JURNAL PEMBELAJARAN FISIKA

http://jurnal.fkip.unila.ac.id/index.php/JPF

Vol 7 (2), 2019, 87-94 ISSN: 2302-0105 (p); 2684-9828 (e)

Development of Assessment Instruments for Physics Problem Solving Skill Based on Polya's Stages

Johan Syahbrudin

Department of Informatics Engineering, Pamulang University, Tangerang Selatan, Banten, Indonesia

* *e-mail: dosen01263@unpam.ac.id*

Received: November 25, 2019 *Accepted: December 30, 2019* *Published: December 31, 2019*

Abstract: This study aims to develop an instrument that measures the physics problem solving skill on the subject of school gas kinetic theory based on Polya's stages. This research is a research & development using a modified 4D development model. Internal trials were conducted on 2 experts and 1 learning practitioner, while empirical trials were conducted on 81 students. Internal trial data were obtained through questionnaires and analyzed using descriptive analysis techniques, while empirical test results were obtained through test techniques and analyzed following the Rasch Model (RM) using the QUEST program. The results showed that at the internal test stage all items were stated to be very good in measuring the physics problems solving skill. Empirical test results, as many as 5 out of 5 items were declared fit, with instrument reliability of 0.63 (high category). The power of distinguishing all items is included in either category. So that it meets the requirements of a good instrument. But the level of difficulty of the questions varies less, where 1 item is in the easy category, 4 other items are in the difficult category, and not in the moderate category.

Keywords: Assessment instruments, Physics problem solving skill, Polya's stages

DOI: http://dx.doi.org/10.23960/jpf.v7.n2.201905

INTRODUCTION

The ability to solve problems is one of the higher order thinking skills that is very important to develop. Because the development of high quality human resources who are skilled in the thought process needs to be improved in every aspect that is in line with the 21st century conceptual framework for education (Boonjeam et al., 2017).

There are four indicators of problem solving ability referring to the effort needed by students in determining problem solutions that are adapted based on the stages of problem solving according to (Polya, 1957), namely understanding the problem, making a problem solving plan, implementing the plan, and re-examining the results obtained. The four stages of problem solving from Polya are very important units to be developed. Understanding the problem (reading the problem) is of course not only reading, but also digesting the problem presented and understanding what is happening. In other words, understanding the problem / reading the problem is the activity of identifying what is asked to be solved and the facts provided. Activities make plans, problem solvers find the relationship between the data provided (known) and unknown (asked). Implementing the plan is concerned with examining each stage of the plan previously made. Check activities related to the truth / certainty of the solution obtained (Baiduri, 2015).

But the facts on the ground show that the ability to solve physics problems is still in the sufficient category (Mustofa & Rusdiana, 2016). This is in line with the finding (Alfika & Mayasari, 2018) that the ability to solve children's problems is included in the inadequate category, with details according to Polya's stages, namely: the ability to understand problems, make plans, and implement plans into less categories, and the ability to re-examine enough categories.

So this ability still needs to be developed, one of them is through problem-based physics learning. Through this learning, educators act as guides for students in solving problems by giving guide questions, motivating, going around the classroom to facilitate discussion, and so on. And students are required to solve the problems presented. So through problem solving in learning physics, it is hoped that later it can develop problem-solving abilities not only in physics, but also in other fields of daily life.

This is in line with (Kemendikbud, 2016), where problem solving skills are very important in learning physics, so that the core competencies of physics apply procedural knowledge to solve problems. And after that, teachers must assess students' problem solving skills to determine the achievement of their competencies. This measurement will also be very important as a learning evaluation material. Because this assessment functions as a controller, control, and align the components of education, such as: goals with standard processes, results with clear achievement criteria, and so on. In addition, assessment is one important component in learning as objective feedback about what students have learned and also used to determine learning effectiveness.

The success of measurement and evaluation is of course closely related to the quality of assessment instruments. The instrument must meet the criteria of validity and reliability in measuring the quality and learning outcomes, said to be valid if the instrument can be used appropriately in measuring what is measured, and said to be reliable if it gives consistent results. results when used repeatedly (Widoyoko & Putro, 2016).

Facts in the field show, the assessment of competency achievement by students such as midterm, final semester exams, even final school exams or national exams even use multiple choice instruments. Whereas multiple choice questions used in physics assessment are rarely able to measure higher order thinking skills (Erfianti et al., 2019) including problem solving skills. In addition, multiple choice questions also allow students to guess the answer choices so that students' thinking abilities cannot be seen clearly (Istiyono et al., 2019).

And the fact that multiple choice questions that measure low-level thinking is used more than other forms of tests (Istiyono et al., 2014) becomes a challenge for researchers to develop questions that measure problem solving abilities because teachers need assessment instruments that can truly measure students 'problem solving skills, instruments to assess problem solving skills will produce accurate data about students' physics solving abilities if quality instruments, which means teachers need quality assessment tools (Asysyifa et al., 2019).

In addition, it was found that there were teachers who were still less varied in using the learning evaluation instrument (Octavia et al., 2017). So there is still a need to develop questions that are used as enrichment by applying indicators that can be used to measure higher-order thinking skills and have good test instrument characteristics to use (Julianingsih et al., 2017).

Based on the background of the problems, it is known that it is very important to measure the extent of students' problems solving skill as learning evaluation materials, so it is necessary to develop a assessment instrument of the physics problem solving skill based on Polya’s stages which include four stages of problem solving namely understanding the problem, making a problem solving plan, implementing the plan, and re-examining the results obtained.

METHOD

Development Procedure

This research is a research & development using a modified 4D development model (Thiagarajan, 1974) which consists of 4 stages, including: Define includes material analysis and learning objectives related to students' solving abilities; The design includes the selection of an instrument rating format that is in based on the Polya's stages; Develop includes the development and trial of products by experts and users; and Dissemination, at this stage is still limited to the use of assessment instruments only at the time of this study.

Population and Sample

The product of the development is in the form of an instrument of problem-solving skill that has been tested for eligibility with internal validity and empirical tests. Internal validity test is related to construct validity and content through expert judgment and at the same time as revision material, which in this case is by two experts and one learning practitioner. Empirical test on 81 XI MAN 1 students in Yogyakarta.

Data Collection and Instrument

The instrument for evaluating internal validity is in the form of a questionnaire with three answer choices according to the content of the question, namely: "In accordance", "Needs revision", and "Not suitable". Revisions are made to the contents of the question given the answer choices "Need revision" or from special input by experts on products that have been made. Whereas for the choices in the "Not suitable" column, the question is deleted if the number of questions still represents the competency being tested, but if not then it needs to be replaced with another question.

Data from the results of internal validity are explained and determined only qualitatively (very good, well, enough, less, and very less) based on criteria adapted from the formula developed by (Azwar, 2013), where susceptible scores for each category are calculated using the formula as in Table 1.

Table 1. Criteria for the Ideal Assessment Category

No.	Score Range	Category
1	$X > Mi + 1,50\ SDi$	Very good
2	$Mi + 0,50\ SDi < X \leq Mi + 1,50\ SDi$	Well
3	$Mi - 0,50\ SDi < X \leq Mi + 0,50\ SDi$	Enough
4	$Mi - 1,50\ SDi < X \leq Mi - 0,50\ SDi$	Less
5	$X \leq Mi - 1,50\ SDi$	Very less

Information:

X : average total score obtained
Mi : mean ideal score
SDi : ideal standard deviation

Data Analysis

The empirical test data is used to test the feasibility in terms of validity, reliability, distinguishing features, and the level of difficulty of the test instrument, which is then analyzed with the help of the Quest program. Validity testing is based on MNSQ Infit values, with valid item criteria if it is at a vulnerable score of 0.77 - 1.30 (Adams & Khoo, 1996). Reliability testing uses the internal consistency method, namely by testing the instrument only once then analyzed, where the reliability coefficient is interpreted based on the criteria in Table 2. Analysis of the level of difficulty test instrument is interpreted based on the criteria in Table 3. While the magnitude of the distinguishing power is interpreted according to (Arikunto, 2012) with the provisions based on the criteria in Table 4.

Table 2. Criteria for the Reliability Coefficient (Rusman, 2012)

No.	Score Range	Category
1	0.80 – 1.00	Very high
2	0.60 – 0.79	High
3	0.40 – 0.59	moderate / sufficient
4	0.20 – 0.39	Low
5	0.00 – 0.19	Very low

Table 3. Criteria for the level of difficulty test instruments (Arikunto, 2012)

No.	Score Range	Category
1	0.00 – 0.30	Difficult
2	0.31 – 0.70	Moderate
3	0.71 – 1.00	Easy

Table 4. Criteria for the magnitude of the distinguishing power (Arikunto, 2012)

No.	Score Range	Category
1	negative	Not good
2	0.00 – 0.20	Poor
3	0.21 – 0.40	Enough
4	0.41 – 0.70	Good
5	0.71 – 0.10	Excellent

RESULT AND DISCUSSION

Expert Judgment Results

The results of this development research in the form of a test instrument for description / physics essay on the subject of the kinetic theory of gas which contains aspects of problem solving based on Polya's stages. The selection of tests in the form of description is intended so that the problem-solving ability by students is easily identified. Preparation begins with the manufacture of questions about the sub-topic according to the syllabus, the ability to be measured as well as the number of items, then proceed with the manufacture of questions along with the answer key and the rules for scoring for each item.

This test instrument consisted of 5 items, which were then validated in stages, namely validity by experts and by users. Data from the results of internal testing were conducted by two lecturers and one teacher, with an expert questionnaire assessment sheet consisting of 20 statements. The score given is 1-3, which means the lowest score is 20 and the highest score is 60, so a Mi score of 40 is obtained, and an SDi score of 6.67. Because the total average score (X) obtained is 58, it falls into the category X> Mi + 1.50 SDi or very good. This is proven by all items that are judged to be very suitable for the competency to be tested and suitable for use in measuring the physics problems solving skill based on Polya's stages.

Empirical Testing Results

After internal validation and the necessary revisions, the first empirical test, the validity test, is continued. Validity is a measure that shows the level of validity or validity of an instrument. The validity of the instrument was empirically carried out by testing it on students who were at the same level and had obtained learning on the gas kinetic theory material, namely at 81 students in class XI of high school. Data from the trial results are then analyzed using the help of the Quest program.

Based on the results of this test it was found that out of 5 items about problem solving skills that were tested based on Infit- MNSQ included in the acceptance criteria / limit of 0.77-1.30 according to the PCM model or 1-PL model, or in other words that the five items proved to be valid. The complete results of this validity test can be seen in Figure 1.

```
Pengujian Instrumen Tes Kemampuan Pemecahan Masalah Fisika
------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order                                      5/ 9/19 13:37
all on all (N = 81 L = 5 Probability Level= .50)
------------------------------------------------------------------------------------------------
    ITEM NAME        |SCORE MAXSCR|  THRESHOLD/S                                | INFT  OUTFT INFT  OUTFT
                     |            |    1        2        3        4        5    | MNSQ  MNSQ   t     t
------------------------------------------------------------------------------------------------
1   item 1           |  339  405  | -1.17    -.81     -.45     -.29     -.09   | 1.08  1.31   .4    .8
                     |            |   .70      .63      .54      .49      .47  |
                     |            |                                            |
2   item 2           |  188  405  |  -.66    -.17      .46      .82     1.02   | 1.07  1.02   .5    .2
                     |            |   .41      .38      .38      .39      .41  |
                     |            |                                            |
3   item 3           |  179  405  |  -.16     .20      .35      .58      .84   | 1.10  1.11   .7    .5
                     |            |   .36      .37      .37      .38      .38  |
                     |            |                                            |
4   item 4           |  192  405  | -2.28    -.19      .34     1.14     1.14   |  .94   .94  -.3   -.3
                     |            |   .81      .41      .37      .31      .31  |
                     |            |                                            |
5   item 5           |  209  405  |  -.48     .11      .19      .42      .80   |  .82   .71  -1.3  -1.3
                     |            |   .39      .38      .37      .37      .38  |
                     |            |                                            |
------------------------------------------------------------------------------------------------
Mean                 |            |   .01                                      | 1.00  1.02   .0    .0
SD                   |            |   .40                                      |  .12   .22   .8    .8
================================================================================================
```

Figure 1. Validity Test Results

After being tested for validity, it is followed by a test of difficulty level, and distinguishing features. A recap of the results of testing the validity, level of difficulty, and distinguishing power of the test instrument for the ability to solve physics problems according to Polya's stages can be seen in Table 5 below.

Table 5. Recap of Empirical Test Results for Problem Solving Ability Test Instruments

No. Item	**Validity (INFIT MNSQ Value)**	**Difficulty Level**	**Distinctive**
Item 1	1.08 (Valid)	0.73 (Easy)	0.46 (Good)
Item 2	1.07 (Valid)	0.21 (Difficult)	0.46 (Good)
Item 3	1.10 (Valid)	0.27 (Difficult)	0.55 (Good)
Item 4	0.94 (Valid)	0.16 (Difficult)	0.40 (Good)
Item 5	0.82 (Valid)	0.30 (Difficult)	0.61 (Good)

Reliability is used to indicate the extent to which a measuring instrument can be trusted or relied upon in research, where a set of measuring instruments is said to be reliable if it provides relatively fixed measurement results. This reliability testing uses the method of internal consistency, because this method is the simplest, namely by testing the instrument once and then analyzed using a particular technique which in this case uses the help of the Quest program. The reliability test results obtained that the Internal Consistency value of 0.63 is included in the high category. The recap of the reliability test results can be seen in Table 6.

Table 6. Test Results of Test Instrument Reliability Ability to Test Problems

Criteria	**Value**
Mean test score	14.28
Standard deviation	5.42
Internal Consistency	0.63

Based on the steps of this development, five items measuring the physics problem solving skill on the subject of school gas kinetic theory based on Polya's stages, have met the requirements and are suitable for use. These requirements include: (a) content validation according to expert judgment included in the excellent category and validity while empirically proven to be fit with the Rasch Model, (b) instrument reliability is included in the high category of 0.63, (c) the distinguishing power of all items is included in either category, so that it meets the requirements of a good instrument, (d) but the level of difficulty of the questions varies less, where 1 item is in the easy category, 4 other items are in the difficult category, and not in the moderate category. Because most items are included in the difficult category, then this assessment instruments may be very suitable for measuring the physics problem solving skill of superior students.

CONCLUSION

A total of five test instrument items developed were stated to be eligible and suitable to be used to measure the physics problem solving skill on the subject of school gas kinetic theory based on Polya's stages. These requirements include: (a) content validation according to expert judgment included in the excellent category and validity while empirically proven to be fit with the Rasch Model; (b) instrument reliability is included in the high category of 0.63; (c) the distinguishing power of all items is included in either category, so that it meets the requirements of a good instrument; (d) but the level of difficulty of the questions varies less, where 1 item is in the easy category, 4 other items are in the difficult category, and not in the moderate category.

REFERENCES

Adams, R. J., & Khoo, S.-T. (1996). *Quest: The interactive test analysis system version 2.1*. Australian Council for Educational Research.

Alfika, Z. A., & Mayasari, T. (2018). Profil Kemampuan Memecahkan Masalah Pelajaran Fisika Siswa MTs. *Prosiding Seminar Nasional Quantum*, *25*, 583–589.

Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan* (2nd ed.). Bumi Aksara.

Asysyifa, D. S., Jumadi, Wilujeng, I., & Kuswanto, H. (2019). Analysis of Students Critical Thinking Skills Using Partial Credit Models (PCM) in Physics Learning. *International Journal of Educational Research Review*, *4*(2), 245–253. https://doi.org/10.24331/ijere.518068

Azwar, S. (2013). *Tes Prestasi*. Pustaka Pelajar.

Baiduri. (2015). Pengaruh Tahapan Polya dalam Pemecahan Masalah terhadap Ketuntasan Belajar Geometri Siswa Sekolah Menengah Pertama. *Jurnal Pendidikan Matematika*, *6*(1), 41–48.

Boonjeam, W., Tesaputa, K., & Sri-ampai, A. (2017). Program Development for Primary School Teachers' Critical Thinking. *International Education Studies*, *10*(2), 131–138. https://doi.org/https://doi.org/http://dx.doi.org/10.5539/ies.v10n2p131

Erfianti, L., Istiyono, E., & Kuswanto, H. (2019). Developing Lup Instrument Test to Measure Higher Order Thinking Skills (HOTS) Bloomian for Senior High School

Students. *International Journal of Educational Research Review*, *4*(3), 320–329. https://doi.org/10.24331/ijere.573863

Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Istiyono, E., Mustakim, S. S., Widihastuti, Suranto, & Mukti, T. S. (2019). Measurement of Physics Problem-Solving Skills in Female and Male Students by Phystepross. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 170–176. https://doi.org/10.15294/jpii.v8i2.17640

Julianingsih, S., Rosidin, U., & Wahyudi, I. (2017). Pengembangan Instrumen Asesmen HOTS untuk Mengukur Dimensi Pengetahuan IPA Siswa di SMP. *Jurnal Pembelajaran Fisika*, *5*(3), 59–68.

Kemendikbud. (2016). *Permendikbud Nomor 24 Tahun 2016 tentang Standar Kompetensi Inti dan Kompetensi Dasar*. Kementerian Pendidikan dan Kebudayaan.

Mustofa, M. H., & Rusdiana, D. (2016). Profil Kemampuan Pemecahan Masalah Siswa pada Pembelajaran Gerak Lurus. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *2*(2), 15–22. https://doi.org/10.21009/1.02203

Octavia, N. R., Nyeneng, I. D. P., & Suana, W. (2017). Pengembangan Kuis Interaktif Tipe Multiple Choice Materi Impuls dan Momentum. *Jurnal Pembelajaran Fisika*, *5*(1), 145–156.

Polya, G. (1957). *Polya How To Solve It: A New Aspect of Mathematical Method* (2nd ed.). Princeton University Press.

Rusman. (2012). *Model-model Pembelajaran: Mengembangkan Profesionalisme Guru* (2nd ed.). PT Rajagrafindo Persada.

Thiagarajan, S. S. (1974). *Instructional Development for Training Teacher of Exeptional Children*. Indiana University.

Widoyoko, S., & Putro, E. (2016). *Teknik Penyusunan Instrumen Penelitian*. Pustaka Pelajar.