Analysis of Based Test Instruments Local Wisdom of Terbang Papat to Measure Mathematical Representation Capabilities in Learning Physics

Danis Alif Oktavia[1], Ahdika Setyadi[2], Warsono[3], Supahar[4]

ARTICLE INFO

ABSTRACT

This study aims to analyze the mathematical representation ability test instruments based on local wisdom of terbang papat to high school students. The research method used by compiling the instrument grid based on aspects of the ability of mathematical representation based on local wisdom of terbang papat, which is then used to compile indicators and items. The test instrument consists of four indicators of the ability of mathematical representation, each of which has five items and has been validated by expert judgment, physics teacher and peer reviewer. The validated instrument was tested on 338 students from two high schools in Kudus Regency. Data obtained in the form of polytomus data, then analyzed using Partial Credit Model (PCM). The results obtained showed that all test instruments of mathematical representation ability were proven to be fit with PCM, reliability of test instruments was declared reliable based on PCM results, total information function was between -2.00 to +1.8, difficulty indexes were more than -2.0 and less than 2.0 ($-2.0 \leq b \leq +2.0$) which means all items are in good category. Thus the ability test instrument based on terbang papat local wisdom can be used to measure the ability of students' mathematical representation according to the partial credit model based on the five categories of polytomus data.

INTRODUCTION

Abstract and difficult learning processes have an impact on the ability to understand concepts and solve problems of students. Problem solving skills are important in achieving learning goals (Merriënboer, 2013). Taale (2011) in his research concluded that students will change the perception that physics is difficult to be interesting if learning is carried out by innovating methods or strategies with problem solving. Based on this problem Docktor and Mestre (2014) states that problem solving ability is one of the goals in physics research and learning. He also said that problem-solving abilities include the ability to categorize, selfdiagnosis, the ability to interpret diagrams, verbal abilities, the ability to read graphics, the ability to analogize and the ability of mathematical equations or symbolic implementations. Based on these problems, it is necessary to develop a mathematical representation ability test instrument to achieve improvement in problem solving skills. Mathematical representation is a form of someone's interpretation of a problem that uses mathematical operations to find solutions. Students who are able to use mathematical representations more easily understand physics subject matter (Pospiech, 2012). Kurnaz and Arslan (2014) explain that the use of representation during physics learning activities takes place, can make learning more meaningful and improve students' conceptual understanding. The representation in question can be presented in the format of images, analogies, graphics, diagrams, vectors and verbally (Cock, 2012).

To monitor the process, progress, and improvement of student learning outcomes on an ongoing basis, assessment is needed. Educational assessment is the process of collecting and processing information to determine the achievement of student learning outcomes (Regulation of the Minister of National Education No. 20, 2007). Assessment can be done orally or in writing. Written assessment is done with a written test. Broadly speaking there are two forms of written test questions, namely: choosing answers and supplying answers. Written test questions that answer by choosing answers include: multiple choice, two choices (right-wrong, yes-no), matchmaking, and causation. At present, many studies examine the ability of mathematical representation. The assessment of course requires a measurement test in order to measure the ability of mathematical representation appropriately. In this study, the development of a description test was chosen to measure the mathematical representation abilities of students.

Physics is learning that is closely related to everyday life where it will become more meaningful when combined with local wisdom (Kun, 2013). The local wisdom of the Kudus district is in the form of terbang papat musical instrument that the surrounding community calls "terbangan". Terbang papat art is a cultural

art that breathes Islam and is very close to the pattern of people's lives on the North Coast of Central Java from the heart to the city.

Assessment in education uses two kinds of measurement theories, namely: classical measurement theory and modern measurement theory. Classical Test Theory is also called Classical True-Score Theory (CTT), called the Classical Test Theory because the elements of this theory have been developed and applied for a long time, but still survive today (Suryabrata, 2002, p.21). According to the classical measurement theory the scoring of test results is usually carried out partially based on the steps that must be taken to answer the item correctly. Scoring is done per step and scores per participant item are obtained by summing the scores of students each step, and the ability to be estimated by a raw score. This scoring model is not necessarily right, because the difficulty level of each step is not taken into account. That weakness triggered a new, more adequate theory, namely Item Response Theory / IRT. If the CTT focuses on information at the test level, the IRT focuses on information at the item level. The implementation of the IRT model is based on several assumptions: (1) the results of participant tests on an item can be predicted by a series of factors called characteristics or abilities; and (2) the relationship between student test results on grains and a set of traits explained by a monotonically rising function called a characteristic curve item (ICC) (Hambleton, Swaminathan & Rogers, 1991; Harvey & Hammer, 1999; Suryabrata, 2000) . Then the ICC curve explains the relationship between the characteristics and results of participant tests on each item. Item response theory is a modern measurement method commonly used in item analysis. Item response theory seen from the characteristics of question items is determined by the test participant's response (both high and low abilities). Based on the IRT model, the relationship between the responses of examinees and test items can be explained by what is called the item characteristic curve (ICC) (Wang, 2006).

The first and most important step in an IRT application is parameter estimation, both the tester's ability parameters (θ) and item characteristics parameters. The approach that can be used to estimate item parameters is the maximum likelihood estimation method (MLE) (Matthew, 2007). The basic principle of the MLE method is if there are random examples $x1$, $x2$, ..., xn of the distribution that has the probability probability function $f(xn; \theta)$, $\theta \epsilon \Omega$. The function of coexistence of opportunity is seen as a function θ (Hogg & Craig, 1978). The graph of the Categorical Response Function (CRF) is the relationship between opportunities for correct answers that get a category k score on item -j with the ability of the test participant (θ) (Toit, 2003). The higher the ability of the test participants, the opportunity to answer the correct items correctly will increase. Based on some of these descriptions, the researcher will guess the critical thinking ability of multiple choice test participants on the grounds of physics subjects with the theoretical granular theoretical approach with the Partial Credit Model (PCM) model.

Partial Credit Model (PCM) is the development of an Item Response Theory (IRT) logistic parameter model (1-PL) and includes the Rasch model. The use of more than two categories is sorted to record the results of individual interactions with an item (usually to recognize the level of truth or completion) (Master, 1999). When it is assumed that items follow PCM, the ability of higher individuals is expected to have a higher score than individuals with low abilities (Widhiarso, 2010). PCM is an analytical model of the IRT format where students' responses to problems can describe certain abilities. This model was developed to describe the relationship between item characteristics and characteristics or characteristics of respondents. PCM was developed to analyze test items that require several steps of completion. PCM can be given in steps that can be done by individuals. Thus, PCM is suitable for wearing on achievement tests, including physics questions that require the identification of the problem to the final solution. So that the development of test instruments by adhering to the PCM model can be used in measuring students' representation abilities.

The test developed in this research and development is in the form of a description of the ability of mathematical representation based on local wisdom of terbang papat to high school students in sound wave chapter. The purpose of this study is to develop a mathematical representation ability test instrument based on the local wisdom of terbang papat high school students in sound wave chapter. Another goal is to know the validity and reliability of the tests developed.

Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

METHOD

The research that will be carried out includes the types of research and development. This study used a 4-D development model (Define-Design-Develop-Disseminate) developed by Thiagajaran in 1974 (Arifin, 2012). The 4-D model consists of Define, Design, Develop, and Disseminate. The development and preparation of mathematical representation test instruments was carried out from November 2018 to January 2019. The trial was conducted in February 2019. The study was conducted in SMA 1 Bae Kudus, SMA 1 Jekulo, Kudus regency, Central Java.

Material

The steps to develop a test instrument follow the 4-D model. The first stage is the definition. At this stage there are five main stages to be carried out namely pre-research analysis, student analysis, concept analysis, task analysis and analysis of learning objectives. The second stage is design. This stage designs the test instrument by formulating indicators. At this stage the mathematical representation indicators are integrated with the local wisdom of terbang papat and the material to be tested.

The third stage is develop. This stage is to produce test instruments that will be validated by material experts and peer reviewers. The mathematical test ability instrument was developed from the four aspects of mathematical representation abilities that became a benchmark. There are four item description test questions developed in this research, each item represents one aspect of mathematical representation ability. The four aspects of the ability of representation and details of the questions are shown in table 1. This validation includes three aspects, namely construction, material and language. Construction aspects consist of indicators of suitability of questions with the subject matter being studied, suitability of the questions with indicators of achievement of learning outcomes, suitability of the questions with the level of understanding of students, and indicators of questions in accordance with indicators of practical ability. Material aspects consist of indicators of statements on the questions formulated briefly and clearly, the items are not dependent on the previous questions, the problem bar describes the ability of the questions, the question bar does not provide clues to the answer key and indicators of images, graphs, tables, or clear diagrams and functions. Language aspects consist of indicators using standard Indonesian, communicative in formulating question sentences, and problem indicators not using words / phrases that lead to multiple interpretations.
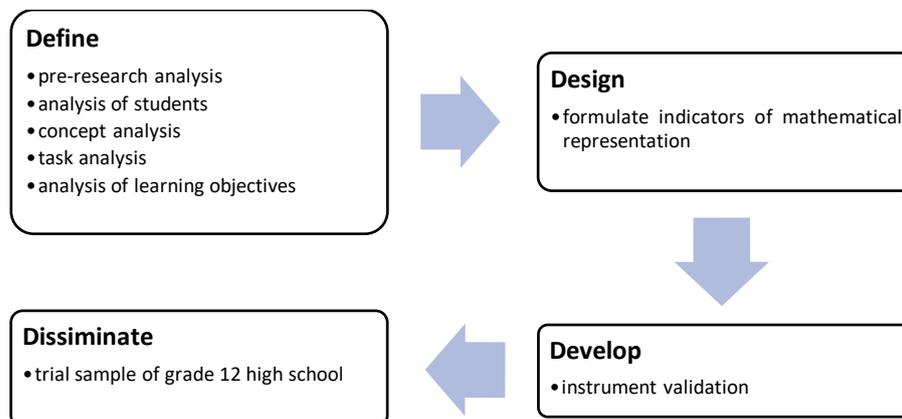

Figure 1. Chart of Research Procedures

After the development of the test instrument was completed, then it was validated, the validation process was carried out by 2 expert judgment, 2 physics teachers, and 4 peer reviewers produced suggestions, comments, and inputs to be used as a basis for analysis and revision of the mathematical representation ability test instrument developed and for product testing on students. The end of this stage is revision based on input and suggestions from expert judgment (revision 1), physics teacher, and peer reviewer (revision 2). The fourth stage is dissemination. At this stage, the results of validation by expert judgment, physics teachers, and peers were then tested on a sample of class XII MIPA SMA.

Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

Table 1. Aspects of Mathematical Representation Ability and Item Details

Mathematical Representation Indicator	Item Indicator	Item Number
Resolve physics problems using mathematical equations	Students can show the relationship to the concept of sound waves	3
	Students can formulate mathematical equations from visual representations in the form of graphics or images.	1
		4
	Students can determine the level of sound intensity at a certain distance.	5
Writing stages of solving mathematical problems with sentences	Interpreting the results of the analysis of the relationship between the distance between the sound wave intensity.	2

Data Analyses

The test response was in the form of politomus data with five categories, namely categories 1, 2, 3, 4, and 5. The response data of the students from the research results was analyzed using the Parcial Credit Model (PCM) model using the Quest and Parscale program packages. The Quest program is used to test: a) fit of items with the PCM model and determination of test reliability, b) estimation of the difficulty of items (delta), and c) estimation of difficulty items. The Quest program is also used to determine the value of the estimated ability of students (θ) expressed on a logit scale. The PARSCALE program is used to describe the parameters of Total Information Function (TFI) and Standard Error Measurement (SEM).

The data analysis technique obtained in determining the fit of the test instrument was developed by looking at the size of the INFIT Mean of Square and its standard deviation. Test items are said to be fit with the model if the average value of INFIT MNSQ approaches 1.0 with a standard deviation of 0.0. The criteria for fit for each item with the model follow the method set by Adam & Khoo, a point fit with the model if the INFIT value of MNSQ is in the range of 0.7 to 1.30.

Determination of difficulty level using criteria according to Hambleton & Swaminathan, that a test item is categorized as very difficult if difficulty > +2 and categorized very easily if difficulty > -2. Therefore, the development of this test instrument is attempted to have a value of $-2 \leq b \leq +2$. The application of PCM scoring requires that the item difficulty level (delta) does not always have to be fulfilled $\delta\_ik < \delta\_i(k+1)$.

FINDINGS

Test instrument validation results

The mathematical representation ability test instrument based on terbang papat of local wisdom developed consisted of 5 items that had been validated by expert judgment, physics teachers, and peer reviewers. Validation analysis of the test instrument refers to the Aiken's V index table (R Aikens, 1985). The

results of the validation of the local wisdom-based mathematical representation ability test instrument can be seen in Table 2. Based on Table 2, it can be concluded that the test instrument for mathematical based on local wisdom of terbang papat ability can be stated as feasible to use.

Table 2. Results of Validation Instrument Mathematical Representation Ability Test

Item	Aiken's V	Information
1	0,86	Valid
2	0,95	Valid
3	0,95	Valid
4	0,86	Valid
5	0,90	Valid

Goodness of Fit

The overall goodness of fit test uses the rules developed by Adam and Khoo (1996, p.30) which are based on the average INFIT Mean of Square (Mean INFITMNSQ) value along with the standard deviation or observing the INFIT t value (Mean INFIT t) along with the standard deviation. Testing the suitability of instrument items is done for the overall test or for each item. Based on Table 3, the average value of INFITMNSQ 1.00 and standard deviation is 0.56, so the whole test is fit with the PCM model. In accordance with the rules developed by Adam and Khoo if the average INFIT MNSQ is around 1.0 and the standard deviation is 0.0 or the average INFIT is close to 0.0 and the standard deviation is 1.0, the overall test is fit with the PCM model.

Table 3. Results of Item Estimates and Tests of Mathematical Representation Ability Test Instruments Based on Local Wisdom of Terbang Papat

o.	Description	Estimates for items	Estimates for testi
	Average value and standard deviation	0.00 ± 0.09	0.32 ± 1.04
	Reliability	0.00	0.00
	Average value and standard deviation of INFIT MNSQ	1.00 ± 0.56	0.96 ± 0.95
	Average value and standard deviation of MNSQ OUTFIT	0.98 ± 0.30	0.98 ± 0.99
	Average value and standard deviation of INFIT t	-1.27 ± 7.85	-0.32 ± 1.64
	Average value and standard deviation of OUTFIT t	-0.82 ± 3.10	-0.12 ± 1.36

Item Characteristics Curve (ICC)

Item characteristics are indicated by item characteristic curves (ICC) and difficulty indexes. Based on the analysis obtained 5 item characteristic curves (ICC). In Figure 1, an example of ICC for item 1 is presented, which can be explained that: (a) score 1 (category 1) is mostly obtained by students with very low abilities (θ = -3), (b) score 2 (category 2) mostly obtained by students with low ability (θ = 2), (c) score 3 (category 3) mostly obtained by students with moderate ability (θ = 0), (d) score 4 (category 4) mostly obtained by students with high ability (θ = 1.5) and (e) score 5 (category 5) is mostly obtained by students with very high abilities (θ = -3). The level of difficulty from small to large sequential categories 1, 2, 3, 4 and 5.
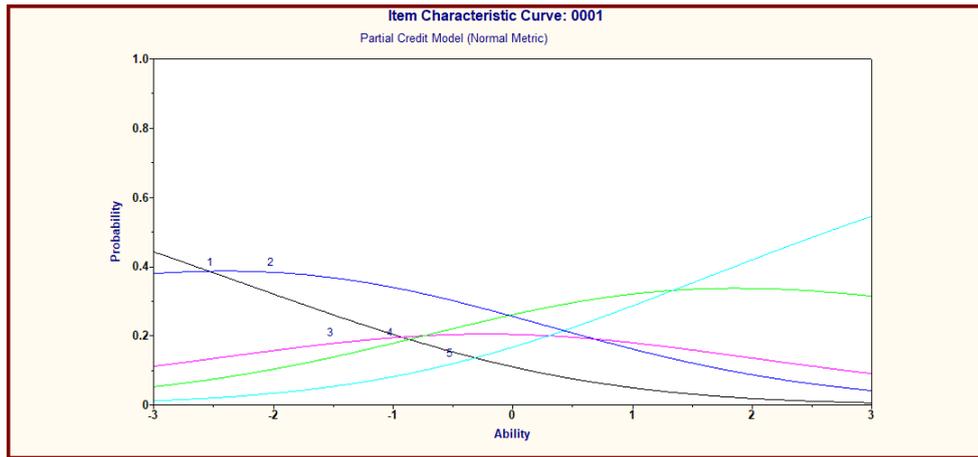
Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

Figure 3. Characteristic Curve item 1

Level of difficulty item (b)

The level of difficulty of the item is indicated by the THRESHOLD parameter on the output of the analysis of the parscale program. Difficulty index or difficulty level (b) for score 0 (b0), score 1 (b1), score 2 (b2), score 3 (b3), score 4 (b4), score 5 (b5), and mean difficulty. Each category has different difficulty levels for each item. This result agrees with the research of Grunert, Raker, Murphy, & Holme (2013) which states that the use of partial credit which is divided into several categories gives a significant impact on the item being tested. The results of the research in Figure 3 on the difficulty diagram show the difficulty level of each item. The difficulty value or the difficulty of the item is in the range of -2 to +2. This value is in accordance with the opinion of Fox & Bond (2015) which states that the level of difficulty for items with good categories is in the range of -2 to +2 (rounding from -1.96 to +1.96) with an error rate of 5%. Fox and Bond's opinion is supported by Hambleton & Swaminathan (1985) which shows that the item is said to be good if it has a level of difficulty from -2 to +2 units of logit.
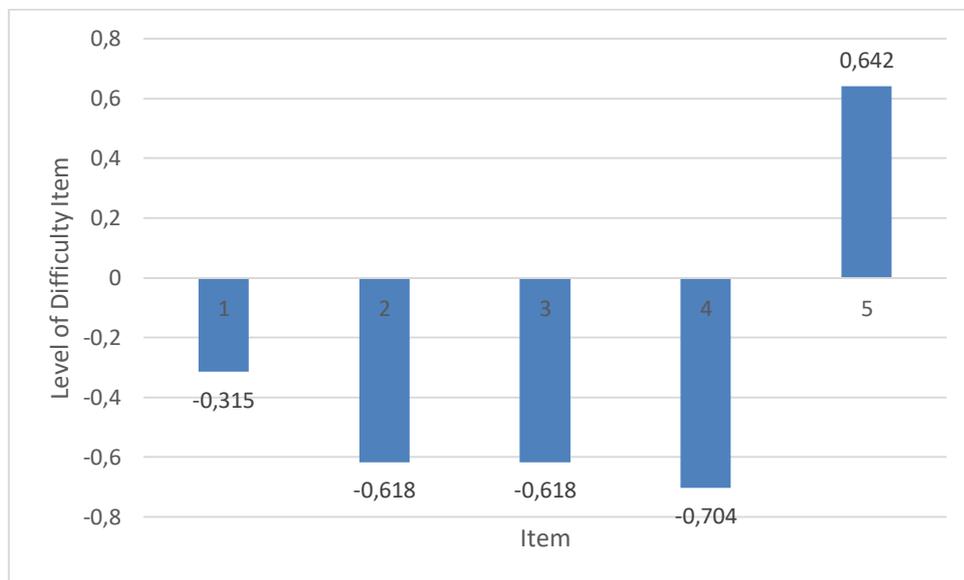

Figure 3. Level of Difficulty Item

The histogram of ability distribution can be seen in Figure 5. Based on Figure 5, the index distribution of difficulties follows the approaching normal distribution. This does not make a problem, because there is no requirement that the difficulty index distribution of items must be normal.
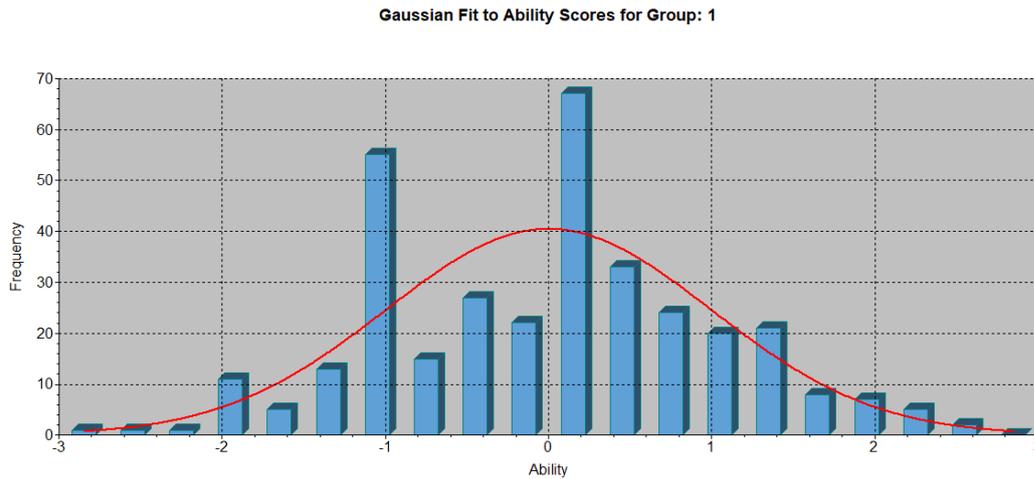
Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

Figure 4. Ability Distribution Histogram
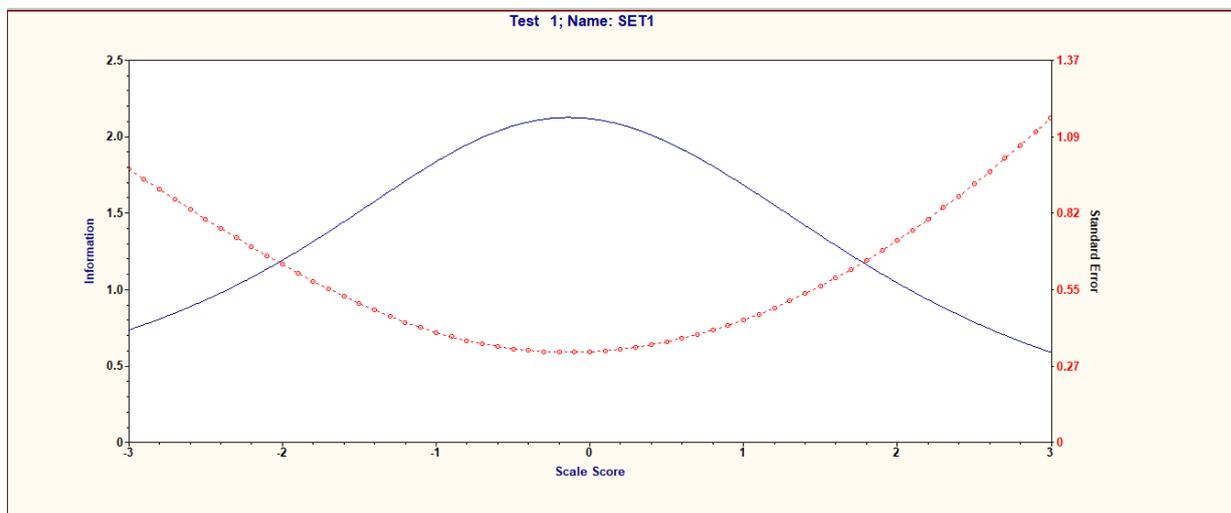
Information function and SEM

Figure 6. Information Function and SEM

Information functions and standard error measurement (SEM) were obtained based on analysis using the parscale program. Figure 5 shows a graph of total functions and SEM. The results of the analysis obtained intersections of information function lines and SEM lines at the point of -2.0 logit up to +1.8 logit. This value indicates that the mathematical representation ability test instrument that was developed reliably when tested on students with moderate (θ) ability is -2.0 <θ <+2.8 logit. The results of the analysis in Figure 5 prove that the test instruments developed have a medium category to measure students' mathematical representation abilities.

RESULT, DISCUSSION, AND SUGGESTIONS

The development of a test instrument for mathematical representation ability based on local wisdom of terbang papat can be 5 points, each of which represents aspects and indicators of mathematical representation ability. Mathematical representation ability test instruments based on local wisdom terbang papat all have met valid criteria. The content validation process is carried out by expert judgment, physics teacher and peer reviewer. While empirical validation is proven by the goodness of fit with the PCM model. According to Supahar, Zuhdan Kun Prasetyo (2015) there are four factors that influence validity and reliability, namely 1) test instruments developed according to the procedure for developing test instruments, 2) test instruments compiled from indicators obtained from measured aspects, 3) test instruments validated by expert judgment by involving a number of Education experts, 4) Respondents who took the test seriously.

Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

The construction of a test instrument for mathematical representation ability based on local wisdom terbang papat consists of four aspects of the ability of mathematical representation with four indicators as an assessment of the ability of mathematical representation. The four indicators of the ability of each mathematical representation are represented by five items. The results of the validation of the test instruments through expert judgment, physics teachers, and peer reviewers have met the content validation requirements.

The test instrument for mathematical representation ability based on local wisdom terbang papat can get the results of fit empirical evidence with a partial credit model based on five categories of polytomus data on 5 items that have been developed. Reliability of the test instruments has met reliable category requirements. Based on the total information function, the test instrument is very appropriate to be used to measure the ability of mathematical representation which has the ability between -2.00 to +1.8. The test instruments are in the criteria because the difficulty index is more than -2.0 or less than 2.0 ($-2.0 \leq b \leq +2.0$). The test instrument can be used to measure the ability of students' mathematical representation according to the partial credit model based on the five categories of polytomus data.

Simple practice tests can affect tests that require complex thinking if the test is associated with learning experiences (Asysyifa, D.S., Jumadi, Wilujeng.I. & Kuswanto, H., 2019). Learning experience can be applied in local wisdom-based test instruments. One learning experience is students doing project learning. Instruments to measure the ability of mathematical representation can also be used in project learning. Students playing terbang papat in groups during project learning to produce the desired tone. The limitations in this study are (1) the items measured are only limited to five aspects of the ability of mathematical representation with 5 items. (2) We only measure the quality of items without comparing with other variables such as gender. Further research is needed which uses a local wisdom-based test instrument to measure mathematical representation abilities. Thus, the ability of mathematical representation as part of problem solving skills still needs to be developed. In addition, quantitative and qualitative research that investigates why students have mathematical representation abilities and appropriate tests must be done. Discussions about the need for and quality test instruments often come to the fore.

REFERENCES

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

Asysyifa, D.S., Jumadi, Wilujeng.I. & Kuswanto, H. (2019). Analysis of student critical thingking skills using partial credit models (PCM) in physics learning. *International Journal of Educational Research Review, 4*(2), 245-253.

Cock, M. De. (2012). Representation use and strategy choice in physics problem solving. *Physics Education Research* 8, 020117(November), 1–15. https://doi.org/10.1103/PhysRevSTPER.8.020117

Depdiknas. (2007). Peraturan Menteri Pendidikan Nasional No 20 Tahun tentang Standar Penilaian

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physics Education Research ,10*, 1–58. https://doi.org/10.1103/PhysRevSTPER.10.020119

Hambleton dan Swaminathan (1991). *Fundamentals of item response theory*. California: SAGE Publications, Inc

Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *The Counseling Psychologist*, 27 (3), 353-383

Hogg R.V. & Craig, A.T. (1978). *Introduction to mathematical statistics*. New York: Macmillan Publishing Co Inc.

Kun, P. Z. (2013). Pembelajaran sains berbasis kearifan lokal. *PROSIDING : Seminar Nasional Fisika Dan Pendidikan Fisika*, 2(1), 246–256. Retrieved from http://jurnal.fkip.uns.ac.id/index.php/prosfis1/article/view/3316

Danis, A.O, Ahdika.S, Warsono & Supahar (2019). Analysis of Terbang Papat local wisdom-based test instruments to measure mathematical representation capabilities in learning physics. International Journal of Educational Research Review,4(4), 555-563.

Kurnaz, M. A., & Arslan, Ğ. L. A. M. (2014). Effectiveness of multiple representations for learning energy concepts : Case of Turkey. *Procedia - Social and Behavioral Sciences*, 116, 627-632

Masters, G.N. (1999). *Partial credit model*. Dalam J.P. Keeves & G.N. Masters (Eds.). Advances in Measurement in Educational Research and Assessment. Amsterdam: Pergamon.

Matthew, S.J. (2007). Marginal maximum likelihood estimation of item response model in R. *Journal of Statistical Software,20(*10),1-24.

Merriënboer, J. J. G. Van. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153–160.

Pospiech, G. (2012). Modelling mathematical reasoning in physics education. *Science & Education,21*(4), 485–506. https://doi.org/10.1007/s11191-011-9396-6

Supahar & Prasetyo, Zuhdan Kun. (2015). Pengembangan instrumen penilaian kinerja kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA. *Jurnal Penelitian dan Evaluasi Pendidikan Volume* 19, No 1, Juni 2015 (96-10)

Suryabrata, S. (2002). Pengembangan alat ukur psikologis. Yogyakarta: Andi Offset.

Taale, K. D. (2011). Improving physics problem solving skills of students of Somanya Senior High Secondary Technical School in the Yilo Krobo District of Eastern Region of Ghana. *Journal of Education and Practice*, 2(6), 8–21.

Wang, F.-H. (2006). *Application of componential IRT model for diagnostic test in a standard conformant elearning system*. In Sixth international conference on advanced learning technologies (ICALT'06). 5-7 July, Kerkrade, Netherlands.

Widhiarso W., (2010). Model politomi dalam teori respon butir. Fakultas Psikologi UGM, Yogyakarta