Developing Assessment Instrument to Measure Senior High School Student's Mathematical Representation Ability in Physics Learning

L P Sari[1], E Istiyono[1]

[1]*Physics Education, Graduate School, Universitas Negeri Yogyakarta, Jl. Colombo No. 1, Yogyakarta, Indonesia*
Correspondent Email: lindahpermata.2018@student.uny.ac.id

ABSTRACT

Physics deals with a lot of quantities that have size and direction and thus require a mathematical representation to describe it. So that in the context of physics learning cannot be separated from the application of mathematical representation. This research aims to: 1) determine the feasibility of assessment instruments by conducting content validation; 2) find out the quality of grading instruments by conducting empirical validation. The test development stage includes: 1) test planning, 2) test trials, and 3) revision and assembly of tests. This assessment instrument contains particle dynamics, the laws of planetary motion and work and energy at the senior high school level. The assessment instrument was developed based on three components of mathematical representation ability. The results of this study are: 1) the results of the content validation data analysis show 39 items of the test items are valid and worth testing; and 2) the results of the empirical validation data analysis are shown from three item characteristics, namely: a) the results of the validity test show that the 39 developed test items are fit with the PCM model with the average INFIT MNSQ items are 1.00 and the standard deviation is 0.09, so all questions are valid; b) the reliability value is 0.46, so the reliability of the questions belongs to the poor category and the test is suitable for students with abilities, $-1.7 \leq \theta < 3.6$; and c) the difficulty value of all items is in the interval -1.28 and 1.23, so that the quality of all items is categorized as good.

Keywords: Developing assessment instrument, Mathematical representation ability, Physics assessment.

1. INTRODUCTION

Physics is closely related to mathematics. Physics deals with a large number of quantities that have size and direction and requires a special mathematical language, vector language, to describe those quantities [1]. Mathematical modeling is also used to describe phenomena and explain the relationships between variables in physics [2]. Therefore, teaching and learning of physics must include the conversion of physics modeling into mathematical modeling (for example, regarding functional relations) and interpretation of mathematical models from a physical point of view [3]. It cannot be denied that the use of mathematical representation in solving physical problems is something that cannot be avoided.

Taking mathematical representation as a variable is based on one of the purposes of mathematical representation namely as a systematic description and study of patterns, so it is not surprising that the world of mathematics is open to many other worlds, such as physics [4]. Mathematical representation is intended as an equation and related symbols [5]. Standards of representation are expected to be mastered by students according to NCTM (2000), namely: 1) making and using representations to recognize, note or record, and communicate mathematical ideas; 2) choose, apply, and translate between mathematical representations to solve problems; 3) using representations to model and interpret physical, social, and mathematical phenomena [6].

Physics, in particular, uses mathematical representations and uses mathematics to describe phenomena, build models, and solve problems [7]. The aspect of mathematical representation can be seen from the presentation of data or information from a problem to the representation of images, diagrams, graphs or Tables,

problem solving that involves mathematical expression and writing steps to solve problems with words [8]. The aspect of mathematical representation ability in this study was synthesized from several opinions above. The aspects of mathematical representation ability can be seen in Table 1.

Table 1. The aspects of mathematical representation ability.

Aspects	Sub Aspects	Indicator of Mathematical Representation Ability
The use of non-mathematical representations to present and communicate mathematical information	Presentation of Information	Present mathematical information with visual representations (pictures, diagrams, graphs or Tables)
	Information communication	Communicating mathematical information with verbal representations (words)
Problem solving involving mathematical expressions	Problem Solving	Maintaining the problem to get the final result by involving mathematical expressions
The use of mathematical representations to model and interpret phenomena	Modeling	Model phenomena using mathematical representations
	Interpretation	Interpretation of phenomena using mathematical representations

In physics class, at least some mathematical material is used routinely, such as, arithmetic, algebra or analytical tools [9]. In some cases they found that students' difficulties with concepts of physics might be related to problems with mathematics, either understanding mathematical concepts, applying concepts to physics, or in understanding some of the mathematical representations used in physics [10]. Based on the results of essay tests conducted by researchers at junior high schools in North Sumatra showed results in the low category on the variable of students' understanding and representation ability [11]. Research on multiple representations also shows that of all types of representational ability, the value of mathematical representation ability is the smallest [12].

From the importance of the role and the difficulties faced by students due to lack of mathematical representation ability, some researchers have applied learning models to encourage students' mathematical representation abilities (e.g. [13]; [14]). Some development of assessment instruments in senior high school physics learning such as higher order thinking skills (e.g. [15]; [16]) and creative thinking abilities (e.g. [17]; [18]) already exist. However, an instrument for evaluating the mathematical representation ability has not yet been developed so there is a limited test

instrument to identify the mathematical representation ability of high school students in learning physics.

Testing physical material to measure the ability of mathematical representation requires a valid and reliable assessment instrument [19]. Validity means the accuracy of interpretation of the results of measurement procedures and reliability means the consistency of measurement results [20]. Assessment instruments can be said to have high validity if the instruments made carry out their measuring functions [21]. These characteristics will be known after the stages of developing a systematic measurement instrument according to the expert. Based on this, it is necessary to develop an instrument to measure the mathematical representation ability of high school students in learning physics that qualifies as a good instrument. This development research aims to: 1) determine the feasibility of the assessment instruments by conducting content validation; 2) find out the quality of grading instruments by conducting empirical validation.

2. METHOD

This research belongs to the type of Research and Development (R&D) research with the development model used for test development is the Oriondo and Antonio model. This research was conducted at Senior High School 1 of KasihanYogyakarta and Senior High

School 1 of Banguntapan Yogyakarta. Subjects numbered 287 students. The subjects in this study were selected by purposive sampling method. Students who become the population are senior high school students who have studied particle dynamics, the laws of planetary motion and work and energy. The test development stage includes: 1) test planning, 2) test trials, and 3) revision and assembly of tests. The test planning stage includes: (a) goal setting, (b) making test grids (c) choosing test formats, (d) writing test items, (e) validating by material experts, and (f) repairing items and assembling 3 test set. The test phase of the test consists of: (a) determining the test subject, (b) conducting the test run (empirical validation), and (c) analyzing the test result data based on Item Response Theory (IRT).

The validation data will be analyzed as follows:

2.1. Content Validity

The content validation sheet is arranged on an interval of 1 to 4 and is calculated using the Aiken validity formula [22]:

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

Information:

Rating scale from l_0 to c

l_0 = smallest scale

r = from $l_0 + 1$ to $l_0 + c - 1$

$s = r - l_0$

s = the number s of n rater.

2.2. Empirical Validity

2.2.1. Test Validity.

Item compatibility with the model (PCM) is used to determine the empirical validity of items or referred to as item fit. The four categories of politomus data which were students' answers were analyzed by PCM model [23]. Scores can be analyzed using the Quest program, to get the estimated value of INFIT mean of square and INFIT t.

2.2.2. Instrument Reliability.

Instrument reliability is carried out with the Quest program, which reads sh output, which is a summary of case estimates. The reliability value is then proven by using graph analysis of the relationship between information functions and Standard Error of Measurement (SEM) [24]. The program used is the Parscale program. The test can be found to be suitable for students with abilities in a certain range θ based on the information function and SEM obtained.

Table 2. Interpretation of reliability values with the Rasch model.

Value of Reliability	Interpretasi Reliabilitas
> 0.94	Excellent
0.91 – 0.94	Very good
0.81 – 0.90	Good
0.67 – 0.80	Moderate
< 0.67	Poor

2.2.3. Item Difficulty (b).

Analysis using the Quest program can produce item difficulty levels (b). The items are stated as good if the difficulty index is more than -2.0 or less than +2.0. If it approaches -2.0 then the item is said to be easy, whereas if it approaches +2.0 then the item is said to be difficult [25].

3. RESULTS AND DISCUSSION

3.1. Result

At the test planning stage, the purpose of the test and the test grid have been compiled as a reference in writing the item. The questions compiled amounted to 39 questions divided into 3 sets, namely Package A, Package B and Package C. The test is in the form of reasoned multiple choice, where there are 5 answer choices and 5 reason choices. Questions with multiple choices alone cannot measure students' ability to solve problems

because they do not consider the steps taken by students [26]. So that in this study the test uses a reasonable multiple choice form.

The material contained in the test items is material about particle dynamics, the laws of planetary motion and

work and energy. Writing test items is developed based on aspects of mathematical representation ability that can be seen in Table 1. The distribution of the number of test items from each aspect can be seen in Table 3.

Table 3. Matrix Math Representation Skill Test.

Aspects	Sub Aspects	Items
The use of non-mathematical representations to present and communicate mathematical information	Presentation of Information	5
	Information communication	9
Problem solving involving mathematical expressions	Problem Solving	9
The use of mathematical representations to model and interpret phenomena	Modeling	9
	Interpretation	7
Total		39

After conducting all stages of the test development, the results of data analysis are obtained, namely:

3.1.1. Content Validity.

Content validation is based on expert judgment. There are 5 experts consisting of one professor, two doctors and two professional physics teachers. The

validation instrument consisted of 4 scales so that a threshold value of 0.87 was obtained. Aiken (1985) explained that the test item is said to be valid if it has a Aiken's V value greater than or equal to a predetermined threshold value [27]. The results of the analysis using Aiken's Value can be seen in Table 4.

Table 4. Aiken's V analysis results for each test item.

Package A			Package B			Package C		
Item Number	Aiken's V	Criteria	Item Number	Aiken's V	Criteria	Item Number	Aiken's V	Criteria
1	0.87	Valid	1	1.00	Valid	1	1.00	Valid
2	1.00	Valid	2	1.00	Valid	2	0.93	Valid
3	0.87	Valid	3	1.00	Valid	3	1.00	Valid
4	1.00	Valid	4	1.00	Valid	4	1.00	Valid
5	1.00	Valid	5	1.00	Valid	5	0.87	Valid
6	1.00	Valid	6	1.00	Valid	6	1.00	Valid
7	1.00	Valid	7	0.93	Valid	7	0.87	Valid
8	1.00	Valid	8	1.00	Valid	8	0.93	Valid
9	1.00	Valid	9	1.00	Valid	9	1.00	Valid
10	0.93	Valid	10	0.87	Valid	10	1.00	Valid
11	1.00	Valid	11	1.00	Valid	11	1.00	Valid
12	1.00	Valid	12	0.87	Valid	12	1.00	Valid
13*	0.93	Valid	13*	0.93	Valid	13*	0.93	Valid

14*	0.93	Valid	14*	0.93	Valid	14*	0.93	Valid
15*	1.00	Valid	15*	1.00	Valid	15*	1.00	Valid

*Anchor Item

Total items developed were 39 items. Three of them are anchor item. In Table 5 the anchor item are marked with the asterisks. From Table 5 we get the results of the content validation of the 3 question packages. It can be seen in Table 5 that the Aiken's V values for each test item are in the interval of values 0.87-1.00. When compared with the threshold value of 0.87, it can be concluded that all questions are valid.

From the results of content validation, the experts also gave suggestions for improvement. The revised questions are then tested on students.

Table 5. The serial number of items in the Quest program.

3.1.2. *Empirical Validity.*

The trial was conducted at Kasihan 1 Senior High School of Yogyakarta and Banguntapan 1 Senior High School of Yogyakarta. The total number of respondents was 287 students. Data from the tested results will produce item characteristics. The intended item characteristics are empirical validity, reliability and item difficulty. Data were analyzed using the Quest program. The serial number in Table 5 is used in the Quest program.

Serial Number	Item Number on Package	Serial Number	Item Number on Package	Serial Number	Item Number on Package
1	13*	14	11A	27	12B
2	14*	15	12A	28	1C
3	15*	16	1B	29	2C
4	1A	17	2B	30	3C
5	2A	18	3B	31	4C
6	3A	19	4B	32	5C
7	4A	20	5B	33	6C
8	5A	21	6B	34	7C
9	6A	22	7B	35	8C
10	7A	23	8B	36	9C
11	8A	24	9B	37	10C
12	9A	25	10B	38	11C
13	10A	26	11B	39	12C

*Anchor Item

Empirical validity is determined based on the goodness of fit of the model (PCM). The estimation results from the Quest Program can be seen in Table 6.

Table 6. The result of empirical validation.

No	Parameter	Item Estimates	Testee Estimates
1	Average and Standard Deviation of INFIT MNSQ	1.00±0.09	1.00±0.22
2	Average and Standard Deviation of INFIT t	0.06±0.77	0.04±0.81
3	Average and Standard Deviation of OUTFIT MNSQ	0.99±0.12	0.99±0.31
4	Average and Standard Deviation of OUTFIT t	0.04±0.64	0.08±0.65

| 5 | Item or testee score 0 | 0 | 0 |
| 6 | Item or testee test perfect | 0 | 0 |

From Table 6, six output parameters are presented from the Quest program. The overall goodnesss of fit characteristic can be determined based on the average and standard deviation of MNSQ INFIT values and the average and standard deviation of OUTFIT t. The average and standard deviation of MNSQ INFIT resulting from the estimated items and tests are 1.00 ± 0.09 and 1.00 ± 0.22. The average and standard deviation of OUTFIT t generated from the estimated items and tests are 0.04 ± 0.64 and 0.08 ± 0.65. In accordance with the explanation of Adam and Kho (1996) that an item is declared fit with the model if the MNSQ INFIT value is in the range of 0.77 to 1.30 [28].

In addition, items that match the Rasch model (1 PL) also have fit requirements if the outfit is ≤ 2 [28]. PCM is the development of the Rasch Model dichotomous into a politomus with one logistical parameter, namely the level of difficulty [15]. Based on this, the results of the average value and standard deviation of INFIT MNSQ and the average value and standard deviation of OUTFIT t in Table 6 satisfy both conditions. The average and standard deviation of MNSQ INFIT values are between 0.77-1.30 intervals, while the average and standard deviation of OUTFIT t value is smaller than 2. So it can be concluded that all items fit test with the model (PCM).

More clearly, the value of INFIT MNSQ for each test item is in the interval 0.77-1.30, so there are no questions that fail. The results of INFIT MNSQ for each test item can be seen in Table 7.

Table 7. INFT MNSQ per test item.

Serial Number	INFIT MNSQ	Serial Number	INFIT MNSQ	Serial Number	INFIT MNSQ
1	1.07	14	1.00	27	0.96
2	0.93	15	1.01	28	0.90
3	0.95	16	1.02	29	1.00
4	0.94	17	1.04	30	1.05
5	1.05	18	1.14	31	0.98
6	0.98	19	1.18	32	1.20
7	1.01	20	0.94	33	0.96
8	0.95	21	0.98	34	1.08
9	0.97	22	1.04	35	1.21
10	0.93	23	0.87	36	0.90
11	0.97	24	0.93	37	0.88
12	1.09	25	0.85	38	1.08
13	1.11	26	1.05	39	0.96

*Anchor Item

The results of the analysis of the trial data also produced another item characteristic, namely the reliability of the questions. The reliability of the resulting problem is 0.46. Based on Table 2, the reliability value is classified as poor. Also in Figure 1 is presented the parscale output to find out which test items developed are suitable or reliable for any ability. In Figure 1 the intersection of the two curves is on a score scale of -1.7. It is estimated that the two curves will re-intersect on a score scale of approximately 3.6. This score shows that the developed mathematical representation test is suitable for students with abilities, $-1.7 \leq \theta < 3.6$.
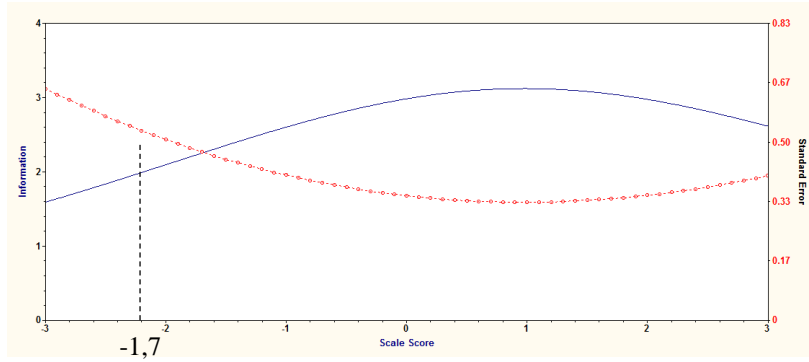
Figure 1. Relation Curve of Information Function and SEM

Next, the last item characteristic is item difficulty. From the results of data analysis, the difficulty value of all items is at intervals of -1.28 and 1.23. The easiest item is number 17 and the most difficult item is number 27. The average overall difficulty is 0.00 with a standard deviation of 0.57. Hambleton and Swaminathan (1985) state that good items have a difficulty index of more than

-2.0 and less than 2.0. Based on this, all test items can be said to be good [25].

Test items divided into three aspects of mathematical representation ability have different difficulties. The level of difficulty index for each aspect and sub-aspect can be seen in Table 8 and Figure 1.

Table 8. Average difficulty index of aspect and sub aspect.

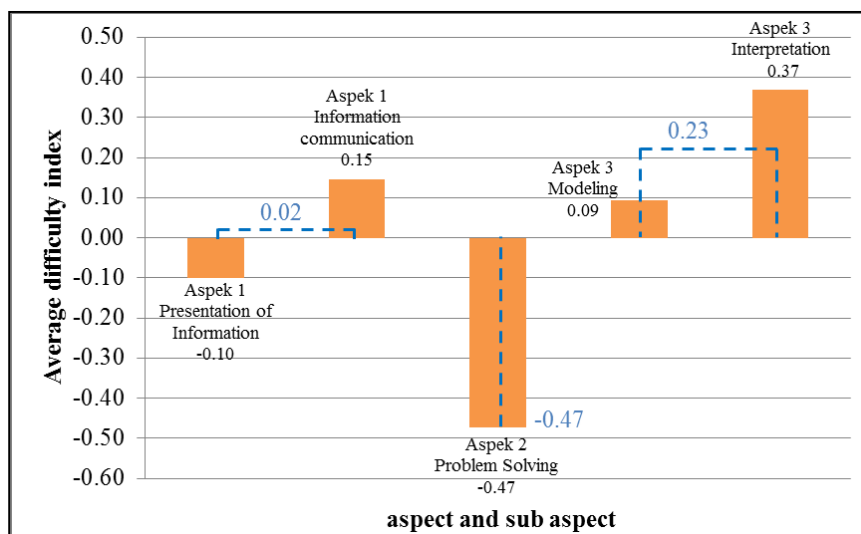| No	Aspects	Sub Aspects	Average Difficulty of Sub Aspect	Average Difficulty of Aspect
1	The use of non-mathematical representations to present and communicate mathematical information	Presentation of Information	-0.10	0.02
		Information communication	0.15	
2	Problem solving involving mathematical expressions	Problem Solving	-0.47	-0.47
3	The use of mathematical representations to model and interpret phenomena	Modeling	0.09	0.23
		Interpretation	0.37	

Figure 2. Average difficulty index of aspect and sub aspect.

Information Figure 2:

• Aspect 1: The use of multiple representations to present and communicate mathematical information

• Aspect 2: Problem solving involving mathematical expressions

• Aspect 3: The use of mathematical representations to model and interpret phenomena

In Table 8 it is known that the average difficulty of each aspect and sub-aspect of mathematical representation ability. The average difficulty aspect of using non-mathematical representations to present and communicate mathematical information is 0.02. The average difficulty of problem solving involving mathematical expressions is -0.47. The average difficulty aspect of using mathematical representations to model and interpret phenomena is 0.23. So the average difficulty aspect from low to high is the problem solving aspect involving mathematical expressions (aspect 2), aspects of using non-mathematical representations to present and communicate mathematical information (aspects 1), and aspects of using mathematical representations to model and interpret phenomena (aspect 3).

Based on Table 8 also explained the average difficulty of each sub-aspect. For more details can be seen in Figure 2. The lowest average sub-aspect difficulty is problem solving, -0.47 and the highest difficulty sub-aspect difficulty is interpretation, 0.37. So that the order of average difficulty sub aspects from low to high is problem solving, presentation of information, modeling, information communication, and interpretation.

3.2. Discussion

3.2.1. Content Validity.

From the above results, 39 test items made have been tested for validity by expert judgment. The results of the content validation of all items are valid. Assessments by experts can be used as an alternative proof of validity, but these results can be subjective depending on the background of their knowledge [19].

3.2.2. Empirical Validiry.

Valid test items are then tried out and analyzed. The results of the analysis showed 39 items of fit test, meaning that no items were fail. This can be seen from the acquisition of goodness of fit that meets the requirements of item compatibility with the model. The estimated level of difficulty of the items shows the quality of all questions in the good category. But the reliability of the questions belongs to the poor category. The developed mathematical representation test is suitable for students with abilities, $-1.7 \leq \theta < 3.6$.

Some factors that affect the reliability associated with the test are: (1) test length, (2) homogeneity of test material, (3) homogeneity of item characteristics, and (4) variability of scores. Reliability related to students is influenced by factors: (1) heterogeneity of groups, (2) the experience of students taking the test, and (3) the motivation of students [20]. Based on this, in this study the estimated number of questions is not enough. The condition of students who have no experience in taking the test in the form of multiple choice reasoned and lack of motivation of students in working on the questions.

Furthermore, the average difficulty of each aspect shows a variety of values. Based on the order of average difficulty, aspect 3 is the most difficult and aspect 2 is the easiest. However, if viewed from the aspect of mathematical representation ability compiled by the author, the three aspects are not a stage, so the average difficulty results are not graded from low to high. This can also be caused by writing items that each aspect has a different level of cognitive domain. This allows items on aspect 1 to be more difficult than items on aspect 2.

Furthermore, in aspect 1 shows higher average difficulty value than aspect 2. If viewed from the indicators of mathematical representation ability, students are expected to present mathematical information with visual representations (pictures, diagrams, graphs or tables). High difficulty value in this sub-aspect can also be influenced by students' difficulty in converting information from mathematical representation to visual representation. Likewise in the information communication sub-aspect, students have difficulty in changing information from mathematical representation to verbal representations.

On the other hand, in aspect 3, problem solving tends to be low. This happens because in solving physical problems that only use solutions with mathematical representations, so students get results through mathematical formulas. This is relevant to research Theasy, Wiyanto, and Sujarwata (2018) that students who are skilled in problem solving categories are accustomed to using non-mathematical representations such as graphs and diagrams. While students who are less skilled in problem solving categories, tend not to use non-mathematical representations. Students have the habit of using mathematical solutions in solving physical problems, so students get stuck in mathematical habits without understanding physical concepts [12]. This reason also answers why in aspect 1 students have difficulty in solving problems. Whereas in aspect 3, to solve problems students need to have ways of solving problems at a higher level of the cognitive domains.

4. CONCLUSION AND SUGGESTION

Based on the results and discussion, all stages of test development have been carried out. The results of this study are: 1) the results of the content validation data analysis show 39 items of the test items are valid and worth testing; and 2) the results of the empirical validation data analysis are shown from three item characteristics, namely: a) the results of the validity test show that the 39 developed test items are fit with the PCM model with the average INFIT MNSQ items are 1.00 and the standard deviation is 0.09, so all questions are valid; b) the reliability value is 0.46, so the reliability of the questions belongs to the poor category and the test is suitable for students with abilities, $-1.7 \leq \theta < 3.6$; and c) the difficulty value of all items is in the interval -1.28 and 1.23, so that the quality of all items is categorized as good. The results of the analysis of this test item can be a guideline in assembling and presenting questions to measure the mathematical representation ability of high school students in physics learning. Suggestions for further development are a larger number of respondents, increased test length, and tests are widely tested with the help of technology.

REFERENCES

Halliday and Resnick 2014 Fundamental of Physics 10th Edition (United States: Wiley)

Opfermann M, Schmeck A, and Fischer H E 2017 Multiple representations in physics and science education–why should we use them? In multiple representations in physics education Models and Modeling in Science Education 10 1-22 https://doi.org/10.1007/978-3-319-58914-5_1

Bing T J and Redish E F 2009 Analyzing problem solving using math in physics: Epistemological framing via warrants Physical Review Special Topics - Physics Education Research 5 1-15 https://doi.org/10.1103/physrevstper.5.020108

Goldin G A 2002 Representation In Mathematical Learning And Problem Solving (New York: Routledge)

Kohl P B and Finkelstein N 2017 Understanding and promoting effective use of representations in physics learning Multiple Representations in Physics Education 10 231-54 https://doi.org/10.1007/978-3-319-58914-5_11

NCTM 2000 Principles and Standards for School Mathematics (Reston, VA: NCTM)

Angell C, Kind P M, Henriksen E K, and Guttersrud Ø 2008 An empirical-mathematical modelling approach to upper secondary physics Physics Education 43 256-64 https://doi.org/10.1088/0031-9120/43/3/001

Sulastri, Marwan and Duskri M 2017 Kemampuan representasi matematis siswa melalui pendekatan pendidikan matematika realistik Jurnal Tadros Matematika 10 51-69 http://dx.doi.org/10.20414/betajtm.v10i1.101

Albe V, Venturini P, and Lascours J 2001 Electromagnetic concepts in mathematical representation of physics Journal of Science Education and Technology 10 97-203 https://doi.org/10.1023/A:1009429400105

Wagner J F, Manogue C A, Thompson J R, Rebello N S, Engelhardt P V, and Singh C 2012 Representation issues: Using mathematics in upper-division physics AIP Conference Proceedings 1413 89-92 https://doi.org/10.1063/1.3680001

Minarni A, Napitupulu E, and Husein R 2016 Mathematical understanding and representation ability of public junior high school in north sumatra Journal on Mathematics Education 7 43-56 http://dx.doi.org/10.22342/jme.7.1.2816.43-56

Theasy Y, Wiyanto, and Sujarwata 2018 Multi-representation ability of students on the problem solving physics Journal of Physics: Conference Series 983 012005 1-4 https://10.1088/1742-6596/983/1/012005

Rezeki S 2017 Meningkatkan kemampuan representasi matematis siswa melalui penerapan model pembelajaran novick Jurnal SAP 1 281-91 http://dx.doi.org/10.30998/sap.v1i3.1203

Widada W, Herawaty D, Jumri R, Zulfadli Z, and Damara B E P 2019 The influence of the inquiry learning model and the Bengkulu ethnomathematics toward the ability of mathematical representation Journal of Physics: Conference Series 1318 012085 1-4 https://10.1088/1742-6596/1318/1/012085

Istiyono E, Mardapi D, and Suparno 2014 Pengembangan tes kemampuan berpikir tingkat tinggi fisika (PysTHOTS) peserta didik SMA Jurnal Penelitian Dan Evaluasi Pendidikan 18 1-12 https://doi.org/10.21831/pep.v18i1.2120

Erfianti L, Istiyono E, and 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** 320-9 https://doi.org/10.24331/ijere.573863

Istiyono E, Dwandaru W B, and Rahayu F 2018 Pengembangan tes creative thinking skills fisika SMA (PhysCreTHOTS) berdasarkan teori tes

modern Cakrawala Pendidikan **38** 190-200 https://doi.org/10.21831/cp.v37i2.19233

Megawan M and Istiyono E 2019 Physics creative thinking measurement using two-tier multiple choice to support science, technology, engineering, and mathematics *Journal of Physics: Conference Series* **1233** 012068 1-7 https://doi.org/10.1088/1742-6596/1233/1/012068

Mardapi D, Kumaidi, and Kartowagiran B 2011 Pengembangan instrumen pengukur hasil belajar nirbias dan terskala baku Jurnal Penelitian dan Evaluasi Pendidikan 15 326-41 https://doi.org/10.21831/pep.v15i2.1100

Depdiknas 2008 Panduan Analisis Butir Soal (Jakarta: Depdiknas)

Azwar S 2011 Reliabilitas dan Validitas (Yogyakarta: Pustaka Belajar)

Aiken L R 1980 Content validity and reliability of single items or questionnaires Educational and Psychological Measurement 40 955-59 https://doi.org/10.1177/001316448004000419

Istiyono E 2018 *Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika dengan Teori Tes Klasik dan Modern* (Yogyakarta: UNY Press)

Mardapi D 2012 *Pengukuran Penilaian dan Evaluasi Pendidikan* (Yogyakarta: Nuha Medika)

Hambleton and Swaminathan 1985 Item Response Theory Principles and Applications (Boston: Kluwer Nijhoff Publishing)

Nadapdap A T Y and Istiyono E 2017 Developing physics problem-solving skill test for grade X student of senior high school Research and Evaluation in Education 3 114-23 https://doi.org/10.21831/reid.v3i2.14982

Aiken L R 1985 Three coefficients for analyzing the reliability and validity of ratings International of Educational and Phsycological Measurement 45 131-42 https://doi.org/10.1177/0013164485451012

Adams R J and Khoo S T 1996 Quest: The interactive test analysis system version 2.1 (Victoria: The Australian Council for Educational Research)

Dadan R and Sukardiyono 2015 Analisis butir dan identifikasi ketidakwajaran skor ujian akhir sekolah untuk standarisasi penilaian Jurnal Kependidikan 45 130-41 https://doi.org/10.21831/jk.v45i2.7490