

ITEM QUALITY ANALYSIS FOR MEASURING MATHEMATICAL PROBLEM-SOLVING SKILLS

Palupi Sri Wijayanti

Universitas PGRI Yogyakarta, Yogyakarta, Indonesia

E-mail: palupi@upy.ac.id

Received 12 August 2020; Received in revised form 29 December 2020; Accepted 30 December 2020

Abstract

This research aims to determine the quality of the questions used to measure students' mathematical problem-solving skills. This type of research includes quantitative descriptive research. The research subject is a grade VII student with a primary data source is scores of mathematical problem-solving skills. The research instrument used was a written test to measure problem-solving abilities. The data analysis technique used is a quantitative analysis by looking at the validity, reliability, differentiation power, and bullet-grain index of the matter. The results showed that the problem has been valid, the reality belongs to the category quite well, the differentiator power for three items, including bad, enough, and good classes. Meanwhile, the difficulty index shows a number in the easy category, and two names are included in the medium category. Thus, it concluded that the instruments used to measure problem-solving abilities are valid and in a suitable category. It can see it from the instrument trial results with the percentage value of students' problem-solving abilities, which is included in the very high category, namely 20%. The number of students who have problem-solving skills in great variety could be presented at 53.33%. Students who have sufficient and low problem-solving abilities were given at 16.67% and 10%.

Keywords: differentiator power; difficulty index; problem; quality; reliability; validity

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas butir soal yang digunakan untuk mengukur kemampuan pemecahan masalah matematika siswa. Jenis penelitian ini termasuk penelitian deskriptif kuantitatif. Subjek penelitian adalah siswa kelas VII dengan sumber data utama adalah skor kemampuan pemecahan masalah matematika. Instrumen penelitian yang digunakan adalah tes tertulis untuk mengukur kemampuan pemecahan masalah. Teknik analisis data yang digunakan adalah analisis kuantitatif dengan melihat validitas, realibilitas, daya pembeda, dan indeks kesukaran butir soal. Hasil penelitian menunjukkan bahwa butir soal telah valid, realibilitas termasuk dalam kategori cukup baik, daya pembeda untuk tiga butir soal termasuk kategori buruk, cukup, dan baik. Sementara itu, indeks kesukaran soal menunjukkan satu nomor dalam kategori mudah dan dua nomor termasuk dalam kategori sedang. Dengan demikian, dapat disimpulkan bahwa instrumen yang digunakan untuk mengukur kemampuan pemecahan masalah telah valid dan berkategori baik. Hal ini terlihat dari hasil uji coba instrumen dengan persentase nilai kemampuan pemecahan masalah siswa yang termasuk dalam kategori sangat tinggi yaitu 20%. Jumlah siswa yang memiliki kemampuan pemecahan masalah dengan kategori tinggi dapat dipresentasikan sebesar 53,33%. Siswa yang memiliki kemampuan pemecahan masalah dengan kategori cukup dan rendah dipresentasikan sebesar 16,67% dan 10%.

Kata kunci: butir soal; daya pembeda; indeks kesukaran; kualitas; realibilitas; validitas



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

The question item is part of the test instrument used for a particular purpose. Tests can be categorized as a

means of measuring an object by collecting interconnected information and fulfilling the characteristics of the object. In this study, the object in

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

question was the cognitive level of students seen from the problem-solving skills.

The question item is part of the test instrument used for a particular purpose. Tests can be categorized as a means of measuring an object by collecting interconnected information and fulfilling the characteristics of the object. In this study, the object in question was the cognitive level of students seen from the problem-solving skills, depending on the results of the stimulus that varies across types of stages or a game that can introduce the significant potential for confusion, namely arousal and cognitive load (Gundry & Deterding, 2019).

An instrument can be said to be a good instrument if the test has been conducted as an analysis of the quality of instruments included in the report of grain items. Several stages are done to obtain a good question that is: to develop a particular purpose test and usability, to compose items, to conduct test tests, required test tests, conducting grain analysis, revising criteria if there is still less precise, rearranging tests, using principles, and performing an interpretation of test results. This is supported by another opinion, which states that the first step in constructing the instrument is identifying the problem and then presenting it in a formula so that it is easy to understand its purpose. The second step is to design an instrument by the capabilities to be measured (Supardi et al., 2019).

Theoretically, it is said to be a good instrument if it can be used to improve the test results that interpret how far away the capabilities are. A study can determine and take a conclusion if the tools used in retrieving data are of good quality to provide the exact picture associated with real and

actual conditions. Therefore, the instrument that has a good variety of data takers is necessary for the study (Aida et al., 2017).

The test instruments used in this focus are instruments used to measure students' problem-solving skills. After knowing the results of the measurement of mathematical problem-solving ability, it will be easy to see the level of understanding of students in associating some mathematical concepts, thus obtaining the right solution.

Problem-solving skills become the underlying thing in learning mathematics from primary to college. The mathematics learning process takes place. It is better to provide a problem that must be resolved by steps that are conveyed orally or in writing, to be able to give provisions to students in improving their ability to think systematically and mathematically (Cholily, Y.M., Kamil, T.R., & Kusgiarohmah, 2020).

Definitive problem-solving capability is one's ability to solve problems with appropriate procedures and steps. The use of methods in the process of solving mathematical problems will facilitate the students to train in logical and systematic thinking (Sumartini, 2018).

As a further form of exercise, the use of mathematical problems faced with the story will provide an experience for students to understand the text of reading and thoroughness in knowing the data elements used in solving the problem.

The process of solving mathematical problems to be directed and systematic can be done with the steps based on the opinion of the experts is the Polya that expressed is: understand the problem, make a plan, carry out our program, look back at the

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

completed solution (Amam, 2017). The teachers and students in solving the question is very well know this workaround, the problem of the story. In general, they hand out the elements known, asked, and the answer process then concludes with a conclusion. Sometimes, students in middle and high school tend to have difficulty rechecking answers. However, it is different from students in low or elementary schools who sometimes have difficulty in planning (Arfiana & Wijaya, 2018).

METHOD

This type of research includes quantitative descriptive. The subject used in the study was a grade VII student at SMP Negeri 3 Jetis school year 2019/2020. The number of students who became the data source consisted of 30 people. All students were adjusted to the research objectives and the learning indicators of achievement goals. Thus, the technique used is purposive sampling. The data source used in this study is quantitative data extracted from students' answers. Data collection techniques were using a test with essay-shaped. This test refers to the indicator and assessment of problem-solving capabilities.

The data analysis technique used is quantitative analysis. Quantitative analysis was done with IBM SPSS 20 and Microsoft Excel to determine the validity, reliability, differentiation power, and index difficulty of the problem. The validity of the instrument in this study was tested using the Pearson product-moment correlation coefficient formula (1):

$$r_{xy} = \frac{N \sum XY - (\sum X) \cdot (\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2] \cdot [N \sum Y^2 - (\sum Y)^2]}} \dots(1)$$

Information:

- r_{xy} : correlation coefficient between item score (X) and the total score (Y)
- N : many subjects
- X : score of an item or statement/question item score
- Y : total score.

In this study, to determine the items' validity, the Correlations test formula was used with SPSS 20 software. In the output, if the Pearson Correlation value is $\geq r_{tabel}$, then the item is valid. However, if Pearson Correlation $< r_{tabel}$, then the thing is invalid. Then, to determine the reliability of the test instrument using the Alpha Cronbach formula (2), namely:

$$r = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum s_i^2}{s_r^2} \right) \dots(2)$$

Information:

- r : reliability coefficient
- n : many items
- s_i^2 : variance of the score of item i
- s_r^2 : the variance of the total score

Where:

$$s_r^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}, \text{ untuk subjek, } n \leq 30 \text{ atau}$$

$$s_r^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n}, \text{ untuk subjek, } n > 30$$

This study to test the reliability using the Cronbach-alpha test with the help of SPSS 20 software. At the SPSS output, if the Cronbach-alpha value is $\geq \alpha$ (0.05), then the instrument is reliable. However, if the Cronbach-alpha value $< \alpha$ (0.05), the instrument is not reliable.

In this study, to determine the distinguishing power using the help of Microsoft Excel with the formula (3):

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

$$DP = \frac{\bar{X}_A - \bar{X}_B}{SMI} \quad \dots(3)$$

Information:

DP = item distinguishing index

\bar{X}_A = average score of the answers of the upper group students

\bar{X}_B = the average score of the answers of the lower group students

SMI = Maximum Ideal Score, which is the maximum score a student will get if he answers the item correctly (perfectly).

In this study, to determine the distinguishing power using the help of Microsoft Excel with the formula (4):

$$IK = \frac{\bar{X}}{SMI} \quad \dots(4)$$

Information:

IK = item difficulty index

\bar{X} = the average score of students' answers on an item

SMI = Maximum Ideal Score, which is the maximum score a student will get if he answers the questions correctly.

RESULTS AND DISCUSSION

Concerning the difficulty level of the item, other researchers defined it as the proportion of test-takers who correctly answered the question (Angriani et al., 2018). The level of difficulty of the items was seen from students' ability or ability to answer them, not from the assumption of the teacher who compiled the questions, because items that are difficult or easy for the teacher are not necessarily difficult or easy for students. An item of question can distinguish between able students and less capable students. The

ability of such an item is called discrimination (Aida et al., 2017).

Based on the results of data analysis obtained, the results of the quality grain problem with the characteristics of validity, reliability, different power, and difficulty level as follows.

1. Validity

Validity is calculated using the product correlation formula of the Pearson moment. The number of students working about 30 students, so it is known $n = 30$, and the value of the R table shows the number 0.361.

A test is called valid or has validity when the test can precisely gauge what to measure. As additional information is, there is an expert opinion which states that the internal validity hierarchy is essential. Overall validity and external validity are considered as second things to consider after internal validity. However, sometimes perspective is not always the right thing (Westreich et al., 2019).

Based on the analysis of the three questions can be found that the three problems are categorized as valid (Tabel 2).

Table 2. Validity calculation results.

No.	Validity	category	Interpretation
1.	0,438	Valid	Good enough
2.	0,882	Valid	Good
3.	0,883	Valid	Good

The results of the analysis of the research data on Question No. 1 indicates the validity of 0.438 so that the criteria is good enough. Student response also showed that students answer this question sufficient in measuring the ability of problem-solving mathematics in daily life by 66.7%, and students answered to have encountered the same type of 60%.

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

In Question, No. 2 indicates the validity of 0.882 so that in proper criteria, the student response also shows that it can measure the ability of math problem solving by 86.7%, and students claimed to have never encountered this type of question before at 53.3%.

In Question, No. 3 indicates the validity of 0.883 so that in proper criteria, the student response also shows that it can measure the ability of math problem solving by 83.3%, and students claimed to have never encountered this type of question before at 56.7%.

Another supportive opinion also states that the discrimination of an item is the item's ability to differentiate between students who score high and score low (Angriani et al., 2018). In terms of distinguishing power, a good question is answered correctly by a test taker who is able/clever / mastering the test material, and cannot be answered correctly by a test participant who has not mastered the test material.

The level of difficulty is a number stating the degree of difficulty of an item of matter. A good question is not too easy or not too difficult (Susanti et al., 2017). The problem is too easy not to stimulate students to heighten the effort to break it. Otherwise, the problem that is too difficult will cause students to become discouraged and have no enthusiasm to try again because it is beyond its reach (Arifin, 2017). The difficulty level is how easy and how difficult a problem is for students (Hayati & Lailatussaadah, 2016). The higher the percentage of students answers to the problem correctly, the easier it is, the smaller the percentage of students answer the question correctly, the harder the challenge.

2. Reliability

The reliability test is used to see the consistency of grain problems in

measuring students' troubleshooting skills. The word reliability comes from the word reliability, from a reliable word that means trustworthy. Tests are said to be credible if they provide fixed results when dealt with repeatedly. The reliability of the items on this instrument is only used on the subject of this study, with conditions adapted to reality (Priyambodo & Marfuatun, 2016). The results of the Reliability test are presented in the Table 3.

Table 3. Reliability calculation results.

Cronbach's Alpha	N of Items
.636	3

The results of test instrument analysis using IBM SPSS 20 program application known that reliability for T.M. 1st problem of 0.636 with the category is good enough. It can be noted that this problem has consistency when given to the same subject even by different people, different times, or different places, it will provide the same relative results.

A test is said to be reliable when the results of the analysis show a decree. The reliability of an instrument is the gift or consistency of the instrument when given to the same subject even by different people, different times, or different places. It will provide the same relative results. The items are said to have a reliable construct if the test results show the Cronbach alpha value, and the construct reliability is 0.7 (Bintarti & Kurniawan, 2017).

3. Power differentiator

The differentiator power analysis by using Microsoft Excel obtained the results in Table 4.

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

Table 4. Differentiator power calculation result

No.	Differentiator Power	Interpretation
1.	0,133	Bad
2.	0,252	Enough
3.	0,437	Good

The distinguishing power analysis on question number 1 shows a characteristic power of 0.133 so that in inadequate criteria, most respondents stated that this problem is straightforward, and data shows that the student score is almost entirely correct. Hence, the problem is bad in distinguishing the students who are highly capable and low-skilled.

Meanwhile, in question number 2 indicates the differentiator power of 0.252 so that insufficient criteria. It means that they can distinguish high-ability and low-skilled students. It is different in question number 3 with proper criteria with differentiation of 0.437. The results showed this problem in both distinguishing high-ability and low-skilled students.

The distinguishing power of a single question expresses how far the ability of the item differentiates between students who can answer questions appropriately and students who are unable to answer the question adequately (Dewi et al., 2019). The distinguishing power of the problem is the ability to distinguish between intelligent (highly skilled) students and ignorant (low-skilled) students (Arifin, 2017). Different power analysis means reviewing test questions in terms of the strength of the test to distinguish students belonging to the low and high category categories (Angriani et al., 2018). The distinguishing power of the problem is the ability of a test item to

differentiate between a highly capable and low-capacity testee.

4. Tribulation Index

Analysis of difficulty levels by using Microsoft Excel obtained the results in Table 5.

Table 5. Success rate calculation result

No.	Tribulation Index	Interpretation
1.	0,785	Easy
2.	0,533	Moderate
3.	0,574	Moderate

Analysis of the difficulty index shows the difficulty level of 0.785, so that in easy criteria. The student response shows that students consider this question to be easy at 63.3%, and students get an answer to this problem independently of 83.3%.

In question number 2 obtained the result of difficulty level 0.533, so that in the medium criteria. The student's response shows that students consider this to be 80%, and on the other, the students get an answer to this problem independently of 56.7%.

The characteristic of number 3 indicates the severity of 0.574 so that it is in moderate criteria. The student response shows that students consider this to be 80%, and students have an answer to this problem independently of 80%.

The analysis of the validity, reliability, level of difficulty, and distinction was intended to reveal the quality of the items so that in this study, the researcher wanted to analyze the validity, reliability, difficulty level, and discrepancy of the items that could be used to reveal students' problem-solving abilities (Almanasreh et al., 2019).

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

The form of the questions analyzed in this study is in the form of problem-solving questions in the form of descriptions with the solution guidelines adjusted to the problem-solving stages proposed by Polya (Lee, 2017). Using these problem-solving steps is straightforward and straightforward so that it is easy to analyze student answers. The items were analyzed for junior high school students to solve geometry, namely flat shapes. The three items that were designed had good quality in terms of reliability, validity, difficulty index, and difference power index (Hidayat et al., 2019).

The results of the reliability calculation gave a Cronbach's alpha value of 0.636. The Cronbach's alpha value is categorized in the reliable items with the moderate category because 0.636 is in the 0.50-0.70 interval. These results indicate that test items' problem-solving ability can be trusted because they tend to provide fixed results (Mohammed, 2019). Although it has been declared reliable, it is not sufficient and must be combined with validity because reliable test questions are not necessarily valid (Martin et al., 2020).

The readability of the test instruments used to measure problem-solving skills (Baştürk, 2016) with the analysis of grain quality problem is to be sought by the characteristics of instruments test problem-solving math problems seen from validity, reliability, differentiation power, and index of tribulations. Reliability analysis was carried out on several test items as a whole, but the validity analysis was carried out on each test item (Priyambodo & Marfuatun, 2016).

The validity of the problem is a degree of precision between data that

occurs on the research object with data that can be reported by researchers. Validity is one of the crucial things in the use of instruments for research and practice (Almanasreh et al., 2019).

A test is said to have validity if the result corresponds to the criteria, in the sense of having a parallel between the test result and the requirements. In scientific research, validity is the primary construct and indicates the quality of the study (Gundry & Deterding, 2019).

The items' validity is the measuring accuracy that is owned by an item, which is an integral part of the test as a totality. The same is the case with the level of difficulty and difference in the calculated items individually. Test the validity, difficulty index, and different power of the three questions as described in Figure 1.



Halaman rumah berbentuk persegi panjang berukuran panjang 70 meter dan lebar 55 meter. Di sekeliling halaman itu, akan dipasang pagar dengan biaya Rp 125.000,00 per meter. Berapakah biaya yang diperlukan untuk pemasangan pagar tersebut?

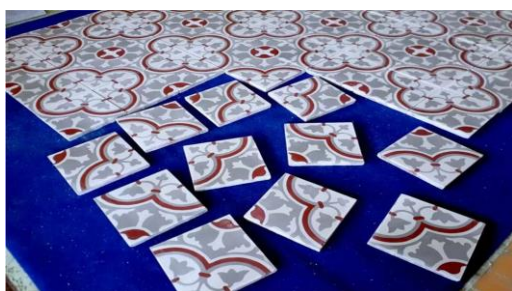
Figure 1. Extract question number 1

The results of the calculation in determining the validity of item number 1 give the Sig. (2-tailed) of 0.438, which means that the item is valid. The result of the difficulty levels was 0.785. These results indicate that the problem difficulty level is in the easy category and the discrimination index is 0.133 in

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

the not good enough category. It provides information that the abilities that will be displayed by students when solving problems in question number 1 can easily find solutions (Hayati & Lailatussaadah, 2016). Based on these results, the researchers were also able to find new information that question number 1 was a routine question for students at SMP N 3 Jetis. Thus, questions with problem types such as number 1 should not be used continuously to differentiate between students' abilities.

Furthermore, item number 2 aims to measure students' ability to add algebraic forms concerning finding the area of a flat shape (Figure 2)



Sebuah lantai berbentuk persegi dengan panjang sisinya 12 m. Lantai tersebut akan dipasang ubin berbentuk persegi berukuran 30 cm x 30 cm. Banyaknya ubin yang diperlukan untuk menutup lantai?

Figure 2. Excerpt from question number 2.

The result of calculating the validity of item number 2 (Figure 2) gives a value of 0.882. Statistically, this means that item number 2 is valid. While the results of the difficulty index of 0.533 (moderate) and the discrimination index of 0.252 were in a relatively good category because they were in the $20 \leq D < 40$ intervals. These questions can provide students with experience in honing what must be prepared in solving the questions as consideration is the size of the floor and

the tiles' size. If students have been able to identify this, it can be said that the students have met the indicator of problem-solving abilities (Putri & Sutarto, 2018). Furthermore, if students can solve it well, they can solve problems with coherent and systematic procedures. Indicator question number 3 measures students' ability to formulate and add algebraic forms such as the following passage.

A pool is rectangular. If the pool area is 32 m² and the pool length is twice the width, then determine the pool's circumference!

Sebuah kolam berbentuk persegi panjang jika luas kolam tersebut adalah 32 m² dan panjang kolam dua kali lebar maka tentukan keliling kolam tersebut !

Figure 3. Excerpt from question number 3.

The result of calculating the validity of item number 3 is 0.883. This statistical value shows that item number 3 is valid. The difficulty index calculation is 0.574 in the moderate category because it is in the interval $0.30 \leq DI < 0.80$, and the discrimination index is 0.437 in the excellent category. It shows that the questions that should be used in measuring problem-solving abilities are of the type as in question number 3.

Based on the description of the research results and the discussion above, it can be concluded that the three test items have good quality in terms of validity, reliability, difficulty level, and distinguishing power and are suitable for measuring students' mathematical problem-solving abilities (Brundle et al., 2019).

The test instrument was then used to measure the students' problem-

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

solving abilities. The analysis used to determine students' mathematical problem-solving abilities, namely the researchers calculated each indicator and the average achievement of the mathematics problem-solving ability steps (Westreich et al., 2019), namely the mathematics problem-solving ability test's average score. The results of students' math problem-solving abilities and average score test are presented in Table 6 and Table 7.

Table 6. Problem-solving ability test results.

No. Item	Indicator	Score	Category
1	A	100,00	Very High
	B	87,00	Very High
	C	95,55	Very High
	D	23,00	Very Low
2	A	90,00	Very High
	B	37,00	Low
	C	62,00	High
	D	20,00	Very Low
3	A	87,00	Very High
	B	41,67	Low
	C	73,33	High
	D	20,00	Very Low

Tabel 7. Average score test.

Indicator Problem-solving skill	Average	Category
A	92,33	Very High
B	55,23	Enough
C	76,96	Very High
D	21,00	Very Low

As an additional explanation, it is called indicator A, which identifies the elements known, asked for, and the adequacy of the elements needed. Indicator B means being able to formulate mathematical problems or compile mathematical models. Indicator C means that it can implement the strategies. Solve the problem.

Furthermore, indicator D describes students who can explain or interpret the results of problem-solving. Based on the research data above, it was known that in test number 1, all students had written indicator A, which explains the elements that are known and asked. In indicator B, most of the students formulate the problem, the rest of the students do not write it down; on indicator C, almost all students are correct in implementing the strategic problem solving, whereas, in the D indicator, not many students explain or interpret the results of problem-solving.

Based on the research data above, it was known that in test number 2, most students had written indicator A, which explains the elements that are known and asked. In indicator B, some students formulate problems for students who do not write them down. Then, in indicator C, some students are correct in applying some students are wrong in implementing the problem-solving strategies, while in indicator D, there are still many students who do not explain or interpret the results of problem-solving.

Based on the research data above, it was known that in test number 3, most students had written indicator A, which explains the elements that are known and asked. On indicator B, some students formulate problems for students who did not write them down. On indicator C, some students are correct in applying some students are wrong in implementing problem-solving strategies, while in indicator D, there are still many students who do not explain or interpret the results of problem-solving.

Based on the results of the test data to measure mathematical problem-solving abilities, it can be seen that the number of students who were included

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

in the category of having a very high level of mathematical problem-solving ability is six students (20%). The number of students with a high level of mathematics problem-solving category is 16 students (53.33%). The number of students who have sufficient mathematical problem-solving ability is five students (16.67%). Three students were described at a low level of mathematical problem-solving skills that represented 10%.

Based on the explanation above, the test instrument developed meets the valid and reliable criteria, then viewed from the level of difficulty and differentiation (Almanasreh et al., 2019). Some questions have an insufficient level of difficulty, but in general, the test instrument has an appropriate level of difficulty to meet the difficulty level criteria. Based on distinguishing power (Gundry & Deterding, 2019). There is one item with insufficient distinguishing power, but the questions have the distinguishing power according to the criteria. Therefore the test instrument to measure the problem-solving ability reaches the final prototype.

CONCLUSION AND SUGGESTION

Based on the results of research and discussion, it can conclude that the problem grain quality has been valid 100%, with the characteristics of the reality shows at 0.636 in a reasonably good category. Besides, the differentiator power for three items each has a lousy grade (33.3%), enough (33.3%), and good (33.3%). Meanwhile, a grain index of problem-solving skills tests shows easy categories and two medium category items. Besides, the number of students included in having a very high mathematical problem-solving ability is six students (20%).

The students who are in the high category were represented in 53.33%. The number of students who have sufficient mathematical problem-solving ability is five students (16.67%). Students with a low level of mathematical problem-solving skills are three students (10%).

Advice for further research is to analyze the problem quantitatively to know the effectiveness of the test itself to improve test quality. Besides, researchers need to analyze questions quantitatively to find out how practical the test is to improve the test's quality. Also, in making the questions, it should not be too difficult and not too easy to be balanced. Furthermore, after analyzing quantitatively, if others encounter items that are not functioning correctly, it should be corrected, and those that are already functioning correctly can be used as references for future tests.

REFERENCES

- Aida, N., Kusaeri, K., & Hamdani, S. (2017). Karakteristik Instrumen Penilaian Hasil Belajar Matematika Ranah Kognitif yang Dikembangkan Mengacu pada Model PISA. *Suska Journal of Mathematics Education*, 3(2), 130. <https://doi.org/10.24014/sjme.v3i2.3897>
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214–221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Amam, A. (2017). Penilaian Kemampuan Pemecahan Masalah Matematis Siswa Smp. *Teorema*, 2(1), 39. <https://doi.org/10.25157/.v2i1.765>

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

- Angriani, A. D., Nursalam, N., Fuadah, N., & Baharuddin, B. (2018). Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Pemecahan Masalah Matematika Siswa. *AULADUNA: Jurnal Pendidikan Dasar Islam*, 5(2), 211. <https://doi.org/10.24252/auladuna.v5i2a9.2018>
- Arfiana, A., & Wijaya, A. (2018). Problem solving skill of students of senior high schools and Islamic high schools in Tegal Regency in solving the problem of PISA based on Polya's stage. *Jurnal Riset Pendidikan Matematika*, 5(2), 211–222. <https://doi.org/10.21831/jrpm.v5i2.15783>
- Arifin, Z. (2017). Kriteria Instrumen Dalam Suatu Penelitian. *Jurnal Theorems (the Original Research of Mathematics)*, 2(1), 28–36.
- Baştürk, S. (2016). Investigating the effectiveness of microteaching in mathematics of primary pre-service teachers. *Journal of Education and Training Studies*, 4(5), 239–249. <https://doi.org/10.11114/jets.v4i5.1509>
- Bintarti, S., & Kurniawan, E. N. (2017). A study of revisit intention: Experiential quality and image of Muara Beting tourism site in Bekasi District. *European Research Studies Journal*, 20(2), 521–537. <https://doi.org/10.35808/ersj/657>
- Brundle, C., Heaven, A., Brown, L., Teale, E., Young, J., West, R., & Clegg, A. (2019). Convergent validity of the electronic frailty index. *Age and Ageing*, 48(1), 152–156. <https://doi.org/10.1093/ageing/afy162>
- Cholily, Y.M., Kamil, T.R., & Kusgiarohmah, P. A. (2020). Secondary school students' error of term of algebraic forms based on mathematical communication. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(2), 252–258.
- Dewi, S. S., Hariastuti, R. M., & Utami, A. U. (2019). Analisis Tingkat Kesukaran Dan Daya Pembeda Soal Olimpiade Matematika (Omi) Tingkat Smp Tahun 2018. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, 3(1), 15–26. <https://doi.org/10.36526/tr.v3i1.388>
- Gundry, D., & Deterding, S. (2019). Validity Threats in Quantitative Data Collection With Games: A Narrative Survey. *Simulation and Gaming*, 50(3), 302–328. <https://doi.org/10.1177/1046878118805515>
- Hayati, S., & Lailatussaadah, L. (2016). Validitas Dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif Dan Menyenangkan (Pakem) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, 16(2), 169. <https://doi.org/10.22373/jid.v16i2.593>
- Hidayat, M. L., Prasetyo, W. H., & Wantoro, J. (2019). Pre-service student teachers' perception of using google classroom in a blended course. *Humanities and Social Sciences Reviews*, 7(2), 363–368. <https://doi.org/10.18510/hssr.2019.7242>
- Lee, C. I. (2017). An appropriate prompts system based on the Polya method for mathematical problem-

DOI: <https://doi.org/10.24127/ajpm.v9i4.3036>

- solving. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(3), 893–910.
<https://doi.org/10.12973/eurasia.2017.00649a>
- Martin, F., Stamper, B., & Flowers, C. (2020). Examining student perception of readiness for online learning: Importance and confidence. *Online Learning Journal*, 24(2), 38–58.
<https://doi.org/10.24059/olj.v24i2.2053>
- Mohammed, L. A. (2019). Validity and reliability. In *Forensic Examination of Signatures*. Elsevier Inc.
<https://doi.org/10.1016/B978-0-12-813029-2.00009-8>
- Priyambodo, E., & Marfuatun, M. (2016). Validity and Realibility of Chemistry Systemic Multiple Choices Questions (CSMCQs). *International Journal of Evaluation and Research in Education (IJERE)*, 5(4), 306.
<https://doi.org/10.11591/ijere.v5i4.5958>
- Putri, D. Y., & Sutarto, S. (2018). The effect of industrial work practice, guidance intensity of industrial side, and vocational competence on working readiness of grade XII students of banking program, Vocational High School Perbankan, Pekanbaru. *Jurnal Pendidikan Vokasi*, 8(2), 132.
<https://doi.org/10.21831/jpv.v8i2.18908>
- Sumartini, T. S. (2018). Peningkatan Kemampuan Pemecahan Masalah Matematis Siswa melalui Pembelajaran Berbasis Masalah. *Mosharafa: Jurnal Pendidikan Matematika*, 5(2), 148–158.
<https://doi.org/10.31980/mosharafa.v5i2.270>
- Supardi, R., Istiyono, E., & Setialaksana, W. (2019). Developing Scientific Attitudes Instrument of Students in Chemistry. *Journal of Physics: Conference Series*, 1233(1).
<https://doi.org/10.1088/1742-6596/1233/1/012025>
- Susanti, S., Musdi, E., & Syarifuddin, H. (2017). Pengembangan Perangkat Pembelajaran Matematika Berbasis Penemuan Terbimbing untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Materi Statistika. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 1(2), 305.
<https://doi.org/10.33603/jnpm.v1i2.561>
- Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R., & Stuart, E. A. (2019). Target Validity and the Hierarchy of Study Designs. *American Journal of Epidemiology*, 188(2), 438–443.
<https://doi.org/10.1093/aje/kwy228>