JERE 9 (1) 2020 1-7

Journal of Educational Research and Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Validity of Content and Reliability of Inter-Rater Instruments Assessing Ability of Problem Solving

Aini Azkiyatu Ulfah✉, **Kartono Kartono, Endang Susilaningsih**

Universitas Negeri Semarang, Indonesia

Article Info

Published 31 March 2020

Keywords:
content validity,
interrater reliability,
ability to solve mathematical problems.

Abstract

In general, problems that occur in the actual are the discovery of instruments that have not been tested for problem solving abilities. The aim of this research is to reveal the content validity and interrater reliability of the instrument for evaluating the problem solving abilities that have been prepared. The research method is used a quantitative description by 3 expert judgments, they are experts in research and evaluation, mathematics education experts, and mathematics teachers. The instrument was developed in the form of an expert observation sheet with 3 aspects of assessment, they are the aspect of content eligibility, construction aspects, and language aspects, which of each aspect has 4 categories, which are very relevant, relevant, quite relevant, and highly irrelevant. Data were analyzed using the Aiken's V formula to determine the level of instrument validity and to determine the level of consistency / constancy between assessors using Instraclass Correlation Coefficient (ICC) analysis with the help of SPSS version 23.0. the results of analysis of the content validity of all items valued above 0.3 which means that all aspects assessed by experts are valid. Interrater reliability test using ICC obtained a value of 0.516, which means that all aspects of the instrument for evaluating problem solving abilities that have been rated have a level of consistency. Thus, the instrument of problem-solving ability that has been tested for validity and reliability can be used by educators to determine the level of students' problem solving abilities appropriately.

✉Correspondence Address :
Kelud Utara 3 kampus Pascasarjana UNNES Semarang, Indonesia
E-mail : ainiazkiya0@gmail.com

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

Assessment according to (Mardapi, 2016: 12) is the process of gathering and processing information to measure the achievement of learning outcomes. In line with the opinion of Anderson (2003) which states that assessment is a process of gathering information that is used to make accountable decisions.

According to Zurqoni (2017: 264) through an assessment can be identified strengths and weaknesses of an institution or institute, besides that it also serves as a basis for determining decisions related to programs that need to be realized, as well as a basis for providing feedback on things that need to be improved . The information can be used as a reference for making decisions about learning, student difficulties and the guidance efforts needed by students.

Assessment instruments commonly used by educators to assess student learning outcomes are by giving tests (Mu'awanah, 2015: 133). Tests are assessments that measure student learning (Roland, 2015: 189). Need to have good test instrument is to do measurements. Ramadhan, et al (2020: 508) explained that a good instrument is an instrument that has high validity and reliability and has the smallest possible error in capturing information.

However, the problem that often occurred in the field is the discovery of instruments that are not yet valid and reliable, but are still used for measurement (Sugiharni, 2017: 679). Meanwhile, to get a good instrument, it needs to be tested for validity and reliability. The test is proved to be valid if the test has measured the actual abilities possessed by students through learning activities (Ramadan, 2019)

The results of observations and interviews conducted by researchers with mathematics teachers at SMP Negeri 3 Lakbok Ciamis Regency, Mr. Dedi Suhardiman explained that instruments and assessment rubrics are needed in the process of learning mathematics, especially in the aspect of problem-solving ability. It caused the assessment of problem-solving ability must be more objective, while the ability of students to solve different problems causes the answer to the problem-solving work so that it is difficult for teachers to assess objectively. Therefore, the right instrument is needed to make the right measurements (Kurniawan, Reyza, & Taqwa, 2018: 1451). Because, the right instrument can minimize errors from measurement (Sinaga, 2016: 171)

Regarding with the description above, an instrument for evaluating the ability to solve SMP problems is important to test the validity and reliability of the assessment instruments. The formula used in the content validity test is the Aiken's V validity index because the instrument was tested or assessed by 3 experts. Whereas the interrater reliability test uses the analysis of the Interclass Correlation Coefficient (ICC) analysis.

METHODS

This research is a part of development research in mathematics learning. The research method is descriptive quantitative because it related to the numbers obtained from the results of the test of content validity and interrater reliability. The instrument validity data was obtained by giving questionnaires to experts. The experts chosen to provide the assessment are 3 experts with different backgrounds, there are research and evaluation experts, mathematics education experts, and mathematics teachers. There are 8 items that were assessed by reviewing 3 aspects of assessment, there are aspects of content eligibility, construction aspects, and language aspects. The results of the assessment of the three experts were then analyzed using the Aiken's V formula, while interrater reliability was analyzed using the Interclass Correlation Coefficient (ICC) analysis in order to gain the value of content validity and interrater reliability assessment instruments for problem solving abilities.

RESULTS AND DISCUSSION

Assessment is the process of gathering and processing information to measure the achievement of student learning outcomes (Hanifah, 2019: 5). Assessment is not tied to the student's assessment characteristics, but also concerns the characteristics of teaching methods, curriculum, facilities and school administration. Assessment instruments intended for students can be either formal or informal methods or procedures to produce information about student progress. The results of evaluations conducted by teachers can also be used by students to support their learning success (Mumu & Tanujaya, 2019: 86)

Development of an instrument for evaluating the ability to solve problems in Mathematics is the development of instruments or measuring instruments used to measure students' problem-solving abilities during class learning. Problem solving ability is the student's attempt to find a solution or answer to a problem. Problem solving ability is a variable that can be observed so that the measuring instrument that is often used is the test instrument.

The preparation of test instruments must be arranged logically and rationally about the main points of what material should be asked as important material for knowledge to be known and understood by students. Not only that, tests made by teachers need to pay attention to the level of difficulty of the items based on the nature or characteristics of students. The tests also need to be tested on large groups (Alam, Japar, & Asnur, 2019: 61).

Mathematical problem-solving ability tests are based on material indicators and indicators of problem-solving abilities in the form of story-based descriptions consisting of 8 items with different levels of difficulty, the goal is that the items can measure accurately the level of students' problem-solving abilities. Before being distributed, the tests were first tested for validity and reliability by the researcher by giving a test assessment questionnaire to 3 experts.

The researcher asked 3 experts in accordance with the field of research, there are research and evaluation experts, mathematics education experts, and mathematics teachers. The three validators were asked to provide an assessment related to the instrument of problem-solving abilities that had been made by researchers.

The validator is given a validation sheet that the researcher has provided for further assess toward instrument being developed. The content validity test was carried out to check the validity of the instrument in terms of 3 aspects, including: the content feasibility aspect, the construction aspect, and the language aspect. The appraiser is enough to give a check mark on each column that has been made by the researcher. There are 4 categories in each aspect, there are (4) very relevant, (3) relevant, (2) quite relevant, and (1) very irrelevant.

Conformance Scores obtained from the results of expert assessments in the form of scoring rubrics then analyzed using the Aiken's V formula to determine whether the developed instrument is proper or not proper to use. The instrument can be categorized as valid if it meets the minimum value requirements specified in Aiken's V. The results of the analysis and analysis of the three experts provide conclusions for each item being usable without revision, can be used with few revisions, can be used with many revisions, and are less suitable for use and must be fixed. The following is the Aiken's V formula used (Ramadhan, Mardapi, Prasetyo, & Utomo, 2019; Ramdhan, Sahabuddin, & Sumiharsono, 2019) :

$$V = \frac{\Sigma s}{n\,(c-1)}$$

Information

S = r - lo

lo = Lowest validity assessment number (Score = 1)

C = Highest validity assessment rate (Score = 4)

R = Number given by an assessors

N = Number of assessors

Checking items that are valid or invalid can be actualized by looking at the criteria determined based on the coefficient of validity ≥ 0.3, the item is valid if the value of validity is 3 0.3 (Azwar, 2014: 134). Expert judgment provides an assessment of the instrument to be tested for the level of validity of its contents. The results of expert judgment are then calculated using the Aiken equation. The following is a table of the results of the calculation of content validation using the Aiken V formula helped by the Microsoft Excel program.

Table 1. Expert Agreement Coefficients

Items	Rater 1	Rater 2	Rater 3	ΣS	Coefisien Aiken	Criteria
1	4	4	4	12	1.0	Valid
2	3	3	4	10	0.8	Valid
3	4	4	4	12	1.0	Valid
4	4	3	3	10	0.8	Valid
5	4	4	4	12	1.0	Valid
6	4	4	4	12	1.0	Valid
7	4	3	3	10	0.8	Valid
8	3	4	3	10	0.8	Valid

Content validity assessment is carried out by expert judgments namely evaluation lecturers, mathematics lecturers and mathematics teachers. The expert conducts an assessment of two main points. First, assessing whether the blue print made shows that the blue print classification has represented the substance to be measured, namely the problem-solving ability. Second, the experts assess whether each item that has been compiled is relevant to the specified blue print classification (Hikmah, et al. 2017: 44).

Based on the content validation calculation using the Aiken V formula, it can be concluded that all items show a valid category, with the lowest index of 0.8 and the highest index of 1. If the index of the agreement is less than 0.4 then the validity is low and if more than 0.8 meant to be very high (Guilford, 1956). This is also in line with the explanation according to Azwar (2014: 143) that an item is proved to be valid if it meets the criteria of the validity coefficient value ≥ 0.3.

In addition to quantitative data, qualitative data were also obtained from expert validators in the form of suggestions and input that became a reference for improvement in the development of instruments for evaluating mathematical problem-solving abilities. The items that have been assessed by experts are then revised by taking into account the suggestions given so that the item is more able to measure students' problem-solving abilities. Suggestions given by the validator on the developed problem-solving ability instrument can be seen in Table 2.

Table 2. Assessment Results and Comments of Validators

Experts	Comments/Suggestion
Prof. Dr. Totok Sumaryanto F, M.Pd (Educational Research and Evaluation Expert)	It is accordance with the rules of instrument development
Hadi Kusmanto, M.Si (Expert in mathematics education)	Generally is good. It would be better if the problem was realistic rather than imaginary. Example "Pak zaenal has a rectangular garden with an area of 40 m2", is that true in the real world? Then for the scoring

	guidelines there is no total score, and conversions (if there is a score conversion).
Saefur, S.Pd (Math Teacher)	The language used for question No. 2 need to improve in terms of language, so that students do not misinterpret the question.

Regarding to suggestion from experts, researchers made improvements to the items in accordance with the advice given by experts, there are in terms of language and the realistic character of the items. In addition, the researcher also changed the scoring or grading rubric according to the level of difficulty of each item and added a conversion score for the whole item in order to get the final score of the mathematical problem-solving ability test.

Next, the researchers conducted an interrater reliability test to determine the level of expert agreement on the instrument of problem-solving ability assessment using the Intraclass Correlation Coefficient (ICC) analysis assisted by SPSS 23.0. Interpretation of high and low reliability coefficients can be seen from the reliability coefficient whose value is in the range 0 to 1.00. If the coefficient value is higher, close to 1.00, the higher the level of agreement. Conversely, the closer to 0, the lower the level of agreement. The following are the results of the calculation of reliability using SPSS 23.0:

Table 3. Reliability Statistics

Cronbach's Alpha	N of Items
.516	3

Based on the table above, it can be concluded that the results of the interrater reliability test calculations with 3 experts and analyzed using the Intraclass Correlation Coefficient (ICC) obtained the results of the assessment with the reliability coefficient rxy = 0.516 and has a Cronbach alpha value between 0.41 $< r_{count} \leq 0.60$ which means the instrument has moderate reliability. Stainer, et al stated that measuring instruments have adequate stability if inter-gauge ICC> 0.50, high stability if inter-measuring ICC> 0.80. (Streiner and Norman: 2000; Polgar and Thomas, 2000). Based on this classification, a reliability of 0.516 has adequate stability.

Thus, the instrument for evaluating problem solving abilities that have been tested for validity and reliability by 3 experts using the Aiken V formula and reliability using Intraclass Correlation Coefficient (ICC) analysis results are valid and reliable.

CONCLUSION

The test instrument and the rubric of the assessment of problem-solving abilities were declared proper based on the value of content validity that has been reviewed by 3 expert judgments and analyzed using the Aiken V formula with results above 0.3. The test instrument was also stated to be quite reliable based on an analysis using the Intraclass Correlation Coefficient (ICC) assisted by SPSS 23.0 with a reliability value of 0.516. Suggestions for future researchers who wants to conduct similar research is to increase the expert judgment / expert, because the more experts who judge the better the quality of the instrument from the aspect of content validity. The researcher would like to thank Mr. Prof. Dr. Totok Sumaryanto F, M.Pd, Mr. Hadi Kusmanto, M.Sc, and Mr. Saefur, S.Pd as

experts who provided an assessment of the instrument of problem solving ability compiled.

REFERENCES

Alam, S., Japar, M., & Asnur, M. N. A. (2019). Pengembangan Instrumen Tes Siswa Tingkat Sekolah Dasar di Kabupaten Kuningan. Qalam : Jurnal Ilmu Kependidikan, 8(1), 59.

Anderson, L. W. (2003). Classroom Assessment: Enhanching the Quality of Teacher Decision Making. London: Lawrence Erlbaum Associates Publisher.

Antonio Roland I. CoPo. (2015). Students' Initial Knowledge State And Test Design: Towards A Valid And Reliable Test Instrument. Journal of College Teaching & Learning – Third Quarter. Volume 12, Number 4

Azwar, S. (2015). Reliabilitas dan Validitas. yogyakarta: Pustaka Pelajar.

Guilford, J. P. (1956). Fundamental Statistics in Psychology and Education. New York: Mc Graw-Hill Book Co. Inc

Gusti Ayu Dessy Sugiharni. (2017). Validitas Isi Instrument Pengujian Modul Digital Matematika Diskrit Berbasis Open Source Di STIKOM Bali. Konferensi Nasional System & Informatika.

Hikmah, Sri Yamtinah, Ashadi, Nurma Yunita Indriyanti. (2017). Analisis Validitas Isi Instrumen Computerized Two-Tier Multiple Choice (CTTMC) Untuk Mengukur Keterampilan Proses Sains Pada Materi Termokimianaili SNPS (Prosiding Seminar Nasional Pendidikan Sains).

Jeinne Mumu & Benidiktus Tanujaya. (2019). Measure Reasoning Skill of Mathematics Students. International Journal of Higher Education. Vol. 8, No. 6

Kurniawan, B. R., Reyza, M., & Taqwa, A. (2018). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Fisika pada Materi Listrik Dinamis. Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan, 3(11), 1451–1457.

Mardapi, D. (2016). Pengukuran, Penilaian dan Evaluasi Pendidikan. Yogyakarta: Nuha Medika.

Mu'awanah, S. (2015). Pengembangan Instrumen Penilaian Problem Solving Pada Materi Larutan Elektrolit Dan Nonelektrolit. 2407–4659, 132–140.

Nurdinah Hanifah. (2019). Pengembangan instrumen penilaian Higher Order Thinking Skill (HOTS) di sekolah dasar . Vol. 1 No. 1.

Permendikbud Nomor 23 Tahun 2016 Tentang Standar Penilaian Pendidikan.

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in Physics. European Journal of Educational Research, 8(3), 743-751.

Ramadhan, S., Mardapi, D., Sahabuddin, C., & Sumiharsono, R. (2019). The estimation of standard error measurement of Physics final examination at senior high schools in Bima Regency Indonesia. Universal Journal of Educational Research, 7(7), 1590-1594

Sinaga, N. A. (2016). Pengembangan tes kemampuan pemecahan masalah dan penalaran matematika siswa SMP kelas VIII. PYTHAGORAS: Jurnal Pendidikan Matematika, 11(2), 169.

Streiner DL, Norman GR (2000). Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press.

Soleh, A., Khumaedi, M., & Pramono, S. E. (2017). Pengembangan Instrumen Penilaian Mata Pelajaran PKn Standar Kompetensi Memahami Kedaulatan Rakyat Dalam Sistem Pemerintahan di Indonesia. Journal of Educational Research and Evaluation, 6(1), 71–80.

Syahrul Ramadhan, Rudy Sumiharsono, Djemari Mardapi, Zuhdan Kun Prasetyo. (2020). The Quality of Test

Instruments Constructed by Teachers in Bima Regency, Indonesia: Document Analysis. International Journal of Instruction. Vol.13, No.2 e-ISSN: 1308-1470.

Zarqoni. (2017). The Development of Learning Services Assessment Instrument Model of Islamic Higher Education. Vol. 17 No. 2. P-ISSN: 1411-3031.