

Development of Vector Representation Test Instrument for Senior High School Students in Realizing Learning Outcomes

Puji Iman Nursuhud¹, Warsono², Supahar³, Rio Sandhika Darma⁴

ARTICLE INFO

Article History:

Received 05.03.2019

Received in revised form

10.06.2019

Accepted

Available online 01.10.2019

ABSTRACT

The study was conducted to develop the vector representation ability test items for senior high school students and obtain Item Curve Characteristics (ICC). Grid test instruments are arranged based on competencies and vector representation indicators which are then used to compile items. The test instrument consists of five question items related to vector representation which includes drawing vectors, determining the magnitude and direction of vectors and performing vector operations. The validated instrument was piloted in 296 students from three schools in Batang Regency, Central Java. Polytomus data were analyzed using the QUEST program for classical analysis and PARSCALE 4 for modern analysis theory based on Partial Credit Model (PCM). The results of the data analysis of the trial items indicate that of the five item questions, only 4 questions were fit with PCM. The reliability of the test instrument is 0.72, the level of difficulty of the items is in the range of 0.79 to 1.35. The information function and Standard Error Measurement (SEM) show that test items are developed reliably to measure the vector abilities of students with the average category in $-1,6 < \theta < +3,2$ logit scale with $SEM \pm 0,5$.

© IJERE. All rights reserved

Keywords:¹

Development of test instruments, vector representation ability, polytomous, PCM.

INTRODUCTION

Education is a means of improving the quality of human resources. Quality human resources prove that the level of education is quality (Baran, 2016). Quality education begins with a learning program that is systematically arranged according to the applicable learning curriculum. A good learning program starts through the preparation of the right learning goals (Guney & Al, 2012).

Learning is a process that occurs repeatedly where students are able to explore, discover and transform experiences that occur in the form of knowledge or a body knowledge (Suyono & Hariyanto, 2014). The Republic of Indonesia Minister of Education and Culture Regulation Number 22 Year 2016 concerning Basic and Secondary Education Process Standards states that the learning process must be conducted interactively, inspiring, interesting, fun and challenging and can motivate students to actively participate. Learning is said to be successful if students can understand the material concepts that are taught very well (Georgiou & Sharma, 2015). Learning activities require a systematic curriculum structure. The learning environment with structured pedagogical concepts, good curriculum design and a comfortable learning atmosphere can make students transform knowledge effectively (Guney & Al, 2012).

Physics is a learning activity that has the purpose of developing logical abilities and inductive and deductive analysis of students using physics concepts so that they can solve problems (BSNP, 2006). Physics focuses on qualitative or quantitative measurements in finding and discovering basic laws relating to phenomena and using them to develop theories (Serway R. & Jewett J., 2004). Baran (2016) states that learning physics provides the ability for someone to solve problems (problem solving) in learning.

Problem solving is the most important basic element in physics learning (Docktor & Mestre, 2014). Merriënboer (2013) suggested four stages of problem solving namely (1) studying the problems raised, (2) exploring and interpreting information with appropriate procedures, (3) looking for references that support to solve problems, (4) the process of trying to solve problems. Whereas according to Dostál (2015), analyzing problem solving must consider several things such as the ability to see problems, perception of problems, ability to solve problems and problem solving strategies.

¹ Corresponding e-mail: nursuhudpujiiman_007.2017@student.uny.ac.id, <https://orcid.org/0000-0002-0563-4649>

²; <https://orcid.org/0000-0003-1988-4881>

³; <https://orcid.org/0000-0002-2486-5549>

⁴; <https://orcid.org/0000-0003-1558-8284>

Universitas Negeri Yogyakarta^{1,2,3,4}

Problem solving strategies are very useful for solving problems in physics learning. Schoenfeld (2013) states that the process of finding a solution to a problem depends on the problem solving strategies used. The use of problem solving strategies must consider several things such as (1) identifying fundamental principles, (2) solving, and (3) checking (checking) (Gok, 2010). Docktor & Mestre (2014) explains that problem solving can be solved by applying representation as a solution strategy.

Vector is one form of representation that plays an important role as a means to understand the concept of physics (Barniol & Zavala, 2014). Vector is a basic concept in physics which is classified as difficult for students to understand (Maries & Singh, 2016). Nguyen & Meltzer (2010) explained that understanding of vectors is an inhibiting factor for students in understanding physics concepts. Problems that are often experienced in understanding vector concepts include (1) drawing vectors and unit notations, (2) determining vector direction, (3) determining vector size, (4) vector addition, (5) vector reduction, (6) vector multiplication with scalar, (7) multiplication of dot between two vectors, and (8) multiplication of cross two vectors (Barniol & Zavala, 2014). Wutchana, et al (2015) and Wutchana & Emarat (2011) explain that there are four aspects that must be considered in understanding the concept of vectors, including (1) knowing the quantity of vectors, (2) understanding that vectors can be moved with direction and magnitude the same, (3) knowing the sum of vectors by bringing together the ends and bases of the summed vectors, and (4) determining the resultant direction of the summed vector.

Understanding of vectors is determined by monitoring the process, progress and improvement of continuous learning outcomes so that an assessment is needed to measure the level of understanding of students' vectors. Educational assessment is the process of collecting and processing information to determine student learning outcomes (Minister of Education Regulation No. 20, 2007). Assessment in the world of education can use two kinds of measurement theories, namely classical theory and modern theory. The use of classical measurement theory in Indonesia to analyze and estimate students' abilities is more desirable than modern measurement theory (Fajrianti et al, 2016). However, classical measurement theory has a weakness in its use. The disadvantages of classical measurement theory include the characteristics of test items such as the level of difficulty and the power of differences that depend on students (Persichitte, 2016). Problems with classical measurement theory will have an impact on the level of ability of students that is difficult to know (Awopeju & Afolabi, 2016). Problems that arise in classical measurement theory can be solved by applying modern measurement theory, namely the item response theory approach (Item Response Theory, IRT) (Baker, 2001).

Item Response Theory (IRT) is a modern measurement theory that has the advantage of being able to find out the abilities and scores of students and have a more complex measurement model (Persichitte, 2016). DeMars (2010) explains that item response theory shows the relationship of ability or level trait measured using instruments and response points with a dichotomous or polytomus scoring model. The scoring model for dichotomous grains consists of: a) 1-PL model (Logistic Parameters) which emphasizes one parameter, namely the level of difficulty of the item, b) the 2-PL model which emphasizes two parameters, namely the level of grain difficulty and power difference, and c) the 3-PL model emphasizes three parameters, namely the level of difficulty of the item, different power and pseudo guessing (Mardapi, 2012). Scoring models for politomus items that are often used include the Graded Response Model (GRM), Modified Graded Response Model (MGRM), and Partial Credit Model (PCM) (Aybek & Demirtasli, 2017).

Partial Credit Model (PCM) is the development of a one-parameter logistic IRT model (1-PL) and is included in the Rasch model (Bacci et al, 2014). PCM is a politomus scoring model that uses several categories to analyze responses to an instrument (Master, 1997). For example, in a vector representation test instrument developed where the process for answering it requires several steps of completion. The Master in Linden (2016) explained that PCM is the easiest and most widely applied polytomus item scoring model to analyze tests and assessments such as measuring critical thinking skills, computer adaptive tests (CAT), measuring conceptual understanding in science and diagnosing mathematical errors. Grunert et al (2013) state that the PCM model is useful for knowing the level of conceptual understanding of students. The Partial Credit Model is an IRT analysis model that was developed with the aim of knowing the relationship of grain characteristics to the natural responses of students (ability or level trait). Fox & Bond (2015) states

that PCM specifically combines the number of different response levels for different items on the same test which can combine dikotomous and politomous items.

METHOD

This research includes the type of development research with a quantitative approach. This development research uses a 4-D development model (Define, Design, Develop, and Disseminate). The study began in November 2018 until February 2019. The development and preparation of vector 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 N 1 Batang, SMA N 2 Batang and SMA N 1 Bandar, Batang regency, Central Java.

Material

The steps to develop a test instrument follow the 4-D model which consists of: 1) Define (Defining Phase), 2) Design (Design Phase), 3) Develop (Development Phase), and 4) Disseminate (Distribution Stage). The defining stage includes: 1) determining the competency tested, 2) determining the material being tested, and 3) determining the indicator of vector representation. The design phase includes: 1) compiling the test grid and 2) arranging items according to vector representation indicators. The development phase includes: 1) validation of test items, 2) improvement of test items, and 3) preparation of scoring guidelines. The deployment phase includes: 1) the determination of the trial subject, 2) the implementation of the trial, and 3) the analysis of the results of the trial data. The stages of test development are presented in Figure 1.

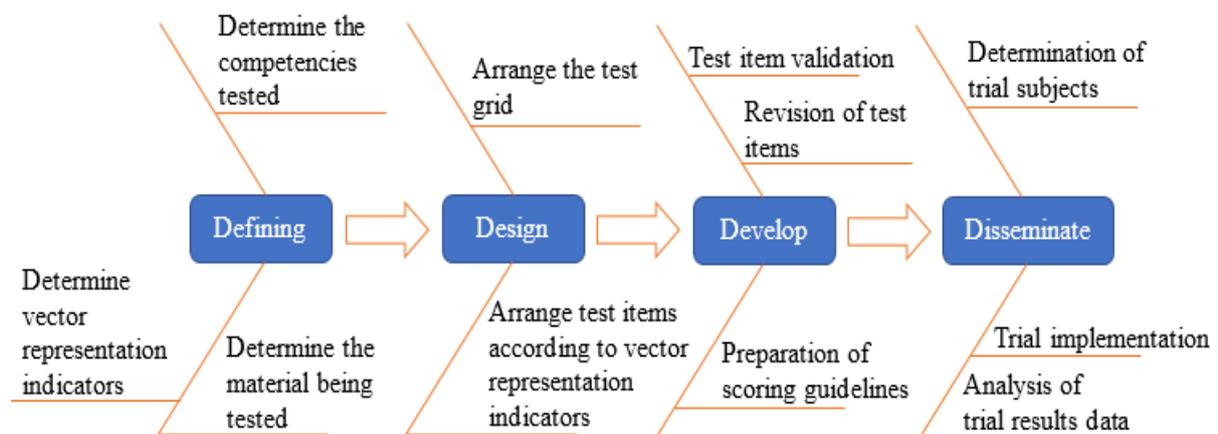


Figure 1. Steps for developing vector representation ability test instruments

The sample size used was 296 students. Fox & Bond (2015) stated that for analysis using the item response theory (IRT), a sample of between 30 and 300 people was used. While Reckase (2000) argues that the sample needed for analysis using 3-PL IRT which includes the level of difficulty, power difference and pseudoguessing is 300 people (Haladyna, 2004). The sample of this study was all students of class XI of the Science specialization program consisting of 3 classes in each school selected using the random sampling method. The number of research subjects was 296 students with 17 years old in the XI MIA Class of academic year of 2018-2019. The schools used for the trial include: SHS 1 Batang, SHS 2 Batang and SHS 1 Bandar, Central Java, Indonesia. The empirical trial was carried out by determining schools in Batang Regency based on national exam scores in physics in 2018 which were in the low, medium and high categories. So that by using the PCM 1-PL model, 296 students were sufficient as subjects for empirical trials.

Data Analyses

Analysis of the data from empirical trials using PCM 1-PL for fit test items for vector representation ability tests. PCM is used to analyze test items that implement a number of completion steps. Consideration of the analysis using PCM, namely 1) can use a sample that is not too large compared to the calibration using the 2-PL logistic model and 3-PL, 2) the response characteristics of the items follow PCM. The analysis

carried out includes: 1) Compatibility of vector representation test items, 2) reliability, 3) item characteristic curve, 4) item difficulty level, 5) item parameter estimation, 6) estimation of learners' abilities, 7) information function and standard error measurement (SEM).

Goodness of fit analysis is carried out to determine item compatibility with the partial credit model (PCM). Goodness of fit is analyzed by interpreting the average MNSQ INFIT value along with the standard deviation or the average INFIT t along with the standard deviation (Adam & Khoo, 1996). If the average INFIT MNSQ approaches 1.0 and the default deviation is 0.0 or the average INFIT t approaches 0.0 and the default deviation is 1.0 then the item is said to be fit with the model. The compatibility of the items with the model is known based on the INFIT MNSQ values in the range of values from 0.77 to 1.30 (Adam & Khoo, 1996). This value if converted using a standard value of t is in the range of -2 to +2 (rounding from -1.96 to +1.96) with an error rate of 5% (Fox & Bond, 2015). The item is said to be good if it has a level of difficulty from -2 to +2 units of logit (Hambleton & Swaminathan, 1985).

The QUEST and PARSCALE 4 programs are used to analyze the results of the trial data. Scores obtained by students were analyzed using the QUEST program to determine the suitability of items for the PCM model and reliability. PARSCALE 4 program is used to analyze data to show parameters of item characteristics such as: 1) item characteristic curve, 2) item parameter estimation, 3) estimation of student ability, 4) information function and standard error measurement (SEM).

FINDINGS

a. Results of Development of Vector Representation Ability Test

The test instruments developed amounted to 5 items in the description of momentum and impulse material. The test instrument was assessed for its feasibility by expert judgment before being used for the trial phase. Tables 1 and 2 show the distribution of test items the ability of vector representation and the results of validation expert judgment.

Table 1. Distribution of Test Item Vector Representation Capability

Vector Representation Indicator	Item Number	Material
Draw a vector	2	
Determine the magnitude and direction of the vector	1,3	Momentum and Impuls
Perform vector operations (addition, subtraction, dot product and cross product)	4,5	

Table 2. Test Item Validation Results based on Aiken's V

Representation	Question Item	Score of Aiken's V	Criteria
Vector	2	0,83	Valid
	1,3,4,5	0,83	Valid

The test instrument developed refers to vector representation indicators which are part of problem solving. Based on the Aiken's V analysis, the test instruments is qualified as valid instrument.

b. Match of the Goodness of Fit Test Item to the PCM Model

Overall testing of goodness of fit is done by analyzing the results of the trial test questions using the Quest program. Goodness of fit is tested according to the rules developed by Adam and Khoo (1996) by looking at the average value of INFIT MNSQ and its standard deviation or by observing the average value of INFIT t and its standard deviation. The test instrument is said to be fit with the 1-PL PCM model if the

average INFIT MNSQ value is around 1.0 and the standard deviation is 0.0 or the INFIT average value is around 0.0 and the default deviation is 1.0. Table 3 shows item and test estimates from the vector representation ability test instrument.

Table 3. Item Estimation and Test of the Test Instrument

Description	Item Estimation	Test Estimates
Average value and standard deviation	0,00 ± 0,61	-0,30 ± 0,86
Reliability	0,72	0,70
Average value and standard deviation of INFIT MNSQ	0,98 ± 0,22	1,01 ± 0,96
Average value and standard deviation of OUTFIT MNSQ	1,02 ± 0,25	1,02 ± 0,99
Average value and standard deviation of INFIT t	-0,32 ± 2,64	-0,19 ± 1,37
Average value and standard deviation of OUTFIT t	0,05 ± 2,18	-0,06 ± 1,12

Testing of goodness of fit for each item follows the rules developed by Adam and Khoo (1996) by looking at the INFIT MNSQ value of each item based on the output of the QUEST program. The item is declared fit or suitable for the model if the MNSQ INFIT value ranges from 0.77 to 1.30. In addition, items are also declared fit to the model if the INFIT t value is in the range of -2 to +2. Table 4 shows the INFIT MNSQ and INFIT values for each item.

Table 4. Distribution of INFIT MNSQ and INFIT t Each Test Item

Item	INFIT MNSQ	OUTFIT MNSQ	INFIT t	OUTFIT t
1	0,91	0,93	-1,10	-0,60
2	0,99	0,96	-0,10	-0,30
3	0,79	0,76	-2,80	-2,50
4	0,86	1,01	-1,70	0,10
5	1,35	1,43	4,10	3,50
Average	0,98	1,02	-0,30	0,00

Table 4 shows that the vector representation ability test items developed have MNSQ INFIT value ranges from 0.79 to 1.35.

c. Reliability

Reliability obtained based on analysis using the QUEST program is 0.72. The reliability value obtained has a medium category. This reliability value indicates that the vector representation ability test instrument developed is qualified as good instrument.

d. Item Characteristic Curve

Item characteristics are indicated by item characteristic curve (ICC). Analysis to find out the ICC was used by the PARSCALE 4 program. The analysis carried out was obtained as many as 5 items characteristic curves. Figure 2 presents ICC item number 1. The ICC chart in Figure 2 shows the opportunity for students to answer test items based on their abilities. Opportunities for students working on item number 1 are as follows: 1) category 1 is $\beta_1 = 0,79$, 2) category 2 is $\beta_2 = 0,31$, 3) category 3 is $\beta_3 = 0,34$, 4) category 4 is $\beta_4 = 0,25$, 5) category 5 is $\beta_5 = 0,78$.

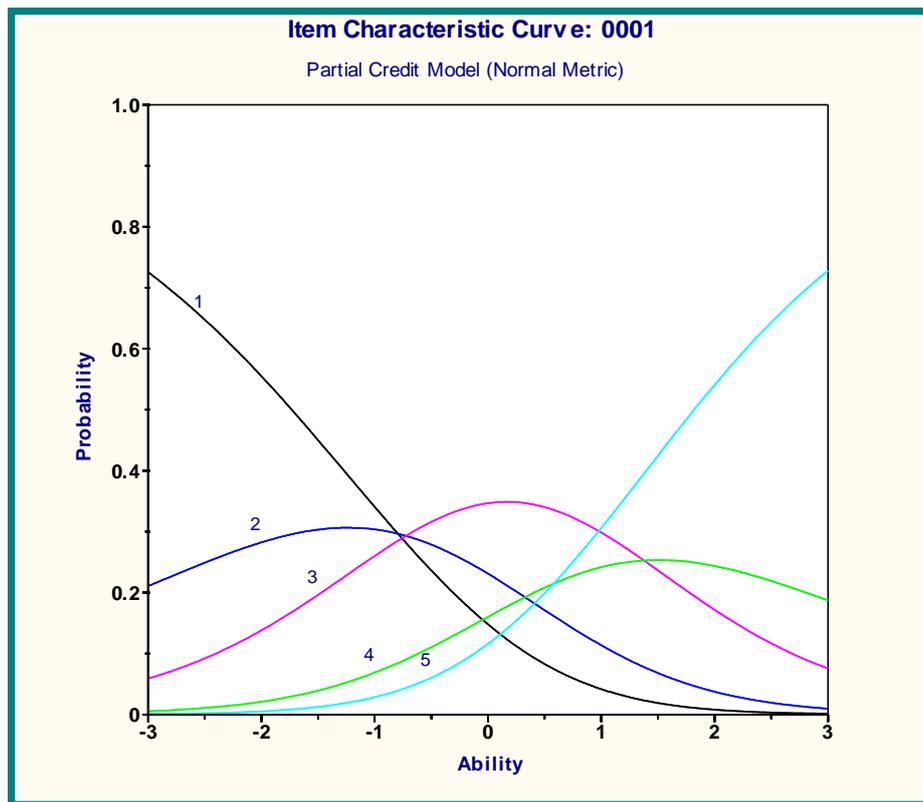


Figure 2. Characteristic curve number 1

ICC for item number 1 contains information as follows: 1) category 1 mostly obtained by students who have ability -3.5 logit scale. 2) category 2 is mostly obtained by students who have the ability of -1.2 logit scale. 3) category 3 is mostly obtained by students who have a capability of 0.2 logit scale. 4) category 4 mostly obtained by students who have the ability of 1.5 logit scale. 5) category 5 is mostly obtained by students who have the ability of 3.5 logit scale.

e. Item Parameter Estimation

The estimated parameters of vector representation ability test items according to the PCM model are indicated by different difficulty levels for each item. Table 5 shows a summary of parameter estimates analyzed using the PARSCALE 4 program.

Table 5. Test Item Parameter Estimation

PARAMETER	MEAN	STN DEV	N
SLOPE	0,561	0,000	5
LOG (SLOPE)	-0,578	0,000	5
THRESHOLD	0,804	0,787	5
GUESSING	0,000	0,000	0

The power estimation of different items is indicated by the SLOPE parameter which has an average value of 0.561. The level of difficulty of the item is indicated by the THRESHOLD parameter which has an average value of 0.804. The pseudo guessing parameter is shown by the GUESSING parameter which has a value of 0,000. Partial Credit Model (PCM) 1-PL refers to one parameter, namely the difficulty level of an item. Table 6 shows the difficulty level of each vector representation capability item for each score category in PCM.

Table 6. Level of Difficulty Test Item Vector Representation Ability

Item Number	Difficulty	Stage Difficulty				
		Category 1	Category 2	Category 3	Category 4	Category 5
1	-0,91	-2,62	-1,07	0,13	2,02	1,53
2	0,51	-2,82	0,32	-0,04	1,09	1,45
3	0,61	-2,37	-0,21	-0,40	1,52	1,46
4	-0,25	-1,46	0,49	0,14	0,39	0,44
5	0,04	-0,43	-1,12	-0,57	1,23	0,89

Table 6 shows that PCM measures the ability of students to work on test items based on the steps taken by dividing them into several categories. Each category has different difficulty levels for each item.

f. Estimated Ability of Learners

Estimating the level of ability of students is shown by the histogram. Figure 3 shows a histogram of students' vector representation abilities. Figure 3 shows that students' vector representation abilities follow a normal curve.

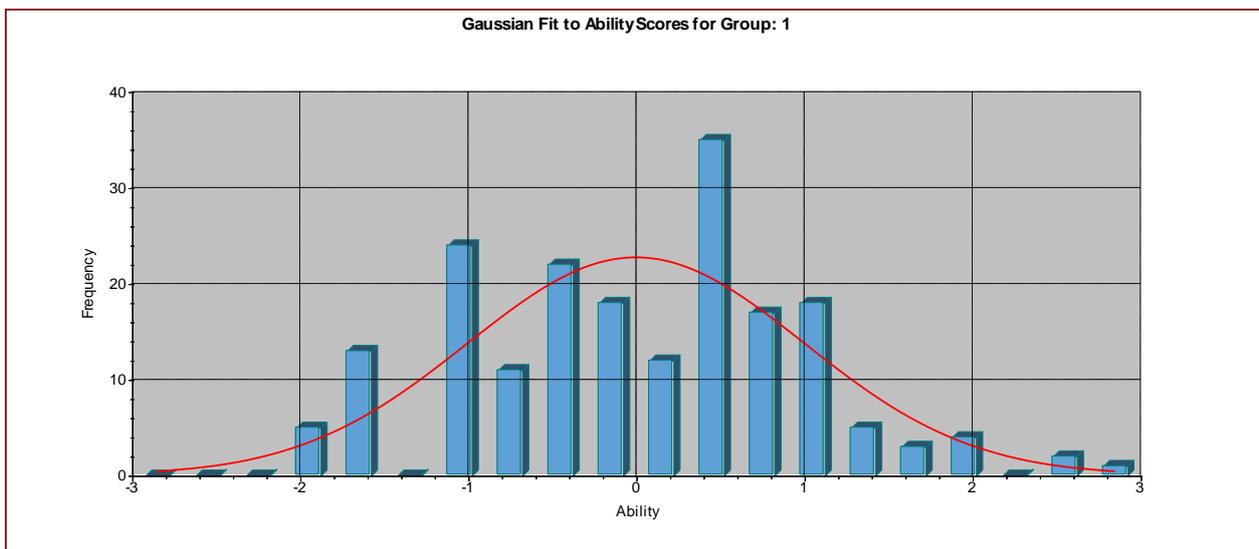


Figure 3. Histogram Estimated Vector Representation Capability

The histogram in Figure 3 can be interpreted by tabulating it into the frequency distribution table. Table 7 shows the interpretation of the ability of vector representation of students on a logit scale.

Table 7. Student Vector Representation Capability Category

Sample	Ability (Logit Scale)	Interpretation
3	+2,00 sampai +3,00	Very High
30	+1,00 sampai +2,00	High
115	-1,00 sampai +1,00	Medium
42	-2,00 sampai -1,00	Low
0	-3,00 sampai -2,00	Very Low

The results of data interpretation based on Table 7 show that there are no students who have very low vector representation abilities. Table 7 shows that 1,57% has a very high vector representation capability, 15,78% has a high vector representation ability, 60,52% has medium vector representation capabilities and 22,10% has a low vector representation ability. The results of the study in Table 7 show that the ability of vector representation of students is in the average to very high category.

g. Information Function and Standard Error Measurement (SEM)

Information functions and standard error measurement (SEM) were obtained based on analysis using the PARSCALE 4 program. Figure 4 shows a graph of total functions and SEM. The analysis results obtained intersection of information function lines and SEM lines at the point -1.6 up to +3.2 logit scale.

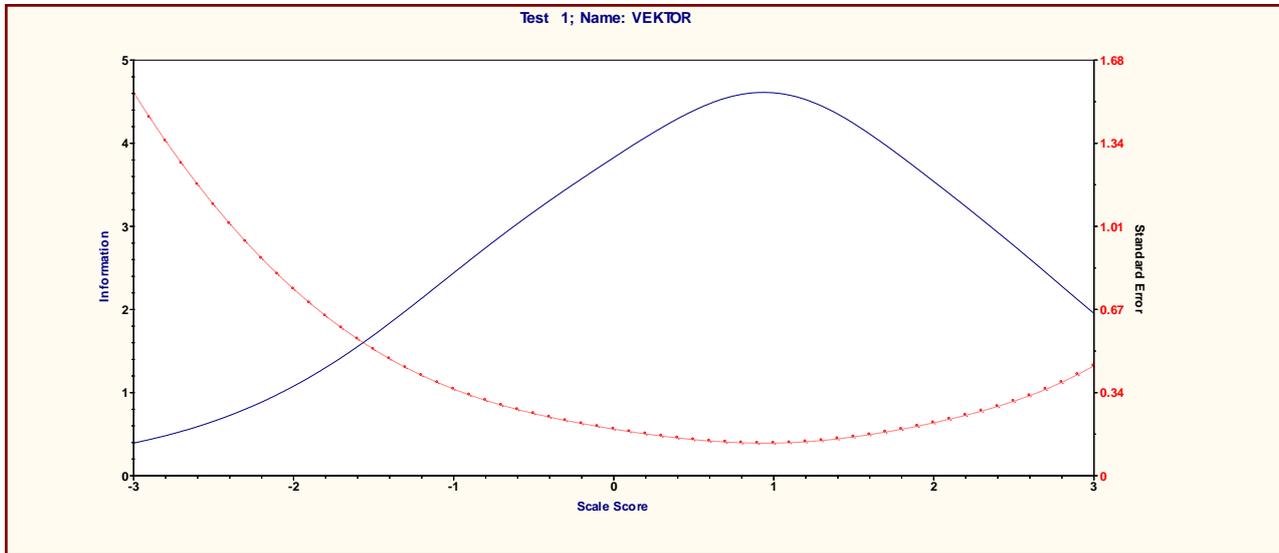


Figure 4. Information Function and Standard Error Measurement (SEM)

This value indicates that the vector representation ability test instrument that was developed reliably when tested on students with medium ability that is $-1,6 < \theta < +3,2$ logit scale with SEM $\pm 0,5$. These results indicate that the ability of vector representation of students is classified as medium.

RESULT, DISCUSSION, AND SUGGESTIONS

The test instruments developed amounted to 5 items in the description of momentum and impulse material. Test instruments are arranged and assembled in accordance with vector representation indicators which include: 1) describing vectors, 2) determining the magnitude and direction of vectors, 3) performing vector operations (addition, subtraction, dot product and cross product). The test instrument developed refers to vector representation indicators which are part of problem solving. Docktor & Mestre (2014) states that problem solving is an important strategy in solving physics problems. This is supported by the results of the Merriënboer (2013) study which suggested that there are four stages of problem solving namely (1) studying the problems raised, (2) exploring and interpreting information with appropriate procedures, (3) looking for references that support solving problems, (4) the process of trying to solve a problem. Vector representation indicators are developed with the aim of students being able to comprehend the concept as a whole so that they are able to apply it to solve physical problems, especially material momentum and impulses. Vector representation indicators developed include (1) drawing vectors and unit notations, (2) determining vector direction, (3) determining vector size, (4) vector addition, (5) vector reduction, (6) vector multiplication with scalar, (7) multiplication of dots between two vectors, and (8) cross multiplication of two vectors (Barniol & Zavala, 2014). This indicator is synthesized into 3 indicators, namely 1) describing vectors, 2) determining the magnitude and direction of the vector, 3) carrying out vector operations (addition, subtraction, dot product and cross product). The test instrument was assessed for its feasibility by expert judgment before being used for the trial phase. The feasibility of the test instrument was assessed based on material aspects and empirical tests (Yadiannur & Supahar, 2017). The results of the analysis of the material aspects using Aiken’s V show that the test instruments developed are in the valid category. This is in accordance with the validation criteria according to Aiken’s V (Aiken, L., 1985) which states that for 8 validators, items were declared valid if they obtained Aiken’s V score ≥ 0.75 .

Goodness of fit is tested according to the rules developed by Adam and Khoo (1996) by looking at the average value of INFIT MNSQ and its standard deviation or by observing the average value of INFIT t and

its standard deviation from output of QUEST program. The test instrument is said to be fit with the 1-PL PCM model if the average INFIT MNSQ value is around 1.0 and the standard deviation is 0.0 or the INFIT average value is around 0.0 and the default deviation is 1.0. The item is declared fit or suitable for the model if the MNSQ INFIT value ranges from 0.77 to 1.30. In addition, items are also declared fit to the model if the INFIT t value is in the range of -2 to +2. Table 4 shows that the vector representation ability test items developed have MNSQ INFIT value ranges from 0.79 to 1.35. But, the item test number 5 who developed has an INFIT MNSQ value outside the range of acceptance of goodness of fit that is equal to 1.35. This value indicates that item number 5 is not in accordance with the PCM model so it needs to be repaired or discarded. Item number 1 to 4 has MNSQ INFIT value which is in the range of acceptance of goodness of fit so it can be concluded that the four test items are suitable and suitable for the partial credit model (PCM). These results are as stated by Adam & Khoo (1996) which states that items are declared fit to the model if they have MNSQ INFIT values in the range 0.77 to 1.30. This opinion is supported by Fox and Bond (2015) that good test items have a range of INFIT values from -2 to +2. Reliability obtained based on analysis using the QUEST program is 0.72. This value obtained that the reliability of test instrument has a medium category. This reliability value indicates that the vector representation ability test instrument developed can be used to make decisions about students (Suryabrata, 2002).

Findings of the study shows that the item characteristics are indicated by item characteristic curve (ICC) in figure 2. Figure 2 presents ICC item number 1. The ICC chart in Figure 2 shows the opportunity for students to answer test items based on their abilities. Opportunities for students working on item number 1 are as follows: 1) category 1 is $\beta_1 = 0,79$, 2) category 2 is $\beta_2 = 0,31$, 3) category 3 is $\beta_3 = 0,34$, 4) category 4 is $\beta_4 = 0,25$, 5) category 5 is $\beta_5 = 0,78$. ICC for item number 1 contains information as follows: 1) category 1 mostly obtained by students who have ability -3.5 logit scale. 2) category 2 is mostly obtained by students who have the ability of -1.2 logit scale. 3) category 3 is mostly obtained by students who have a capability of 0.2 logit scale. 4) category 4 mostly obtained by students who have the ability of 1.5 logit scale. 5) category 5 is mostly obtained by students who have the ability of 3.5 logit scale. Politomus data scoring using PCM 1-PL produces characteristic curves such as Figure 2. This is as explained by Master (1999) that PCM is a politomus scoring model that uses several categories to analyze the response to an instrument. Politomus newspapers use partial credit divided into several categories. Categories are sorted from easy to difficult categories, namely 1,2,3,4,5. 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 categories on PCM specifically combine the number of different response levels for different items on the same test which can combine dichotomous and politomous items (Fox & Bond, 2015). The same thing is shown by Ayanwale (2018) that IRT approaches with PCM 1-PL in item analyze more better than (Classical Test Theory) CTT. Problems with classical test theory will have an impact on the level of ability of students that is difficult to know (Awopeju & Afolabi, 2016).

The estimated parameters of vector representation ability test items according to the PCM model are indicated by different difficulty levels for each item. Partial Credit Model (PCM) 1-PL refers to one parameter, namely the difficulty level of an item. Findings of the study shows that the difficulty level of each item vector representation ability divide for each score category in PCM. PCM measures the ability of students to work on test items based on the steps taken by dividing them into several categories. Each category has different difficulty levels for each item. This result agrees with the research of Grunert et al (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 Table 6 on the difficulty column 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 difficulty level from -2 up to +2 logit scale.

Estimating the level of ability of students is shown by Figure 3. Figure 3 shows a histogram of students' vector representation abilities. Table 7 shows the interpretation of the ability of vector representation of students on a logit scale based on histogram in Figure 3. Table 7 shows that there are no students who have very low vector representation abilities. There are 1,57% students has a very high vector

representation capability, 15,78% has a high vector representation ability, 60,52% has medium vector representation capabilities and 22,10% has a low vector representation ability. The results of the study in Table 7 show that the ability of vector representation of students is in the average to very high category. This proves that the vector representation ability test item developed is able to measure the level of ability of students. The results of this study agree with the Master's statement in Linden (2016) which explains that PCM is the easiest and most widely applied polytomus item scoring model to analyze tests and assessments such as measuring critical thinking skills, computer adaptive test (CAT), measuring conceptual understanding in science and diagnose mathematical errors. In addition, DeMars (2010) explains that the use of item response theory in assessment can show the relationship between ability or level trait measured using instruments and response items with dichotomous or polytomous scoring models. The same thing is shown by Diena (2019) that IRT can show relationship between ability measured using instruments with polytomous scoring models. Based on findings, IRT approach shown that vector representations ability of students in physics learning have a medium category. But, students are almost never drilled to apply vector representations test to solve problems in physics learning. So, the vector representations skill needs to be developed.

Findings of the study about information functions and standard error measurement (SEM) were obtained based on analysis using the PARSCALE 4 program. Figure 4 shows a graph of total functions and SEM. The analysis results obtained intersection of information function lines and SEM lines at the point -1.6 up to +3.2 logit scale. This value indicates that the vector representation ability test instrument that was developed reliably when tested on students with medium ability that is $-1,6 < \theta < +3,2$ logit scale with $SEM \pm 0,5$. These results indicate that the ability of vector representation of students is classified as medium. This agrees with the research of Maries & Singh (2016) which states that vectors are basic concepts in physics which are classified as difficult to understand by students. Whereas, Nguyen & Meltzer (2010) explained that understanding of vectors is a inhibiting factor for students in understanding the concepts of physics. But, students are almost never drilled to apply vector representations test to solve problems in physics learning. The results of the analysis in Figure 4 prove that the test instruments developed have a medium category for measuring the ability of vector representation of students. Thus, the vector representations skill needs to be developed.

ACKNOWLEDGEMENTS

Thank you to the Ministry of Research and Higher Education of Indonesia who has helped in financing this research.

REFERENCES

- Adams, R.J. & Khoo, S.T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research.
- Adekunle Ayanwale, M., Oluwatoyin Adeleke, J., & Iyabode Mamadelo, T. (2018). An assessment of item statistics estimates of basic education certificate examination through classical test theory and item response theory approach. *International Journal of Educational Research Review*, 3(3), 93–105. <https://doi.org/10.24331/ijere.452504>
- Aiken, L.R. (1985). *Three coefficients for analyzing the reliability and validity of ratings*. *Educational and Psychological Measurements*, (45), 131-142.
- Asyisyifa, D. S., 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>
- Awopeju & Afolabi. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *12(28)*, 263–284. <https://doi.org/10.19044/esj.2016.v12n28p263>

- Aybek, E. C., Demirtasli, R. N., & Computerized, R. N. (2017). Computerized adaptive test (Cat) applications and item response theory models for polytomous items : Computerized Adaptive Test (CAT) Applications and Item Response Theory Models for Polytomous Items. <https://doi.org/10.21890/ijres.327907>
- Bacci, S., Bartolucci, F., Gnaldi, M. (2014). *A class of Multidimensional Latent Class IRT models for ordinal polytomous item responses. (February)*. <https://doi.org/10.1080/03610926.2013.827718>
- Baker, F.B. (2001). *The basic of item response theory*. USA: ERIC Clearinghouse on Assessment and Evaluation.
- Baran, M. (2016). An analysis on high school students ' perceptions of physics courses in terms of gender (A Sample from Turkey). *Journal of Education and Training Studies* 4(3), 150–160. <https://doi.org/10.11114/jets.v4i3.1243>
- Barniol, P., & Zavala, G. (2014). Students ' difficulties in problems that involve unit-vector notation. *Latin American Journal of Physics Education*, 8(4).
- Bond, T.G, & Fox, C.M. (2015). *Applying the rasch model. Fundamental Measurement in the Human Sciences (3rd edition)*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika untuk SMA dan MA*. Jakarta: BSNP-Depdiknas.
- DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional No 20 Tahun 2007 tentang Standar Penilaian*.
- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics, *020119*, 1–58. <https://doi.org/10.1103/PhysRevSTPER.10.020119>
- Dostál, J. (2015). Theory of problem solving. *Procedia - Social and Behavioral Sciences*, 174, 2798–2805. <https://doi.org/10.1016/j.sbspro.2015.01.970>
- Fajrianthi., Hendriani, W., & Septarini, B.G.. (2016). Pengembangan tes berpikir kritis dengan pendekatan item response theory. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(1), p. 45–55.
- Georgiou, H., & Sharma, M. D. (2015). Does using active learning in thermodynamics lectures improve students' conceptual understanding and learning experiences? *European Journal of Physics*, 36(1), 015020. <https://doi.org/10.1088/0143-0807/36/1/015020>
- Gok, T. (2010). The general assessment of problem solving processes and metacognition in physics education. *Eurasian Journal of Physics and Chemistry Education*, 2(2), 110–122. <https://doi.org/10.1007/s11409-008-9026-0>
- Grunert, M. L., Raker, R., Murphy, K. L., & Holme, T. A. (n.d.). polytomous versus dichotomous scoring on multiple-choice examinations: Development of a rubric for rating partial credit. *J. Chem. Educ*, 90(10), 1310-1315.
- Guney, A., & Al, S. (2012). Effective learning environments in relation to different learning theories. *Procedia - Social and Behavioral Sciences*, 46, 2334–2338. <https://doi.org/10.1016/j.sbspro.2012.05.480>
- Haladyna, T.M. (2004). *Developing and validating multiple choice test items*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton, R.K, & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Maries, A., & Singh, C. (2016). *To use or not to use diagrams : The effect of drawing a diagram in solving introductory physics problems*. AIP Conf. Proc. 1513, 282.
- Masters, G.N. & Wright, B.D. (1997). *The partial credit model*. In: van der Linden w.j., Hambleton, R.K. (eds) *Handbook of Modern Item Response Theory*. New York: Springer.
- Merriënboer, J. J. G. Van. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64,

153–160. <https://doi.org/10.1016/j.compedu.2012.11.025>

Nguyen, N., & Meltzer, D. E. (2010). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics*, 71, 630–630(2003). <https://doi.org/10.1119/1.1571831>

Persichitte, K. A. (2016). *Educational technology to improve quality and access on a global scale*. Papers from the Educational Technology World Conference (ETWC 2016) Editors: Persichitte, Kay, Suparman, Atwi, Spector, Michael (Eds.)

Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*: 10(1).

Serway, R.A. & Jewett, J.W. (2012). *Physics for scientists and engineers (6th ed)*. United States: Brooks-Cole Publishing.

Suryabrata, S. (2002). *Pengembangan Alat Ukur Psikologis*. Yogyakarta: Andi Offset.

Wutchana, U., Bunrangsri, K., & Emarat, N. (2015). Teaching basic vector concepts : A Worksheet for the Recovery of Students ' Vector Understanding. *Eurasian J. Phys. & Chem. Educ.* 7(1): 18-28.

Wutchana, U., & Emarat, N. (2011). Students ' understanding of graphical vector addition in one and two dimensions. *Eurasian J. Phys. Chem. Educ.*, 3(2):102-111.

Yadiannur, M. (2017). Mobile learning based worked example in electric circuit (WEIEC) application to improve the high school students. *Electric Circuits Interpretation Ability*, 12(3), 539–558. <https://doi.org/10.12973/ijese.2017.1246p>