Thus, students' average positive response was 79.16%, while the average negative response questionnaire of students' responses was 20.83%. It can be concluded that more than 50% of students gave positive responses. Moreover, the students' response questionnaire has met the "achieved" criteria, and there is no improvement/ revision of the measuring instrument to be developed.

The most important stage in this research was to do the field test to see the level of reliability and difficulty of measuring tools for students' understanding, application, and mathematical reasoning abilities. Field tests were conducted at 3 MI schools in South Sulawesi in class IV, consisting of 29 students of MI Al Abrar Makassar, 39 stu-

dents of MI Madani Pao-Pao, and 16 students of MIN 2 Sinjai. Each student must complete 30 multiple choice questions and 10 filling gap question items in 2 x 60 minutes. Based on the results of the student's work, the level of test reliability can be calculated. The following data is the results of the calculation of the test reliability test shown in the following Table 1.

Table 1. Reliability Test Result

Reliability Statistics	
Cronbach's Alpha	N of Items
.731	30

Regarding table 1, it appears that the reliability value of items that have been made and tested amounted to 0.71. Therefore, it showed that the analysis of the reliability level of the measurement tools for comprehension, application, and reasoning for multiple choice questions and the tested contents could be categorized as reliable.

In addition to the reliability test, the researcher also tests the difficulty level and distinguishing features of measuring devices for comprehension, application, and reasoning. Points of measuring instruments can be good if the test items have a difficulty level at intervals of 0.30-0.70. It was found that the items were not too difficult and not too easy. The difficulty level of measuring instruments developed was also obtained from data on student work results in trials (field tests). The difficulty level analysis results of the ability of understanding, application, and reasoning can be seen in appendix A.

To determine the decision that an easy, medium, and hard problem can be used to categorize that a good level of difficulty was between 0.3 - 0.7. In addition to that, the problem needed to be revised or rejected. In this case, a quality distribution of questions can be presented, as shown in Figure 2.

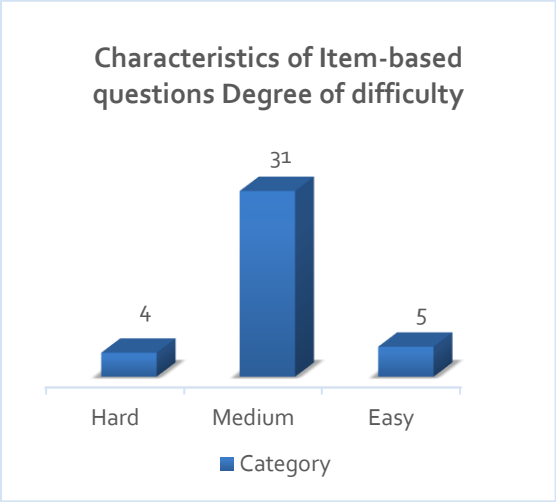

Figure 2. Diagram of Difficulty Levels of Multiple Choice Questions

As what is projected on the table, it can be concluded that the results of the analysis of the difficulty level measuring instrument of understanding, application, and mathematical reasoning of students in MI obtained 4 items of hard category, 31 items of the medium category, and 5 items of the easy category. Based on this empirical data, the 9 items needed to be reviewed or revised.

In addition to the difficulty level of the items, another aspect used to determine the items' quality is the power of differentiation. The item analysis results to determine the distinguishing power of the questions are presented in **appendix B**. Meanwhile, the distribution of the analysis results of the differences in power can be shown in Figure 3 below.

The distribution of different power analysis results can be presented in Figure 3.

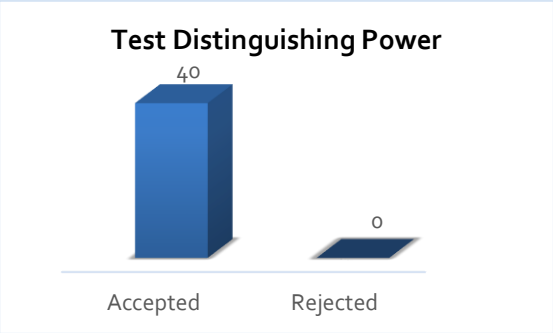

Figure 3. Diagram of Distribution of Difficulty Degree for Multiple Choice Items

Based on the table, it can be concluded that the analysis results of the distinguishing power of measurement tools for students' understanding, application, and mathematical reasoning in MI with the criterion that if the differentiating power was greater than or equal to 0.3. Then the items that were accepted will be revised if the distinguishing power was between 0.2 - 0.29. Whereas if the distinguishing power were below 0.2, then the item would be rejected. Following the above analysis, it was obtained that 40 items were accepted and 0 items were rejected. Hence, from this empirical data, the 3 question items needed to be reviewed or revised.

Finally, the last stage was to analyze the level of students' understanding, application, and mathematical reasoning skills by calculating each student's score after answering the questions given. The analysis results of students' understanding, application, and reasoning tests are projected in the Table 4.

Regarding the data analysis of measuring the ability of understanding, application, and reasoning of MI-Madani Pao-Pao students, it is known that from 84 students subjected to measuring instruments, from the results of the trials there were no students who had very good levels of comprehension, application, and reasoning abilities, there were 28 students (33.33%) included in the category of having good comprehension, application, and reasoning categories, 35 students (41.67%) included in the category of having sufficient (medium) comprehension, application, and reasoning abilities. Meanwhile, 19 students (22.62 %) were included in the category of having the low ability of understanding, application, and reasoning, and 2 students (2.38%) were included in the category of having very low ability of understanding, application, and category reasoning.

The low ability of understanding, application, and reasoning is due to several factors including, formal education that is taking place now tends to get stuck just struggling with aspects of remembering, and understanding which is a low order of thinking. This is in line with Ardhana's (2017) opinion, who concluded that students could only work on questions at the remembering stage. It is presumably because students are asked to swallow only the things that the teacher delivers. Learning activities with a pouring system can cause the control of children's potential, even though each child is born with extraordinary potential. Students'

Table 4. The Analysis of IV Grader MI Field Test Results

Total Questions	Students Score	Frequency	Percentage (%)	Category
40 questions	80< *score* ≤100	0	0	Very Good
	60< *score* ≤80	28	33,33	Good
	40< *score* ≤60	35	41,67	Medium
	20< *score* ≤40	19	22,62	Low
	0≤ *score* ≤20	2	2,38	Very Low
	Total of Subjects	84	100	
	Mean Score		**50,71**	Medium

difficulty in understanding abstract concepts with learning methods dominated by the teacher, as in general students are often successful in solving certain problems but fails if the problem's context is modified. It is one of the reasons students are not accustomed to thinking at a high level, namely using comprehension, application, and reasoning abilities.

In the implementation of this development research, the researcher has several limitations. Some of them are mathematics teachers who do not participate in developing this problem, there are still many students who are not serious in working on the questions, and HOTS questions cannot be followed evenly by all students who basically each of them have different abilities, students are not used to dealing with HOTS questions.

CLOSING

The process of developing an instrument for testing the ability of understanding, application, and mathematical reasoning of Grade IV Madrasah Ibtidaiyah students through 4 stages, namely: (a) preliminary stage is the initial or preliminary stage of the development process. At this stage the researcher seeks references about test instruments to measure students' understanding, application, and mathematical reasoning abilities and determine the place of research trials; (b) the self-evaluation stage is this stage in which the researcher designed the initial instrument that was developed based on the results of the curriculum 2013 analysis, student analysis, MI grade IV material analysis and also indicators of comprehension ability, application, and mathematical reasoning; (c) the prototyping stage, which is to test the validity of the test instruments to 6 validators who are lecturers in the Faculty of Mathematics Education in the Faculty of Tarbiyah, and teacher training and one-to-one trials to 3 students to be asked for comments about the test questions; (d) field test stage, namely field trials in 3 MI schools in South Sulawesi in grade IV namely MI Madani Pao-Pao, MIN 2 Sinjai, MI Al-Abrar Makassar to determine the level of comprehension, application, and reasoning ability, reliability, the level of difficulty of the test that was created. In terms of the analysis results obtained, an average ability of understanding, application, and reasoning of MI students was 50.71 with enough categories. Thus, based on the analysis results, it can be concluded that the teacher still needs further efforts to provide questions that can stimulate students' understanding, application, and reasoning abilities.

REFERENCES

Anisah, A., Zulkardi, Z., & Darmawijoyo, D. (2011). Pengembangan soal matematika model PISA pada konten quantity untuk mengukur kemampuan penalaran matematis siswa sekolah menengah pertama. *Jurnal Pendidikan Matematika*, *5*(1).

Bakker, M., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2015). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability. *Contemporary Educational Psychology*, *40*, 55–71.

Brookhart, S. M. (2010). Asses Higher-order Thinking Skills in Your Classroom. In *ASCD*.

Danielson, C. (2007). Enhancing Professional Practice: A Framework for Teaching (2nd Edition). *Adolescence*.

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. Bin, & 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.

Hikmah, R. (2017). Penerapan Model Advance Organizer untuk Meningkatkan Kemampuan Pemahaman Siswa. *SAP (Susunan Artikel Pendidikan)*, *1*(3), 271–280.

Jablansky, S., Alexander, P. A., Dumas, D., & Compton, V. (2016). Developmental differences

in relational reasoning among primary and secondary school students. *Journal of Educational Psychology*, *108*(4), 592.

Karim, K., Hidayanto, T., Kamaliyah, K., & Arrasyid, M. F. (2019). Bimbingan Pembuatan Soal Berorientasi HOTS Bagi Guru Peserta MGMP Matematika SMP Kabupaten Banjar. *Bubungan Tinggi: Jurnal Pengabdian Masyarakat*, *1*(2), 100.

Lie, A., Tamah, S. M., Gozali, I., Triwidayati, K. R., Utami, T. S. D., & Jemadi, F. (2020). Secondary School Language Teachers' Online Learning Engagement During The Covid-19 Pandemic In Indonesia. *Journal of Information Technology Education: Research*, *19*, 803-832.

Mardapi, D. (2008). Teknik penyusunan instrumen tes dan nontes. *Yogyakarta: Mitra Cendekia*.

Mathematics, N.C.T. (2000). Principles and Standards for School Mathematics. *Journal of Equine Veterinary Science*, *18*(11), 719.

Moss, J., Hawes, Z., Naqvi, S., & Caswell, B. (2015). Adapting Japanese Lesson Study to enhance the teaching and learning of geometry and spatial reasoning in early years classrooms: a case study. *Zdm*, *47*(3), 377–390.

Novikasari, I. (2009). Pengembangan Kemampuan Berpikir Kritis Siswa melalui Pembelajaran Matematika Open-ended di Sekolah Dasar. *Insania*, *14*(2), 346–364.

Nursalam, N. (2016). Diagnostik Kesulitan Belajar Matematika: Studi pada Siswa SD/MI di Kota Makassar. *Lentera Pendidikan*, *19*(1), 1–15.

Nursalam, N., Angriani, A. D., & Usman, H. (2017). Pengembangan Tes Kemampuan Penalaran Matematis Peserta Didik Madrasah Tsanawiyah Di Makassar. *Lentera Pendidikan : Jurnal Ilmu Tarbiyah Dan Keguruan*, *20*(1), 85–97.

Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education*, *24*(3), 251–283.

Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher order thinking skills and academic performance in physics of college students: A regression analysis. *International Journal of Innovative Interdisciplinary Research*, *1*(4), 48–60.

Saido, G. M., Siraj, S., Bin Nordin, A. B., & Al Amedy, O. S. (2015). Higher Order Thinking Skills among Secondary School Students in Science Learning. *Malaysian Online Journal of Educational Sciences*, *3*(3), 13–20.

Sani, R. A. (2013). *Inovasi Pembelajaran*. Bumi Aksara.

Sumaryanta. (2015). Pedoman Penskoran. *Indonesian Digital Journal of Mathematics and Education*, *2*, 181–190.

Tasni, N., & Susanti, E. (2017). Membangun Koneksi Matematis Siswa dalam Pemecahan Masalah Verbal. *Beta Jurnal Tadris Matematika*, *10*(1), 103.

Tessmer, M. (1993). Planning and conducting formative evaluations: Improving the quality of education and training. In *Planning and Conducting Formative Evaluations*. Psychology Press.

Tjalla, A. (2010). *Potret Mutu Pendidikan Indonesia Ditinjau dari Hasil-Hasil Studi Internasional*. Jakarta: UT.

Yeni, E. M. (2011). Pemanfaatan Benda-Benda Manipulatif Untuk Meningkatkan Pemahaman Konsep Geometri dan Kemampuan Tilikan Ruang Siswa Kelas V Sekolah Dasar. *Jurnal Edisi Khusus*, *1*, 63–75.

Yuliani, E. N., Zulfah, Z., & Zulhendri, Z. (2018). Kemampuan Pemahaman Konsep Matematis Siswa Kelas Viii Smpn 1 Kuok Melalui Model Pembelajaran Kooperatif Tipe Group Investigation. *Jurnal Cendekia*, *2*(2), 91–100.

Appendix A

Difficulty level analysis is intended to find out if the question is relatively easy or difficult. Difficulty is a number that indicates the difficulty or ease of a problem (Arikunto, 1999). The difficulty level category is relatively difficult to be at the interval of kindergarten <0.30, medium with a difficulty level of 0.30 ≤TK≤0.7, and easy with TK>0.70. Meanwhile, the details of the measuring instrument question can be said to be good if the test items have a difficulty level at intervals of 0.30-0.70 (Allen and Yen, 1979). The following results of the analysis of the difficulty level of the double problem measuring instrument are shown in table A1.

Table A1. Results of Analysis of Difficulties in the Level of Measurement of Students' Understanding, Application and Reasoning For Multiple Choice Questions Mathematics at Madrasah Ibtidaiyah in South Sulawesi

NO	DIFFICULTY INDEX	CATEGORY	NO	DIFFICULTY INDEX	CATEGORY
1	0,43	MEDIUM	16	0,45	MEDIUM
2	0,43	MEDIUM	17	0,45	MEDIUM
3	0,50	MEDIUM	18	0,44	MEDIUM
4	0,26	HARD	19	0,48	MEDIUM
5	0,64	MEDIUM	20	0,48	MEDIUM
6	0,61	MEDIUM	21	0,73	EASY
7	0,38	MEDIUM	22	0,39	MEDIUM
8	0,43	MEDIUM	23	0,44	MEDIUM
9	0,31	MEDIUM	24	0,37	MEDIUM
10	0,51	MEDIUM	25	0,60	MEDIUM
11	0,31	MEDIUM	26	0,45	MEDIUM
12	0,38	MEDIUM	27	0,63	MEDIUM
13	0,56	MEDIUM	28	0,46	MEDIUM
14	0,44	MEDIUM	29	0,43	MEDIUM
15	0,20	HARD	30	0,69	MEDIUM

Based on table A1 it is known that for multiple choice questions, there are 1 question item that is relatively easy, 2 points of questions that are classified as difficult and 17 questions that are classified as moderate.

Table A2. Results of Analysis of Difficulties in the Level of Measurement of Students' Understanding, Application and Reasoning for Mathematics Fill in Questions at Madrasah Ibtidaiyah in South Sulawesi

NO	DIFFICULTY INDEX	CATEGORY	NO	DIFFICULTY INDEX	CATEGORY
1	0,83	EASY	6	0,30	HARD
2	0,79	EASY	7	0,54	MEDIUM
3	0,64	MEDIUM	8	0,64	MEDIUM
4	0,51	MEDIUM	9	0,99	EASY
5	0,25	HARD	10	0,95	EASY

Based on table A1 it is known that for the questions of choice of stuffing, there are 4 items that are classified as moderate, 2 items that are classified as difficult, and 4 items that are classified as easy.

Appendix B

Table B1. Results of the Analysis of Distinguishing Power of Measuring Instruments for Understanding, Application, and Reasoning of Student Mathematics Multiple Choice Questions at Madrasah Ibtidaiyah in South Sulawesi

NO	ITEM DISCRIMIN-ATION	CRITERIA OF ITEM DISCRIMINATION	CRITERIA	NO	ITEM DISCRIM-INATION	CRITERIA OF ITEM DISCRIMINATION	CRITERIA
1	0,24	NOT SATISFACTORY	ACCEPTED	16	0,33	SATISFACTORY	ACCEPTED
2	0,10	NOT VERY SATISFACTORY	ACCEPTED	17	0,12	NOT VERY SATISFACTORY	ACCEPTED
3	0,21	NOT SATISFACTORY	ACCEPTED	18	0,05	NOT VERY SATISFACTORY	ACCEPTED
4	0,31	SATISFACTORY	ACCEPTED	19	0,19	NOT VERY SATISFACTORY	ACCEPTED
5	0,24	NOT SATISFACTORY	ACCEPTED	20	0,26	NOT SATISFACTORY	ACCEPTED
6	0,24	NOT SATISFACTORY	ACCEPTED	21	0,12	NOT VERY SATISFACTORY	ACCEPTED
7	0,02	NOT VERY SATISFACTORY	ACCEPTED	22	0,26	NOT SATISFACTORY	ACCEPTED
8	0,26	NOT SATISFACTORY	ACCEPTED	23	0,21	NOT SATISFACTORY	ACCEPTED
9	0,10	NOT VERY SATISFACTORY	ACCEPTED	24	0,14	NOT VERY SATISFACTORY	ACCEPTED
10	0,29	NOT SATISFACTORY	ACCEPTED	25	0,26	NOT SATISFACTORY	ACCEPTED
11	0,29	NOT SATISFACTORY	ACCEPTED	26	0,14	NOT VERY SATISFACTORY	ACCEPTED
12	0,19	NOT VERY SATISFACTORY	ACCEPTED	27	0,31	SATISFACTORY	ACCEPTED
13	0,33	SATISFACTORY	ACCEPTED	28	0,00	NOT VERY SATISFACTORY	ACCEPTED
14	0,31	SATISFACTORY	ACCEPTED	29	0,24	NOT SATISFACTORY	ACCEPTED
15	-0,02	NOT VERY SATISFACTORY	ACCEPTED	30	0,33	SATISFACTORY	ACCEPTED

Based on the results of the differentiation analysis in table B1, 12 questions have an unsatisfactory discrimination index with accepted criteria, and 12 questions have a very unsatisfactory index with accepted criteria, and 6 questions have a satisfactory discrimination index with accepted criteria.

Table B2. Results of the Analysis of Distinguishing Power of Measuring Instruments for Understanding, Application, and Reasoning of Student Mathematics Multiple Choice Questions at Madrasah Ibtidaiyah in South Sulawesi

NO	ITEM DISCRIMIN-ATION	CRITERIA OF ITEM DISCRIMINATION	CRITERIA	NO	ITEM DISCRIM-INATION	CRITERIA OF ITEM DISCRIMINATION	CRITERIA
1	0,05	NOT VERY SATISFACTORY	ACCEPTED	6	0,19	NOT VERY SATISFACTORY	ACCEPTED
2	0,17	NOT VERY SATISFACTORY	ACCEPTED	7	0,24	NOT SATISFACTORY	ACCEPTED
3	0,26	NOT SATISFACTORY	ACCEPTED	8	0,17	NOT VERY SATISFACTORY	ACCEPTED
4	0,36	SATISFACTORY	ACCEPTED	9	0,02	NOT VERY SATISFACTORY	ACCEPTED
5	0,02	NOT VERY SATISFACTORY	ACCEPTED	10	0,00	NOT VERY SATISFACTORY	ACCEPTED

Based on the results of the discriminating power analysis in table B1, it can be seen that 2 items have an unsatisfactory discrimination index with accepted criteria, and 7 items have a very unsatisfactory index with accepted criteria, and 1 item has a satisfactory discrimination index with accepted criteria.