PE - 14

THE DEVELOPMENT OF PHYSICS ESSAY TEST FOR HIGHER ORDER THINKING SKILLS IN JUNIOR HIGH SCHOOL

Edi Istiyono

Physics Education Department, Mathematics dan Natural Science Faculty, Yogyakarta State University

email: edi_istiyono @uny.ac.id; edi_istiyono_uny@yahoo.co.id

Abstract

This research has been done to develop an instrument for measuring junior high school students' physics higher order thinking skills (PhysETHOTSS) and to obtain the characteristics of the PhysETHOTS. The instrument blue print has been developed based on the aspects and sub-aspects of higher order thinking skills, then it was used to develop the items. The instrument consisting of 24 items were validated by physics educational measurement experts. The validated instrument was tried out on students of junior high school (SMPN 1 Sewon). The polytomous data were analyzed according to the partial credit model (PCM). The results show that the 24 items of PhysETHOTS were fit to the PCM, the reliability of the test was 0.75, the items' difficulty indexes ranged from -1.22 to 0.34. Therefore, the PhysETHOTS is qualified for the measurement of junior high school students' physics higher order thinking skills.

Keywords: *instrument development, physics essay test of higher order thinking skills, polytomous, and PCM*

INTRODUCTION

Today the world is in an era of globalization that needs quite tight competition. In this era the competition is quite tight, the competition of human resources (HR). The quality of the nation's human resources is determined by the education level of the nation. Improving the quality of education can begin from improving the learning quality. Improving the learning quality can begin by setting appropriate learning objectives.

One of the aims Science learning in junior high school so that learners have the ability to develop reasoning skills in the analysis of inductive and deductive thinking using concepts and principles of physics to explain the events of nature and solving problems both qualitatively and quantitatively (BSNP, 2006: 160). Thus, through the study of physics students are expected to develop themselves in thinking. Learners are required not only have the ability of lower order thinking, but the higher order thinking skills (HOTS). With regard to the higher order thinking skills, the fact remains that the Indonesia physics achievement as measured on the reasoning aspect is ranked 40^{th} of 42 countries (TIMSS & PIRLS International Study Center, 2012:48). Thus, the physics achievement of Indonesian high school students that requires of HOTS in the international level is low. The low physics achievement can be caused by an improper learning process or assessment model. In this case, only assessment will be discussed, because proper assessment can encourage students to learn higher order thinking skills.

Based on Piaget's development theory, the formal operational stage is a stage of children beginning from eleven years old. At this stage the children have begun to develop the ability to manipulate abstract concepts through the use of propositions and hypotheses (Piaget, 2005: 122 and Reedal, 2010:7). The junior high school students are between 12 to 15 years, so that higher order thinking skills of junior high school students have been established.

The revised Bloom's taxonomy divide the cognitive aspect into lower order thinking skills (LOTS) and higher order thinking skills (HOTS). LOTS include the ability to remember, understand, and apply, while the HOTS include the ability to analyze, evaluate, and create

(Anderson and Krathwohl, 2001:30). Bloom's taxonomy has been applied in the education. Bloom's taxonomy is still used in many curricula and teaching materials (Brookhart, 2010: 39; Schraw and Robinson, 2011: 158-159). Thus the HOTS in physics includes physics abilities, i.e. analyzing, evaluating, and creating.

According to Brookhart (2010:5) higher order thinking skills (HOTS): (1) high-order thinking is at the top of Bloom's cognitive taxonomy, (2) The purpose behind teaching cognitive taxonomy to equip learners to transfer knowledge, (3) able to think meaning that learners are able to apply the knowledge and skills they developed while studying in a new context. In this case the term "new" is a concept application that has not been thought of before by learners, this means that the universal is not necessarily something new. Higher-lorder thinking means the ability of learners to connect learning with other things that have never been taught.

To monitor the process, progress, and improvement of students' learning outcomes on an ongoing basis, the necessary assessment. Educational assessment is the process of information collecting and processing to determine the achievement of students' learning outcomes (Regulation of the Minister of National Education, No. 20, 2007). Assessment can be done orally or in writing. Written assessment is conducted by a written test. There are two forms of the written tests, namely selecting and supplying the answers. Written test by selecting answers include: multiple choice, two- choice (true - false, yes - no), matching, and cause and effect.

Keep in mind that the evaluation model also effects the thinking skills of students. According to van den Berg (2008:15) that the curriculum has a rich potential for developing higher-order thinking skills of learners. Teachers have to plan well and engage learners in activities that encourage and develop the higher order thinking. Assessment can be implemented to bring students in improving their higher order thinking skills. This is supported by the other opinions, higher-order thinking questions that encourage students to think deeply about the subject matter (Barnett & Francis, 2012 : 209). Based on this argument means that the assessment, test of higher-order thinking skills, provide stimulation of students to develop high order thinking skills as well.

Nitko and Brookhart (2011:223) describe that the basic provisions of the assessment is the ability to use higher order thinking tasks that require the use of knowledge and skills in new situations. Must use new materials to assess the higher order thinking skills. One way done use sets of items that depend on the context .

There are disadvantages of multiple choice test, namely: (1) students chances to guess the answer is still quite large and (2) the students' thinking process cannot be seen exactly (Sujana, 1990:49). Therefore, essay test is good alternative test.

Assessment are based on the stages can be completed examinees. Although only just completed the initial stage, the examinees had to get the value. The highest value of course obtained when the examinee has completed all phases of the exam in point. The assessment procedure is the same with how individuals respond to the items in the psychological scale. For example, an item that provides four categories of response of 'never', 'rarely', 'often', and 'always ' analogous to the completion stage. Just about to finish the first stage is analogous to the category of 'never', while when it comes to the final stage, analogous to the category of 'always'. This assumption was later developed into a partial credit model (PCM). When it is assumed that a partial credit item then follow the pattern of higher ability individuals are expected to have higher scores than individuals who have a low ability (Widhiarsa, 2010: 6). According to Wright & Masters, PCM is also appropriate to analyze the response to the measurement of critical thinking and conceptual understanding in science (Linden & Hambleton, 1997: 101-102)

Based on the above, of various types of written test, essay test is a good one to explore the physics higher order thinking skills in junior high school. To measure higher-order thinking skills is used test that called Physics Essay Test for Higher Order Thinking (PhysETHOTS). So that we need to develop of physics essay test for higher order thinking skills (PhysETHOTS). Based on the description in the future, the goal is : (1) to develop an instrument for measuring junior high school students' physics higher order thinking skills (PhysETHOTS); and (2) to obtain the characteristics of the PhysETHOTS

RESEARCH METHOD

This research is the development research with quantitative approach. This instrument development research was done with the modified of the Wilson Model and Antonio Oriondo Model.

The test instrument development used a modified form of the Wilson and Antonio Oriondo model, consisting of: (1) the design of the test and (2) the test tryout. The test design phase included: (1) the determination of objective tests, (2) the determination of competency to be tested, (3) the determination of the tested material, (4) the preparation of test blue print, (5) the writting of items based on the principles of HOT test development, (6) the preparation the scoring guidelines, (7) test validation and (8) the repairing the items and assembling the test, The stages of the development of the test are presented in Figure 1. The try out included: (1) the establishment of try out subjects and (2) the implementation of the tryout.

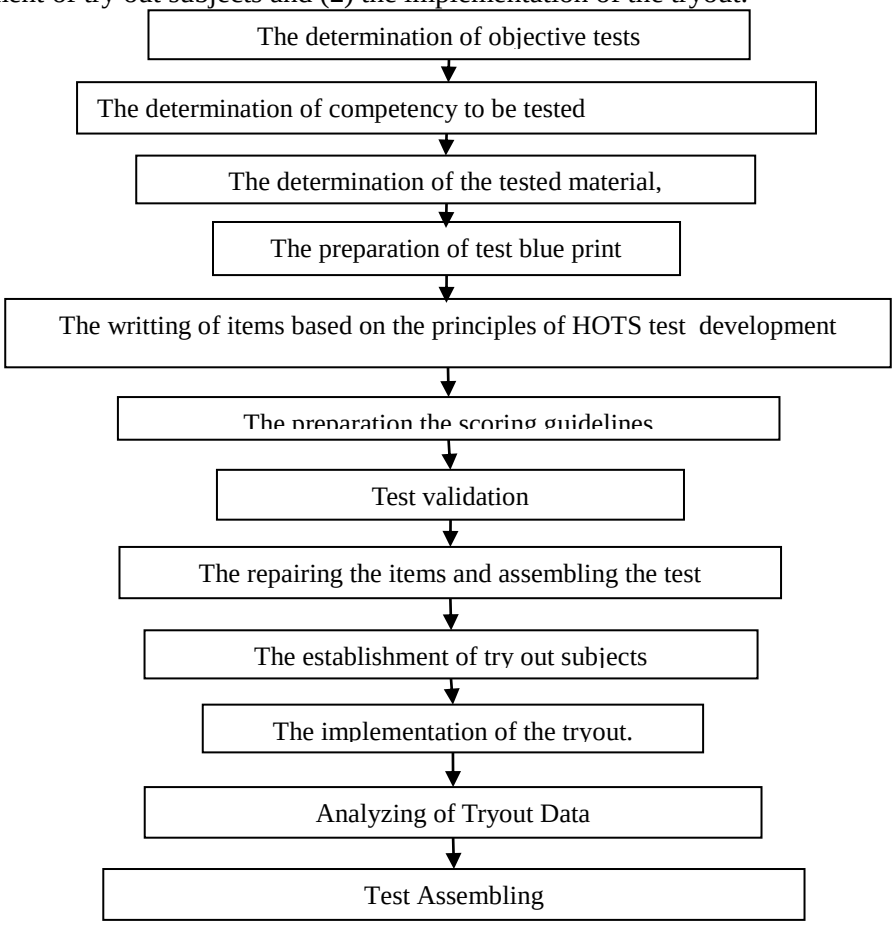

Figure 1. Steps of The Instrumen Development

Related to the sample number, according to some measurement experts IRT analysis requires 200 to 1000 people (Seon, 2009: 3). Reckase (2000) concluded that the minimum sample size for estimating the three parameters, which include discrimination, the difficulty index, and pseudoguessing, is 300 (Haladyna, 2004: 206). So with the PCM model of 1PL, the students for the tryout subjects as many as 100 are more than enough.

RESULT AND DISCUSSION

Result of The Test developed

The PhyTHOTS instrument consisted of 24 items, The test included sub physics matter: force, Newton's law, work and energy, simple machines, pressure, vibrations and waves, sound, light, and optical instruments and sub-aspect of HOTS: analyze, evaluate, and create. The items distribution is presented on Table 1. The PhyETHOTS was validated by experts judgment.

Description of the Physics instruments higher order thinking skills (PhysETHOTS) in JHS validation has been done then do the next step is try out. Tests conducted on 100 students in grade VII of SMP N 1 Sewon. The response of students then assessed and given a score on respondents (examinee). Score of the students are coded in note pad for analysis preparation.

Table 1. Distribution of Item PhyETHOTS in Grade VII of Junior High School

Dimension Kognitive Category	**Kognitive Process**	**Kompetency Standard and Physics Matter**								
		5. Understanding the role of work, force, and energy in daily life					6. Understanding the concept and application of vibration, waves and optics in daily technology products			
		Force	Newton's law	Work and energy	Simple machines	Pressure	Vibrations and waves	Sound	Light	Optical instrument
Analyze	Differenting	1					15	18		
	Organizing		4			12				23
	Attributing			7	10				21	
Evaluate	Checking	2		8				19		
	Critiquing		5			13	16			
Create	Generating			9	11				22	
	Planning	3	6				17			
	Producing					14		20		24

Goodness of fit of Instrument

Testing for goodness of fit for the overall test and each item is carried out. Testing goodness of fit the fit for the overall test developed Adam Khoo (1996:30) based on the mean value of INFIT Mean Square (Mean INFITMNSQ) and its standard deviation or average values INFIT t (Mean INFIT t) and its standard deviation. If the average INFIT MNSQ approximately 1 and 0.0 standard deviation or mean INFIT t close to 0 and standard deviation 1.0, then the

overall fit test with PCM. The INFITMNSQ is 1.01 (about 1) and a standard deviation is 0.13 (approximately 0.0), therefore the overall test fits with 1 PL PCM model.

Testing for goodness of fit Item and testee is determined that an item or testee is fit by models with boundary MNSQ INFIT range of 0.77 to 1.30. The MNSQ INFIT values of items between 0.78 to 1.27. Thus, 24 items are fit with PCM model.

Reliability

Based on the analysis, the reliability of the instrument (test) is estimated at 0.75. Reliability value is qualified as good instrument.

Item Characteristic Curva (ICC)

The characteristics of the item indicated by the curves characteristic of the item (ICC) and the index of difficulty. Based on the analysis, there were obtained item characteristic curves (ICCs) as many as 24 pieces. Figure 2 presented the characteristic curve item 1, that means: (1) score of 1 is largely for very low ability students ($\theta = -3$), (2) score 2 mostly to moderate ability students ($\theta = 0$), (3) score 3 mostly for high ability students ($\theta = 1$), (4) a score of 4 and 5 mostly for very high ability students ($\theta = 3$). The items' difficulty index from the small to the large ones sequential categories 1, 2, 3, and 4.

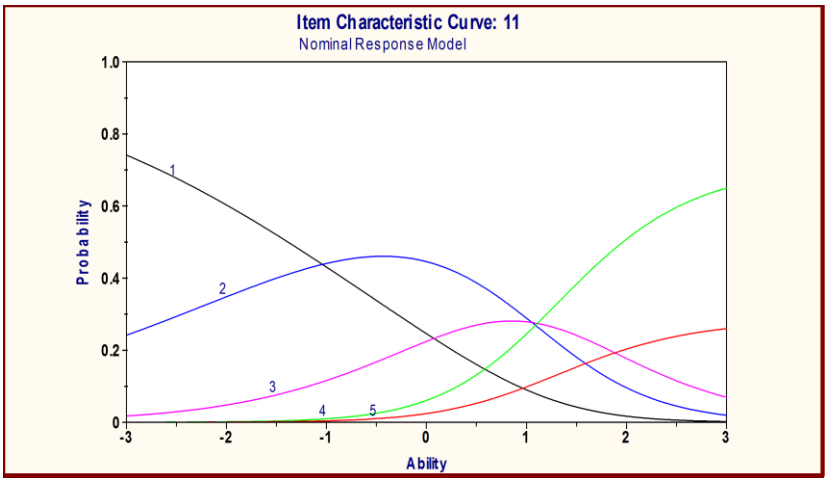

Figure 2. Item Characteristic Curve 11

The Difficulty Index

The items' difficulty index were from -1.22 to 0.34 with an average of 0 and a standard deviation of 0.35. So that based on difficulty ($-2.0 < b < 2.0$), all of 24 items were good. For more details, please see diagram distribution of items according to index difficulty and subaspek aspects of the instrument in Figure 3. Based on Figure 3, the order of item difficulty index of each aspect is analyzing, evaluating, and creating.

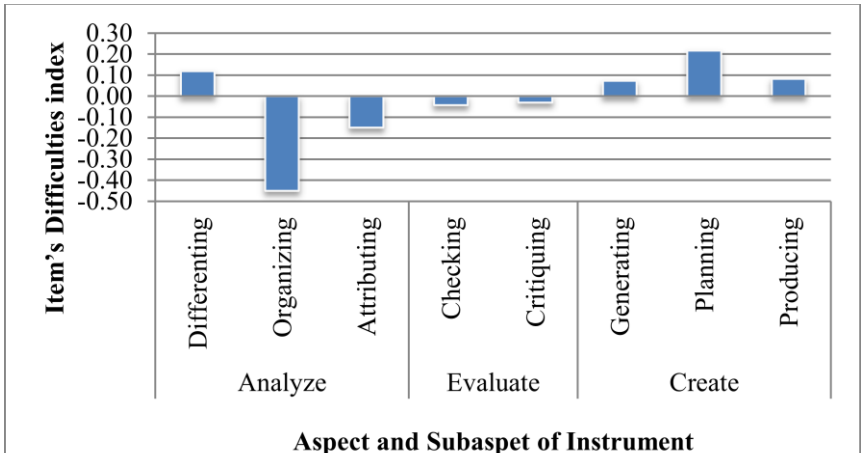

Figure 3. Item's difficulty Index of Each Aspect and Subaspect Instruments

Information Function and SEM

Based on the data analysis, it was obtained information and standard error of measurment (SEM). Based Functions information and SEM presented Figure 4, the test is suitable for the students that whose ability (θ) is high, ie $1 \leq \theta \leq 2.8$. This is consistent with the purpose of the developed instrument to measure Physics higher-order thinking skills.

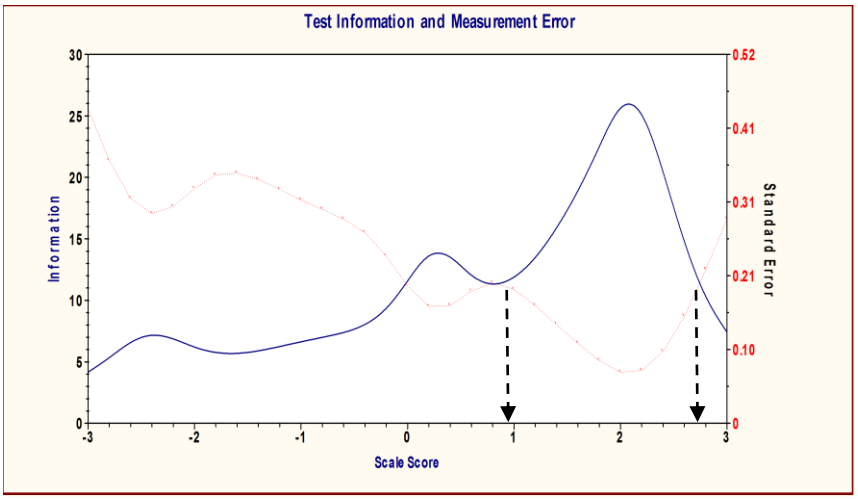

Figure 4. Information Function and SEM

Discussion

PhysETHOTS reliability is 0.75 that mean the test is good. It is said to be good, if the test has a reliability coefficient of more than 0.65 (Mchrens & Lehman (1991: 263). In addition, the information function is relatively for high ability between 1.0 to 28. This means that this instrument has high strength and reliable because it is composed of items that have high information function (Hambleton and Swaminathan, 1985: 94). Based on the reliability coefficient, the test information functions, and parameter estimation, this means PhysETHOTS

realiabel and has high stability.

Content validity of the test has been proven by expert judgment. Empirically verified the validity of the goodness of fit of the partial credit model (PCM). Based on Table 2, the average value and the standard deviation INFIT MNSQ 1.01 each (about 1) and 0.13 (about 0) , then the fit test with 1 PL PCM. This means that the test empirically valid. This is supported by all the items have a value between 0.78 INFIT MNSQ up with 1.27 which lies between the limits of receipt of the item using INFIT MNSQ or fit according to the model (between 0.77 to 1.30) means that all items fit many as 24 items of all. This is caused by several things, among others: (1) the items were developed according to the procedure correct item development instruments, (2) the items were developed from indicators derived from aspects of high order thinking skills and materials physics, (3) test which consists of 24 items that has tested the content validity by expert judgment, and (4) the respondents (students) were tested in earnest in doing because it involves supervisors of their physics teacher.

According to Hambleton & Swaminathan (1985:36), the item's difficulty index are good if they varied between 2.00 to 2.00. Items whose difficulty index of -2.00 indicates this is very easy, while the difficulty index of 2.00 means that the item is very difficult. Thus, based on the item's difficulty index the instruments (from -1.22 to 0.34) are good.

CONCLUSION AND SUGESSTION

Conclusion

Based on the analysis, the conclusions are as follows:

1. PhysETHOTS instrument was developed on junior high school students' abilities to analyze, evaluate, and create and on sub physics matter: force, Newton's law, work and energy, simple machines, pressure, vibrations and waves, sound, light, and optical instruments. The PhysETHOTS instrument is essay test that consisted of 24 items.
2. Characteristic of PhysETHOTS are:
   a. PhysETHOTS has content validity provided by expert judgment and empirical evidence has been getting fit with Partial Credit Model (PCM) based on polytomous data five categories.
   b. All items of PhysETHOTS on the criteria well as the difficulty index is in the range between 2.00 to 2.00.
   c. PhysETHOTS reliability is qualified
   d. Based on the information function and SEM, PhysETHOTS is very appropriately used to measure students' higher order thinking skills physics of 1.0 to 2.8.

Sugesstion

Based on the analysis, it are recommended:

1. Teachers can implement physics tests of high order thinking skills in junior high school.
2. Training for the development of physics test of higher order thinking skills is required for teachers.
3. Further research can be done using the data analysis by generalized partial polytomus credit model (GPCM 3PL).

REFERENCES

Anderson, L.W., & Krathwohl, D.R. (2001). *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Barnett, J. E & Francis, A.L. (2012). *Using higher order thinking questionsto foster critical thinking: a classroom study.* Educational Psychology: An International Journal of Experimental Educational Psychology. http://www.tandfonline.com/loi/cedp20. Diakses tanggal 10 Desember 2012

Bloom, B.S., et al. (1979). *Taxonomy of Educational Objectives: Handbook I Cognitive Domain*. London:Longmans Group Ltd.

Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom.* Alexandria: ASCD

BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP-Depdiknas

Depdiknas. (2003). *Undang-undang No. 20 tahun 2003 tentang Sistem Pendidikan Nasional*

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

Haladyna, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*. Boston: Allyn and Bacon

Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New Jersey: Lawrence Erlbaum Associates, Inc.

Hambleton & Swaminathan (1991): *Fundamentals of Item Response Theory*. Los Angeles: SAGE Publications, Inc

Mchrens, W.A & Lehman, I. J. (1991). *Measuement and Evaluation in Education and Psycology.* Belmont: Wadsworth

Sudjana, N. (1990). *Penilaian Hasil Blajar Mengajar.* Bandung: PT Remaja Rosdakarya

Oriondo, L.L.and Dallo-Antonio, E.M. (1998). *Evaluation Educational Outcomes*. Manila: Rex Printing Compagny, inc

Piaget, J. (2005). *The psychology of intellegence* [Versi elektronik]. Taylor & Francis e-Library.

Reedal, K.E. (2010). *Jean Piaget's Cognitive Development Theory in Mathematics Education*. Department of Mathematics and Computer Science – Ripon College. Summation, May 2010, pp. 16-20 http://ripon.edu/macs/summation.

Schraw , G, & Robinson, D.H. (2011). *Assessment of Higher Order Thinking Skills*. New York: Information Age Publishing, Inc

TIMSS & PIRLS International Study Center. (2012). *TIMSS 2011 international results in science.* Boston: The TIMSS & PIRLS International Study Center, Boston College. Diambil tanggal 5 Januari 2013, dari http: timss.bc.edu/timss2011/release.html

Van der Linden, Wim J & Hambleton, Ronald K. (1997). *Handbook of Modern Item Response Theory* . New York: Springer-Verlag New York, Inc

Widhiarso, W. (2010). *Model politomi dalam teori respons butir.* Yogyakarta: Psikologi UGM

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.