Validity and Reliability Content of Physics Problem Solving Test Instrument Based on Local Wisdom

Dwiana Octafia ✉, Supriyadi Supriyadi, Sulhadi Sulhadi

Universitas Negeri Semarang, Indonesia

Info article	Abstract
	This research was part of the research development a instrument test for the ability to solve physics problems based on problems that occur in the field, that is an instrument for assessing the ability to solve problems that has not been tested. The purpose of this study was to reveal the validity and reliability of the contents of the physics problem solving ability test instrument that had previously been compiled. The research method used is a quantitative description by 5 expert judgments. The instrument developed in the form of an expert observation sheet with 3 aspects of assessment, namely aspects of content feasibility, aspects of construction and aspects of language. Analysis of the content validity of the observation sheet using the V coefficient by Aiken and the reliability of the instrument content using the Inter-class Correlation Coefficient (ICC) analysis with the help of SPSS version 16.0. Based on the results of the study, it shows valid results with all item items valued> 0.3, those are with the lowest index 0.6 and the highest 1.0. Inter-rater reliability test using ICC obtained a value of 0.630, which means that all aspects of the instrument for assessing the ability of physics problem solving that have been assessed have a level of consistency. Based on the results of these studies, the test instrument for the ability to solve physics problems is feasible to use.

✉Correspondence address :
  Kelud Utara 3 kampus Pascasarjana UNNES Semarang, Indonesia
  E-mail : dwianaoctafia3@gmail.com

INTRODUCTION

UNESCO determines acompetencies for students in the 21st century, one of which is the ability to think critically and do problem solving (Sani, 2014: 24). There are four groups of skills that are important to master in the 21st century, there are ways of thinking, ways for working, tools for working, and living in the world (Barak, 2017: 284). In this research, the emphasis is on ways of thinking. Physics learning and problem solving skills are an important part of the learning process.

Problem solving abilities are abilities that must be possessed by students. This is emphasized in the Partnership for 21st Century Skills (2017) in Kurniawan (2018: 1451) which stated that problem solving ability is one of the abilities that needs to be mastered in order to be successful in life and work. Polya (1945) in Robson (1945: 7) stated that problem solving ability is an attempt to find a way out of a goal that is not so easy to achieve immediately using the power and benefits of mathematics. Problem solving is one of the key factors in learning physics in addition to mastery of concepts (Riantoni et al., 2017: 41). Physics concepts that are owned or have been understood will not be mastered properly without practicing questions and solving problems related to the subject being studied (Silaban, 2014: 68).

The results of preliminary studies through interviews with several teachers who teach physics subjects revealed that students still have difficulty when given problem solving questions so they need to be guided first, even repeating some material to understand the concept. The test used to measure problem-solving abilities also still uses the questions in the books used in classroom learning such as textbooks and student worksheets (student worksheets). The test used was also developed again from the questions that already exist in the book, then given to students without being tested first, and the scoring used was still estimated according to the weight of the answer.

The number of research that examines problem-solving abilities, of course, requires instruments that can be used to measure problem-solving abilities (Kurniawan et al., 2016: 1451). The test results can also be used to monitor the development of education quality with the smallest possible measurement error (Mardapi, 2017: 94). The right instrument is needed to make the right measurements in order to minimize the error of measurement results (Sinaga, 2016: 171). Therefore, a test instrument was developed in the form of an essay test to measure students' problem solving abilities.

Before do implementation, a test must be analyzed in advance to determine its quality. The instrument used in the test has several requirements to be fit for use (valid and reliable). Therefore, the previously developed local wisdom based physics problem solving ability test instrument must meet these two requirements. It is necessary to test the validity and reliability. Lawrence (1994) in Retnawati (2016: 17) stated that content validity is a representation of the question of the ability to be measured so that the content validity is related to the domain to be measured. Testing the validity of the test instrument for the ability to solve physics problems based on local wisdom that will be used is the content validity test. To find out the content validity, questions are validated by experts, there are lecturers and teachers who have competence in the fields of development and physics (Kereh et al., 2015: 41).

The reliability of a test is related to the strength of the test to measure skills or the achievement of what will be measured (Mardapi, 2017: 46). Reliability is related to the consistency of learner performance (Gipps, 1994). Reliability is also related to measurement error which is indicated by the greater the reliability of an instrument, the smaller the error in measurement, on the other hand if the smaller the reliability of an instrument, the greater the measurement error. Measurement errors can be caused by several factors, there are the characteristics of the instrument used, the quality of the items in the instrument is not good, the existence of cooperation in carrying out the test, the condition of the participants while responding to the test,

and the environment in which the test is held is not supportive (Retnawati, 2016: 85).

In line with the description above, it is important that the instrument for assessing the problem solving ability of student is tested for validity and reliability of the assessment instrument. The formula used in the content validity test is the Aiken's V validity index because the instrument was tested or assessed by 5 experts. Meanwhile, the inter-rater reliability test used the Inter-class Correlation Coefficient (ICC) analysis.

METHODS

This research is part of the research development on physics learning. The research method used is descriptive quantitative because it relates to the numbers obtained from the results of the content validity and reliability tests. The instrument validity data was obtained by giving a questionnaire to the expert. The experts chosen to give the assessment were 5 experts with different backgrounds, namely 2 research and evaluation experts, physics education expert, and 2 high school physics teachers. There are 8 items that were assessed by reviewing 3 aspects of the assessment, namely aspects of content feasibility, aspects of construction, and aspects of language. The results of the three experts' assessment were then analyzed using the Aiken's V formula, while the content reliability was analyzed using the Instraclass Correlation Coefficient (ICC) analysis. ICC (Inter-class Correlation Coefficient) is a reliability that is used for rating results from observations of several raters (Shrout & Fleiss, 1979: 422). Based on these calculations, the value of content validity and inter-rater reliability of the problem-solving ability assessment instrument can be obtained.

(Aiken, 1985: 133) formulated a formula for calculating the content-validity coefficient which is based on the results of an expert's assessment of n people on an item regarding the extent to which the existing items can represent the construct being measured or commonly referred to as Aiken's V. 1 to 5, the number 1 to represent the assessment is irrelevant until 5

represents the most relevant (Azwar, 2015: 112-113). Aiken-V is formulated as follows:

$$V = \frac{\sum(r - l_o)}{[n(c - 1)]}$$

Notes:
l_o : Very low number of validity assessments (in this case 1)
c : The highest number of validity assessments (in this case 5)
r : Number is given by assessor

A high value means that the item has high content validity according to the assessor or rater. The results of the scoring of the validation sheet are then said to be valid in terms of content when checking the V index obtained by seeing the specified criteria> 0.3 (Azwar, 2015: 134). If these criteria have met, then the item is considered valid.

The reliability of the instrument content developed using the ICC approach with the help of SPSS 16.0 to calculate the level of agreement between the five Expert Judgments, before using the ICC approach, the reliability value was estimated using the Alpha coefficient. The coefficient value must be> 0.6 so that the Intraclass Correlation Coefficient (ICC) analysis can be continued (Suharsimi, 2007: 75). The interpretation of the high and low reliability coefficient can be seen from the reliability coefficient whose value is in the range of 0 to 1.00. If the coefficient value is closer to 1.00, the higher the agreement level is. Conversely, the closer to 0, the lower the level of agreement (Streiner & Norman, 2000: 156).

RESULTS AND DISCUSSION

Test instrument was developed in the form of an essay test to measure students' problem solving abilities. The problem-solving ability applied to physics learning can not only measure aspects of knowledge but also know the understanding of students' physics concepts. Before do implementation, a test must be analyzed in advance to determine its quality. The instrument used in the test has several requirements to be fit for use (valid and reliable). Therefore, the previously developed local

wisdom based physics problem solving ability test instrument must meet these two requirements. It is necessary to test the validity and reliability.

The validity assessment is carried out by the expert to provide an assessment of the instrument. Assessor are 5 experts with different backgrounds, there were 2 research and evaluation experts, physics education expert, and 2 high school physics teachers. Experts make an assessment of two main things. First, assessed whether the grid is made showing that the grid classification represents the substance to be measured, namely problem solving. Second, the experts assessed whether each item that has been arranged is relevant to the classification grid that has been determined. Analysis of the content validity of the instrument used the Aiken coefficient V, where V index> 0.3. The following is Table 1 of the results of the content validity test using the Aiken'V formula.

Table 1. Content Validity Results using Aiken'V

No. Butir	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	V indeks	Kesimpulan
1	3	3	3	3	3	1.0	Valid
2	3	2	2	2	2	0.6	Valid
3	3	3	3	3	3	1.0	Valid
4	3	3	3	2	2	0.8	Valid
5	3	3	3	3	3	1.0	Valid
6	3	3	3	3	3	1.0	Valid
7	3	3	3	3	3	1.0	Valid
8	3	3	3	3	3	1.0	Valid

(Source: Researcher Data, 2020)

In Table 1, the results of the Aiken'V analysis produced all items that are declared valid with the lowest index 0.6 and the highest index of 1.0. This is in accordance with the criteria stated by Azwar (2015: 34) if the validity coefficient ≥ 0.3 means the item is declared valid. The same thing is explained by (Guilford, 1956: 113) if the agreement index is less than 0.4 then it has low validity and if it is more than 0.8 then the validity is very high.

The results of expert validation also showed qualitative data from the assessor in the form of suggestions and input which became improvements in the development of an instrument of physics problem solving ability. The results that have been assessed by experts are then revised by paying attention to the suggestions given so that the items are more biased to measure students' physics problem solving abilities. The suggestions given by the assessor on the instrument can be seen in Table 2.

Table 2. Assessment Results and Suggestions from Assessor

Expert Name	Comments/Suggestion
Expert 1	Can be used to retrieve data
Expert 2	The RPP Assessment Rubric and so on are written in a uniform / relevant level and the number of words is the same or close to the same
Expert 3	In substance, the instrument "test the ability of physics problem solving based on local wisdom of the Dayak Tomun community" is suitable for use with minor revisions
Expert 4	
Expert 5	Revise the narrative structure so as not to create multiple interpretations. The arrangement / order in arranging paragraphs or sentences is improved.

(Source: Researcher Data, 2020)

Suggestion that is provided by the expert is used as a reference for improving the test instrument before continuing to be tested on a small and large scale. The improvements made by the researcher were on the item points, in terms of language and question construction. In addition, researchers also changed the instrument's lattice in terms of language.

Reliability test between assessor is to use ICC (Inter-class Correlation Coefficient) using SPSS 16.0 to estimate inter-rater reliability by showing the comparison between variations caused by the attributes measured with the overall measurement variation. The results of the content reliability calculation using SPSS 16 can be seen in Table 3.

Table 3. Reliability Test Results of 5 Experts

Cronbach's Alpha	N of Items
.895	5

(Source: Researcher Data, 2020)

The ICC analysis results obtained a reliability of 0.675. According to Rusilowati (2014: 29), the reliability criteria are considered moderate in the value range $0.4 \leq r < 0.6$. Reliability is considered high if it is in the range $0.6 \leq r < 0.8$. In line with the opinion (Khumaedi, 2012: 29), which states that the reliability coefficient > 0.5 is quite accepted as good reliability. Based on Cronbach's Alpha analysis, it shows that the results of the agreement between the raters, the coefficient of $r_{xx} = 0.874$, means that the coefficient of r_{xx} is > 0.7 so that it can be continued with the ICC (Inter-class Correlation Coefficient) analysis with the help of SPSS 16. Following are the results of the ICC test for agreement between raters in Table 4.

Table 4. Inter-class Correlation Coefficient Inter-Rater Agreement

Inter-class Correlation Coefficient	
	Inter-class Correlation
Single Measures	.630[b]
Average Measures	.895[c]

(Source: Researcher Data, 2020)

According to (Suharsimi, 2007: 75), the assessment instrument is said to be reliable if rxx > 0.6. Based on the ICC output results above, seen from Single Measures = 0.630. In line with Stainer's opinion that the measuring instrument has adequate stability if the ICC between gauges > 0.50, high stability if the ICC between gauges > 0.80 (Streiner & Norman, 2000: 156). Based on this classification, the reliability of 0.630 is included in adequate reliability and is ready for field testing.

Thus, the problem solving ability test instrument that has been tested for validity and reliability by 5 experts uses the Aiken V formula and reliability using Inter-class Correlation Coefficient (ICC) analysis, the results are valid and reliable.

CONCLUSION

The test instrument and problem solving ability assessment rubric were declared feasible based on the content validity value that was reviewed by 5 experts (expert judgment) and analyzed with the Aiken V formula with results above 0.3, namely with the lowest index of 0.6 and the highest index of 1.0. it shows that the instrument items developed are valid.

The test instrument was also declared quite reliable based on the analysis using the Inter-class

Correlation Coefficient (ICC) assisted by SPSS 16.0 with a reliability value of 0.630. Thus the instrument for testing the ability to solve physics problems based on local wisdom that was developed has a fairly high agreement between the rater.

ACKNOWLEDGMENT

REFERENCES

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *5(1)*, 131–142.

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

Barak, M. (2017). Science Teacher Education in the Twenty-First Century: a Pedagogical Framework for Technology-Integrated Social Constructivism. *Research in Science Education*, *47*, no.*2*, 283–303.

Gipps, C. V. (1994). *Beyond Testing: Educational, Towards a Theory of Falmer, Assessment*. Washington DC: The Press.

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

Kereh, C. T., Liliasari, Tjiang, P. C., & Sabandar, J. (2015). Validitas Dan Reliabilitas Instrumen Tes Matematika Dasar Yang Berkaitan Dengan Pendahuluan Fisika Inti. *Jurnal Inovasi Dan Pembelajaran Fisika*, *2(1)*, 36–46.

Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin*, *12(1)*, 29.

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

Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan* (2/April 20). Yogyakarta: Parama Publishing.

Retnawati, H. (2016). *Validitas, Reliabilitas, dan Karakteristik Butir* (Pertama). Yogyakarta: Parama Publishing.

Riantoni, C., Yuliati, L., & Mufti, N. (2017). Kemampuan Pemecahan Masalah Materi Listrik Dinamis pada Pembelajaran Guided Inquiry Berbantuan PhET pada Mahasiswa S1 Pendidikan Fisika. *Jurnal Riset Dan Kajian Pendidikan Fisika*, *4*(1), 40.

Robson, A. (1945). Reviewed Work: How to Solve It by G. Pólya. *The Mathematical Gazette*, Vol. 30, p. 181.

Rusilowati, A. (2014). *Pengembangan Instrumen Penilaian*. Semarang: Unnes Press.

Sani, R. A. (2014). *Pembelajaran Saintifik untuk Implementasi Kurikulum 2013*. Jakarta: Bumi Aksara.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses Psychological, assessing rater reliability. *Bulletin*, *86(2)*, 420–428.

Silaban, B. (2014). Hubungan Antara Penguasaan Konsep Fisika Dan Kreativitas Dengan Kemampuan Memecahkan Masalah Pada Materi Pokok Listrik Statis. *Penelitian Bidang Pendidikan*, *20*(1), 65–75.

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

Streiner, D., & GR, N. (2000). *Health measurement scales: A practical guide to their development and use*. Oxford: Oxford University Press.

Suharsimi, A. (2007). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.