International Journal of Advanced Science and Technology
Vol. 29, No. 6, (2020), pp. 1483-1492

Test Instrument Development of Mathematical Problem Solving Skills

Siti Annisah[1,4], Zulela[2], Endry Boeriswati[3], Yunita Wildaniati[4], Atin Supriatin[5]
[1,2,3]*State University of Jakarta, Indonesia.*
[4]*State Islamic Institute of Metro Lampung, Indonesia*
[5]*State Islamic Institute of Palangkaraya, Indonesia.*

Abstract

Instruments that can measure mathematical problem-solving skills are very limited. The test instruments used in mathematics learning often measure memorization and knowledge skills. The purpose of this research is to develop a test instrument that can measure the ability to solve mathematical problems in prospective teachers in elementary school. This research is a research and development. Data collection techniques using questionnaires and tests. Data analysis using qualitative descriptive techniques and the Content Validity Ratio (CVR) formula. The results of this research are a test instrument consisting of a five-grain essay that measures math problem-solving skills. The instrument is of good quality and valid according to expert judgment.

Keywords: *Test Instruments, Problem-Solving Skills, Mathematics*

I. INTRODUCTION

Test as one of the instruments that can measure problem-solving capability, it failed to run its function. Many test instruments used in mathematics learning only measure memory skills. Some of the research shows that the problems used in mathematics learning are regular problems and low-level questions (Nur Atikah Khairun Nisa; Rani Widyastuti; dan Abdul Hamid, 2018). Most of the available mathematical tests test memorization skills (Kosko, 2019). Besides, the tests used in the school have not been oriented towards creative thinking and problem-solving skills (Hidayat, Susilaningsih, & Kurniawan, 2018). Furthermore, the results of observations that researchers do that in general mathematics books or materials provide very little examples of problem-solving, and educators give tests or questions on books or materials mathematics Students so that the math problems given in learning more measure low levels of proficiency. The limited test instruments assessment of problem-solving skills due to the creation of the test is not easy work. Problem-solving is one of the high levels of thinking Skills (HOTS) (Abdullah, Abidin, & Ali, 2015; Brookhart, 2010). Problem-solving capabilities involve the ability to analyze, interpret, reasoning, predict, evaluate and ponder (Anderson, 2009). Therefore, making troubleshooting questions means making challenging questions for learners to be able to perform analysis, reasoning, and evaluation in resolving a problem.

In general, the test instrument has two functions. First, as a tool for measuring or assessing learning outcomes such as problem-solving capabilities. The test is one of the tools teachers can use to understand the success rate of learners in learning (Muri Yusuf, 2015). Assessing the ability of high-level thinking is certainly different from assessing low-level abilities. Rofiah explained that high-level thinking skills need to be measured with the appropriate test language because a good test should be able to measure the skills being measured (Rofiah, Aminah, & Ekawati, 2013). The most important requirement in a test instrument is to measure what will be measured (Purwanto, 2006). Secondly, a test or troubleshooting problem can be used to practice problem-solving skills. A specially designed test to practice problem-solving skills can affect the enhancement of these troubleshooting capabilities. The results showed that one of the factors causing low levels of high-level thinking such as problem-solving skills was lacking learners in resolving the questions that measured HOTS (Budiman & Jailani, 2014). Problem-solving capabilities are the ability in mathematics that needs to be trained to learners as long as they learn mathematics in class (Sugiman, 2009). Of these two functions, it appears that the test instrument specifically designed to measure problem-solving skills is needed in mathematics learning, both as an evaluation tool as well as a tool for training learners so that they can Improve problem-solving capabilities.

So far, the study on the development of test instruments assessment of problem-solving skills seen from two things. First, the study examines the development of test instrument problem-solving skills

ISSN: 2005-4238 IJAST

as part of a high-level thinking ability or higher Order Thinking Skill (HOTS). Such as the results of the development of the test instrument assessment of Higher Order Thinking Skill (HOTS) on mathematics subjects for students in grade X SMK Linear equation System (C. F. Lestari, Kristiana, & Kurniati, 2016), for students of grade VIII JUNIOR algebraic material and Geometry (Budiman & Jailani, 2014), for students of class IX accelerated SMP Material line and series (Lewy, 2013), and for high school students material magnitudes and units (Pratiwi & Fasha, 2015). The results of the development are questions that measure high-level thinking skills that include critical thinking, creativity, analysis, reasoning, evaluation, and problem-solving. Secondly, the study examines the development of problem-solving skills tests as a separate ability. As a result of the development of test instruments problem-solving skills for junior high school students in vibration, wave, and sound material and number material (Bidasari, 2017). Of the two studies, it appears that the development efforts of assessment of the problem-solving capabilities have been conducted but are still limited to certain materials. Therefore, it is necessary to develop problem-solving problems of mathematics for other materials.

Besides, studies on the development of test instruments assessment of problem-solving skills are carried out on two levels of education. First, in the first or JUNIOR secondary education level, such as the development of problem-solving ability test to material & function, straight-line equation, system Linear equation two Variable, circle, Pythagoras, and build flat side chamber (Sinaga, 2016). Second, in upper or high school levels, such as the development of test problem solving for function material (Ferdianto, Fadiyah, & Sunawan, 2019). These two trends seemed to lack the development of test instruments of mathematical problem-solving skills at the elementary and higher education level. The development of problem-solving test instruments in the higher education level is also needed. Widjajanti explained that the obstacles faced by a lecturer in developing mathematical problem-solving skills of candidates for mathematics teachers include choosing the appropriate problem to guide the lecture (Widjajanti, 2009).

The research seeks to complement the shortcomings of the study on the development of test instruments that measure problem-solving skills in students. Therefore, the purpose of this research is to develop a test instrument that can measure the ability to solve math problems to prospective teachers in elementary school. The test instrument developed is an instrument that has a good quality that is stated to be valid, reliable, has a good difficulty level, and can distinguish the students ' ability in solving mathematical problems.

II. METHOD

This research is a research and development that adopts the 4D (four D Model) development, model. Model 4D includes four stages namely the definition stage (define), the design level, the development stage, and the spread (disseminate). The 4D development Model used in the study was modified to 3-D without the deployment stage.

The defining stage (define) is a stage that defines the requirements in the manufacture of assessment instruments. The activities undertaken in this stage are the analysis of student learning results, mathematical learning process analysis, and analysis of the mathematical test instruments used in the PGMI Prodi in the city of Metro Lampung.

The design phase aims to design the assessment instruments of good problem-solving skills according to the criteria and rules of manufacture. The activities undertaken at this stage are to develop a test instrument that refers to the indicator of access to mathematics, problem-solving indicators, and criteria for problem-solving. The indicators of the problem-solving skills referred to in this research are the ability of students to understand the problem, make a plan of completion, perform settlement, and re-examine (K. E. Lestari & Yudhanegara, 2017; NCTM, 2000).

The developing stage aims to generate a draft test instrument assessment of a valid problem-solving ability. Activities conducted at this stage include the validation of assessment instruments by experts who followed by revisions and trials of problems solving the problem to students. The test results are used to measure the validity, reliability, difficulty level, and the varying power of grain items developed.

ISSN: 2005-4238 IJAST

This research data includes qualitative and quantitative data. Qualitative data in the form of advice or input from experts related to the validity of the test instruments, while quantitative data includes the validation results of test instruments and test results of instruments on students of study program PGMI IAIN Metro Lampung Indonesia. The data collection techniques used are questionnaires and tests. Questionnaires are used to obtain qualitative data in the form of feedback from experts and quantitative data in the form of validation of the test instrument. The poll measures the suitability for items with 1) Learning achievement indicator, 2) problem-solving indicator, 3) criterion problem solving, 4) good and correct Indonesian grammar, and 5) correct answer/completion. Further tests are used to obtain the data on the result of problem-solving skills (instrument test results).

Qualitative data is analyzed in a simple qualitative descriptive, while data relating to the validation of test instruments are analyzed using the CVR (Content Validity Ratio) formula (Susetyo, 2015). The validity and reliability of item grain items are calculated by using the Cronbach Alpha formula. The difficulty level of test grain is calculated by dividing the number of scores of students that answer correctly with the student number, and the difference in the rate of the test is calculated from the result of reduction of the difficulty level of an upper group with a lower group difficulty level (Bagiyono, 2017).

III. RESULT

The results of the research and development of test instruments are described according to the following phases.

3.1. Define

Preliminary study results in the study included mathematical learning results, a description of the process of learning mathematics, and the description of test instruments of mathematics learning results used in the PGMI in the city of Lampung.

Math Learning results illustrate students ' ability to master concepts, procedures, and solve mathematical problems. Data of learning Mathematics is obtained through the test and observation of the learning process. Based on the results of the tests that have been conducted showed that students generally have difficulties in resolving non-routine mathematical problems. To strengthen the results of the observation was conducted a test to 10 students PGMI who have taken the mathematics course. The test is a question of a non-routine story or problem-solving. Based on the results the test is known to have 30% of students due and 70% of students are not complete. Data shows that most or 70% of students have poor mathematical problem-solving skills.

Based on the results of observations during the lecture process shows that the practice of problem-solving still a little. This is because in general, the books or materials of mathematics used very little to give sample problem-solving. The problems that are contained in mathematics books or materials are more routine mathematics. This situation is one of the factors that affect students ' low problem-solving skills. It is, therefore, necessary for the development of teaching materials and test instruments that can measure problem-solving skills in students. This research focused on the development of the instrument test problem-solving. The results of the development of the test instruments can be used to measure and train problem-solving skills, to improve the ability of mathematical problem solving for students.

3.2. Design

The product design phase includes designing the problem-solving test instruments by the criteria for problem-solving. The problem indicators problem solving developed in this study are: 1) mind-challenged (challenging), 2) not automatically known how to solve it (non-routine), and 3) allows there is a variation of completion. Products or troubleshooting tests that have been compiled are then validated by experts.

3.3. Develop

The results of the development of test instruments include the results of the validation of assessment instruments by experts who followed by revisions and trials of problems solving a

problem to students. The test results are used to measure the validity, reliability, difficulty level, and the different power-related problems solving developed.

2.1.1. Validation result of expert test instruments

The process of instrument validation by experts is twice the validation of the 1st and 2nd instruments. Validation or assessment of the test problem-solving skills by the expert based on 1) the conformity of the problems with the achievement indicators, 2) the conformity of items with the indicator of problem-solving capabilities, 3) The conformity of grain problems with the criteria problem solving, 4) Alternative truths of answers The result of the validation/assessment of the expert was analyzed quantitatively and qualitative descriptive. Quantitative data analysis using a Content Validity Ratio (CVR) formula. Based on the validation result of the test instrument obtained the following data:

Table 1. Test instrument Validation Result Troubleshooting (1st validation)

Number items	Expert Assessment Results						Amount	Indeks CVR	Assessment Results
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6			
1	0,8	1	1	1	1	1	5.8	0.933	Valid/worthy of use
2	1	1	1	1	1	1	6	1	Valid/worthy of use
3	0,8	0,8	1	1	1	1	5.6	0.867	Valid/worthy of use
4	0,8	1	1	1	1	1	5.8	0.933	Valid/worthy of use
5	0,8	1	1	1	1	1	5.8	0.933	Valid/worthy of use
6	0	1	1	1	1	1	5	0.667	Valid/worthy of use
7	0	1	1	1	1	1	5	0.667	Valid/worthy of use
8	0	1	1	1	1	1	5	0.667	Valid/worthy of use
9	0	1	1	1	1	1	5	0.667	Valid/worthy of use
10	0	1	1	1	1	1	5	0.667	Valid/worthy of use

Table 1. above shows that the Content Validity Ratio Index (CVR) each item is greater than 0,5. This indicates that all item items are valid or eligible to be used. Thus the test is considered valid and can be used to measure mathematical problem-solving skills.

Although quantitative data analysis results suggest that the ten items above are considered valid but there are some suggestions or feedback from experts to correct the problem. As for some advice from experts related assessment Test instruments problem solving the following:

Table 2. Expert advice/expert to repair Item question

Question item	Comments or suggestions for improvement from experts	Conclusion
1	Item number 1 is fixed to a more challenging question (Expert 1), questions do not make points (expert 1, expert 2)	A question worthy of use with revisions
2	Item number 2 is good and challenging, according to Criterion Question solving (Expert 1, expert 2, expert 3, expert 4, expert 5, expert 6)	Problem worthy of Use
3	Item number 3 is often present and less challenging (expert 1, expert 2, expert 3, expert 4), can be used if the number is made larger to be more complex (expert 1, expert 2, expert 4)	Problem worthy of use with revisions
4	Items about number 4 are often and less challenging (expert 1, expert 4), can be replaced with question number 8 (Expert 3)	Problem worthy of use with revisions
5	Item question number 5 is less challenging. Editorial sentences and questions about number 5 are fixed as advised to become more challenging question solving for students (expert 1)	Problem worthy of use with revisions

Expert suggestions or comments as in Table 2. above are assigned a matter of number 1 to 5. Next, to question numbers 6, 7, 9, and 10 there are no comments or suggestions of improvements from experts. It is because in general, the validator argues that concerning the allocation of time and difficulty types of problem-solving questions, it is recommended that the number of 10 items prepared is reduced. In addition to reducing the number of questions, it is suggested that it is made more challenging and has a higher level of difficulty adapted to the student's intellectual level. Here are examples of items that are corrected according to expert advice, which is more challenging for students, how the solution is not known automatically and has several alternative solutions.

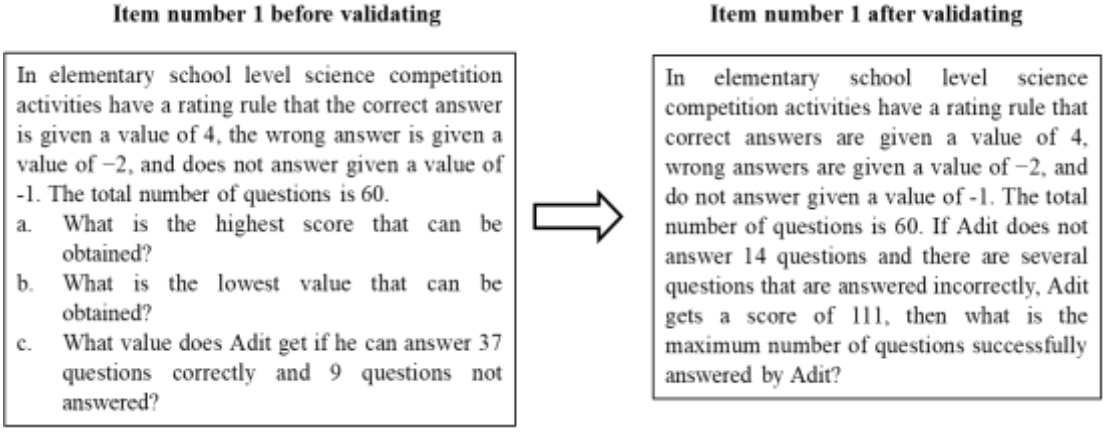

Figure 1. Sample Item question number 1 before and after validated

Figure 1. shows item change number 1 before validated and after revised. According to experts that the initial condition of item number 1 is not challenging and completion can be known in advance. Therefore, item number 1 should be changed to be more challenging so it can be categorized into problem-solving. Improvements can be initiated by changing the word editor and eliminating points on the item. In the question item number 1 (before validated) asked is the final value obtained by Adit, and the number of questions that are correct and answered incorrectly is known. The information will make the student answer the questions directly or the completion can be found easily. In contrast, the item (revision result) asked is the numbers that make up the final value obtained by Adit is 111. To answer the question, the student must specify several different numbers so that if they are aggregated and adjusted to the information about generating a value of 111. Furthermore, students can also use different settlement strategies, such as retry and test strategies table-making strategies, making pictures, and others.

Based on the results of the first validation of the problem-solving test it can be decided that the problem solving that will be used as many as 5 items, about number 1, 2, 3, 4, and 5 that have been fixed according to expert advice. After all five items were corrected according to expert advice, a second validity test was conducted on the five points. This is done to get a decision on whether the problem that has been corrected is considered valid by all experts. The second validation result related to the troubleshooting test can be seen in Table 3.

Table 2. Test result validation Instrument problem Solving (2nd validation)

Question item	Validator Rating						Amount	Indeks CVR	Assessment Results
	1	2	3	4	5	6			
1	1	1	1	1	1	1	6	1	Valid/worthy of use
2	1	1	1	1	1	1	6	1	Valid/worthy of use
3	1	1	1	1	1	1	6	1	Valid/worthy of use
4	1	1	1	1	1	1	6	1	Valid/worthy of use
5	1	1	1	1	1	1	6	1	Valid/worthy of use

The result of the 2nd validation described in Table .3 above shows that all of the question items of numbers 1, 2, 3, 4, and 5 are considered valid or well-used. Besides, all the validators assess the five items according to the indicator of problem-solving problems without any suggestions. Thus, the developed test instrument can be declared valid according to the mathematics education experts to measure mathematical problem-solving skills.

2.1.2. Trial results

At the stage of development, the instrument tested to 30 students. The test results are analyzed for the validity, reliability, difficulty level, and the varying power of grain items.

2.1.2.1. Validity and reliability test results

A good test instrument must meet the characteristics of a good test of validity, reliability, objectivity, and practicality (Purwanto, 2006). Of the four characteristics, at least two characteristics are considered very important and often made the basis of determining the reliability of a test of validity and reliability (Raya, 2016). The validity and reliability of the test instruments are calculated using the Cronbach Alpha formula. The results of the validity and reliability test of the test instrument are as follows:

Table 3. Validity Test Output with SPSS

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Item_1	21.00	53.724	0.867	0.763	0.927
Item_2	21.43	52.944	0.839	0.738	0.932
Item_3	20.40	49.352	0.811	0.679	0.943
Item_4	21.37	54.447	0.890	0.798	0.924
Item_5	21.80	56.234	0.875	0.778	0.928

The value of Corrected item-total Correlation in Table 4. above shows the validity value of the problem item and Cronbach's Alpha if Item Deleted shows the reliability value of the problem item. To find out if the item is valid and reliable, it should be compared to r_tabel. On r_tabel can be seen for degrees of freedom or DF = 28 and the probability 0.05 is 0.374. After the value of Corrected item-total Correlation with the value, r_tabel indicates that all Correlation Item Corrected values for items of item number 1, 2, 3, 4, and 5 are larger than 0.374. This means that the numbers 1, 2, 3, 4, and 5 items are valid.

The next value of Cronbach's Alpha if Item Deleted is also greater than the r_tabel value of 0.374. It shows that all the items of the problem are the numbers 1, 2, 3, 4, and 5 are reliable. To know the reliability of the tests as a whole can be done by comparing the value of Cronbach's Alpha Based on Standardized Items with r_tabel of 0.374. In table 5 can be seen the value of Cronbach's Alpha Based on Standardized Items greater than r_tabel or 0.948 > 0.374. It shows that overall the test instrument developed is reliable.

Table 4. Reliability Test Output with SPSS

Cronbach's Alpha	*Cronbach's Alpha Based on Standardized Items*	*N of Items*
0.944	0.948	5

Based on the results of the validity and reliability test, the above describes that the test instrument developed is proven to be valid and reliable, either in whole or every item of matter. This results in explaining that item items of numbers 1, 2, 3, 4, and 5 can measure the ability to accurately solve students ' math problems.

2.1.2.2. Difficulty level and different power

A good instrument is also determined by the level of difficulty and differentiation of each item items. Analysis of the difficulty level of the item is to examine item items in question from the difficulty side, so that items can be obtained which is a matter of easy, moderate, and difficult. Items must also be able to distinguish the clever students from less clever students. The distinguishable power of an item can be known by looking at a large number of item discrimination indexes. Based on the test result of difficulty level and differentiation power obtained the following data:

Table 5. Difficulty level test result and differentiator power

	Difficulty level test Result				**Differentiator Power Test Results**				
Question item	**Number of correct answers**	**Ideal score**	**Difficulty level**	**Criteria**	**Average upscale**	**Average Lower class**	**Daya Beda**	**Criteria**	**Conclusion**
1	165	10	0.55	Medium	7.78	4.00	0.38	Medium	Used
2	152	10	0.51	Medium	7.33	3.44	0.39	Medium	Used
3	183	10	0.61	Medium	9.11	4.33	0.48	High	Used
4	154	10	0.51	Medium	7.22	3.78	0.34	Medium	Used
5	141	10	0.47	Medium	6.67	3.78	0.29	Medium	Used

In Table 6. above, it is obvious that all item items have difficulty level with "medium" criteria. This is very good because a good item problem is if the items are not too difficult and also not very easy (Muri Yusuf, 2015). Therefore, from the aspect of difficulty level can be explained that the test instrument developed can be used to measure the ability of problem-solving mathematics.

Further from the differentiator power test results can be seen that items number 1, 2, 4, and 5 have moderate criteria, and item number 3 has high criteria. The criteria are that a matter of item is enough to distinguish students who are good at less clever students. High criteria mean those item items are good or can distinguish clever students with less clever students. Therefore, from the distinguishing aspect of the power item can be explained that the test instrument developed can be used to distinguish students who have good problem-solving skills and the less good.

IV. DISCUSSION

Based on the research results outlined above can be analyzed into several points. First, this research is backed by the need for math learning about test instruments capable of measuring the ability to solve mathematical problems. Because preliminary study results showed that the test instruments that have been used have more measured the low level of ability. However, problem-solving capability is one of the competencies that will be achieved in the elementary/MI Mathematics course. Therefore, instruments that can measure and assess problem-solving capabilities.

Secondly, the process of the development of test instruments begins with creating a test instrument design. The test instruments on this research are designed by the indicators of mathematical learning access, problem-solving indicators, criteria for problem-solving, rules of good and correct Indonesian grammar, and preparing alternatives to correct answers/solutions. The basic test preparation is the result of reviewing the problem-solving variables and drafting an assessment instrument. Thus, in theory, the test instrument developed can be determined to be valid to fulfill the validity of content validity and the validity of the construction (construct validity). A test is said to have the validity of the contents when it can measure the competencies that are developed along with their indicators and learning materials (Widoyoko, 2012). The validity of the construction refers to the extent to which an instrument measures the concept of a theory that is the preparation of the instrument (Widoyoko, 2012).

Thirdly, to prove if the test instrument is valid both in content and construction, then the validity test is carried out by the experts (expert judgment). Based on expert assessment results shows that all item items developed are valid. The process of validation of test instruments assessed by six

experts and performed 2 times indicates that the test instrument assessment of this mathematical problem-solving ability is valid and reliable. Thus, in theory, the test instrument developed is valid by experts and can be used to assess or measure the ability of mathematical problem-solving.

Fourth, the test result test instrument is used to determine the validity of the test item. The validity of item items is very important because if there is an invalid item, it will reduce the overall validity of the instrument. The result of the instrument is to show that all item items of matter i.e. number 1, 2, 3, 4, and 5 are considered valid and reliable. Thus, from the results of the validation of the test instruments shows that the question of number 1, 2, 3, 4, and 5 is valid and reliable either in theory (expert judgment) or whole or item items.

Grain quality is also determined by the difficulty level and differentiation power of each item items. Based on the test results shows that all item items have difficulty levels with moderate or very good criteria. As seen from the differentiation power aspect indicates that the test instrument developed indicates that items number 1, 2, 4, and 5 have moderate criteria, and item number 3 has high criteria. The criteria meant that item items could distinguish good students with less clever students. Therefore, the aspect of distinguishing power can be explained that the test instruments developed can distinguish students who have good problem-solving skills and who are not good.

The important findings of the products produced in this research are five essay problems capable of measuring problem-solving skills. Problem essay has the potential to measure student performance comprehensively (Srihari et al., 2008). The novelty of the five problem problems solving this is challenging the students to complete, having a solution that has not been known before, and has some alternative solutions. Product problem Solving can also be used to help educators implement troubleshooting learning.

Solving mathematical problems is a complex cognitive activity that involves several processes and strategies. From the results of the analysis of the answer/resolution of problem-solving shows several factors that affect the success of students solve problems solving a problem that is the ability to understand the problem and do calculations. It is in line with Fung DKK's research results that reading and calculating skills have a direct effect on the accuracy of problem-solving (Fung, Swanson, & Orosco, 2014). Lack of linguistic, strategic, and procedural knowledge is the main source of mistakes experienced by students in solving mathematical problems (Cardelle-Elawar, 1992). These findings can be used as an evaluation material to design more effective troubleshooting learning.

Scenario is of three squares of instructive grounds. State any client john of type training needs to utilize the administrations gave by universal processing than he need to initially enroll himself with the server through the organization. State after enlistment client is furnished with a novel client id and introductory trust estimation of level 3. He is presently permitted to utilize all the typical administrations and paid administrations gave by the specialist organization. State following a half year his legitimate access proportion is 1:3 than his trust level will be decreased to level 2 as it ought to be among ½ and 4/5 and on the off chance that his lawful access proportion is 4:5, at that point his trust level will be moved to level 4. Right now trust level will fluctuates as per the lawful gets to he have made in most recent a half year. His most recent trust level will be utilized to give him approval to the utilization of different administrations. As indicated by the most recent trust level he will be approved to next degree of services.

V. CONCLUSION

This research resulted in a test instrument capable of measuring mathematical problem-solving skills for class teacher candidates in elementary school. The test instrument produced as many as five items in the form of essay problems. The results of the validation show that the problem belongs to the category of problem-solving which has challenging nature, indirect resolution, and some correct workaround alternatives. The test results also show that the test instrument developed is an instrument that is of good quality, meaning it has been declared valid and reliable, has an excellent level of difficulty, and can distinguish the students ' ability in Solve Math problems.

Preliminary study results show that problem-solving is very limited. Though it takes variations of

the question that can assess high-level thinking skills such as problem-solving. Therefore, the products of this research result can be used to assess problem-solving capabilities. Also, problem-solving can be used in mathematics learning to practice students ' ability to solve non-routine problems.

ACKNOWLEDGMENTS

Our gratitude goes to the promotor, co-promotor, mathematics lecturer, PGMI student, Jakarta State University, Ministry of Religion of the Republic of Indonesia, IAIN Metro who have supported the research activities carried out correctly.

REFERENCES

1. Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving Higher Order Thinking Skills (HOTS) problems for the topic of fraction. *Asian Social Science*, *11*(21). https://doi.org/10.5539/ass.v11n21p133
2. Anderson, J. (2009). Mathematics Curriculum Development and the Role of Problem Solving. *ACSA Conference*, 1–8.
3. Bagiyono. (2017). Analisis Tingkat Kesukaran dan Daya Pembeda Sial Ujian Pelatihan Radiografi Tingkat 1 The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination. *Bagiyono*, *16*(No. 1), 1–12. Retrieved from http://repo-nkm.batan.go.id/140/1/05_analisis_tingkat_kesukaran.pdf
4. Bidasari, F. (2017). Pengembangan Soal Matematika Model PISA pada Konten Quantity untuk Mengukur Kemampuan Pemecahan Masalah Matematika Siswa Sekolah Menengah Pertama. *Jurnal Gantang*. https://doi.org/10.31629/jg.v2i1.59
5. Brookhart, S. M. (2010). *How to assess higherorder thinking skills in your classroom*. Alexandria: ASCD.
6. Budiman, A., & Jailani. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) ... (Agus Budiman, Jailani) - 139. *Riset Pendidikan Matematika*, *1*(November 2014), 139–151. https://doi.org/https://doi.org/10.21831/jrpm.v1i2.2671
7. Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education*, *8*(2), 109–121. https://doi.org/10.1016/0742-051X(92)90002-K
8. Ferdianto, F., Fadiyah, F. S., & Sunawan, M. D. (2019). Pengembangan Perangkat Pembelajaran Melalui Model Problem Based Learning Berorientasi Kemampuan Pemecahan Masalah Matematis pada Materi Fungsi Kelas X SMA. *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, *3*(2), 165. https://doi.org/10.31331/medivesveteran.v3i2.799
9. Fung, W. W., Swanson, H. L., & Orosco, M. J. (2014). Influence of reading and calculation on children at risk and not at risk for word problem solving: Is math motivation a mediator? *Learning and Individual Differences*, *36*, 84–91. https://doi.org/10.1016/j.lindif.2014.10.011
10. Hidayat, T., Susilaningsih, E., & Kurniawan, C. (2018). The effectiveness of enrichment test instruments design to measure students' creative thinking skills and problem-solving. *Thinking Skills and Creativity*. https://doi.org/10.1016/j.tsc.2018.02.011
11. Kosko, K. W. (2019). A multiplicative reasoning assessment for fourth and fifth grade students. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2018.11.003
12. Lestari, C. F., Kristiana, A. I., & Kurniati, D. (2016). Pengembangan Paket Tes Matematika Berbasis Kemampuan Berpikir Tingkat Tinggi Siswa Kelas X TKJ SMK Materi Sistem Persamaan Linier. *Jurnal Edukasi*, *3*(2), 34. https://doi.org/10.19184/jukasi.v3i2.3527
13. Lestari, K. E., & Yudhanegara, M. R. (2017). *Penelitian Pendidikan Matematika*. Bandung: Refika Aditama.
14. Lewy, L. (2013). Pengembangan Soal Untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Pokok Bahasan Barisan Dan Deret Bilangan Di Kelas Ix Akselerasi Smp Xaverius Maria Palembang. *Jurnal Pendidikan Matematika*, *5*(1). https://doi.org/10.22342/jpm.5.1.821.
15. Muri Yusuf. (2015). *Asesmen dan Evaluasi Pendidikan* (Edisi Pert). Jakarta: Prenadamedia Group.
16. NCTM. (2000). *Principles and Standards for School Mathematics*. Resto VA: NCTM.

17. Nur Atikah Khairun Nisa; Rani Widyastuti; dan Abdul Hamid. (2018). Pengembangan Instrumen Assesment Higher Order Thinking Skill (HOTS) Pada Lembar Kerja Peserta Didik Kelas VII SMP. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika*, *3*(3), 543¯ 556. Retrieved from http://www.ejournal.radenintan.ac.id/index.php/pspm/article/view/2465/1963
18. Pratiwi, U., & Fasha, E. F. (2015). Pengembangan Instrumen Penilaian Hots Berbasis Kurikulum 2013 Terhadap Sikap Disiplin. *Jurnal Penelitian Dan Pembelajaran IPA*, *1*(1), 123. https://doi.org/10.30870/jppi.v1i1.330
19. Purwanto, N. (2006). *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
20. Raya, P. (2016). Institut Agama Islam Negeri (IAIN) Palangka Raya – 11 Agustus 2016 1, (2011), 1–4.
21. Rofiah, E., Aminah, N. S., & Ekawati, E. Y. (2013). Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP. *Jurnal Pendidikan Fisika*, *1*(2), 17–22. Retrieved from http://www.jurnal.fkip.uns.ac.id/index.php/pfisika/article/view/2797/1913
22. Sinaga, N. A. (2016). Pengembangan tes kemampuan pemecahan masalah dan penalaran matematika siswa SMP kelas VIII. *PYTHAGORAS: Jurnal Pendidikan Matematika*, *11*(2), 169. https://doi.org/10.21831/pg.v11i2.10642
23. Srihari, S., Collins, J., Srihari, R., Srinivasan, H., Shetty, S., & Brutt-Griffler, J. (2008). Automatic scoring of short handwritten essays in reading comprehension tests. *Artificial Intelligence*, *172*(2–3), 300–324. https://doi.org/10.1016/j.artint.2007.06.005
24. Sugiman. (2009). Kemampuan Pemecahan Masalah Matematik Siswa Smp: Problematika dan Cara Melatihnya. *Prosiding Seminar Nasional Penelitian, Pendidikan Dan Penerapan MIPA*, 531–538.
25. Susetyo, B. (2015). *Prosedur Penyusunan & Analisis Tes*. Bandung: Refika Aditama.
26. Widjajanti, D. B. (2009). Kemampuan Pemecahan Masalah Matematis Mahasiswa Calon Guru Matematika: Apa Dan Bagaimana Mengembangkannya, 978–979.
27. Widoyoko, E. P. (2012). *Teknik Penyusunan Isntrumen Penelitian* (Edisi Pert). Yogyakarta: Pustakan Pelajar.