ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 295

1st International Conference on Educational Sciences and Teacher Profession (ICETeP 2018)

An assessment of Higher Order Thinking Skills (HOTS) Based on Rasch Models of Student in Physics Learning

Nirwana Nirwana, Syaiful Rochman, Alif Yanuar Zukmadini
Department Mathematics and Science Education
University of Bengkulu
Bengkulu, Indonesia
srochman@unib.ac.id

***Abstract*—This study I am instrument used for high-level learning at senior high school. Type of research is development of three stages, namely: initial development of instruments, content and empirical validation, and measurement. Instruments that have been validated by experts, tested on 95 students. The next instrument that has been applied to measure the ability of students. The number of students who use the total number of sampling is 148 students. The measurement results in this study were data analyzed by Rasch model using program Winsteps 3.37. This measurement is a model of suitability, degree of difficulty of grain, and characteristic curve. The results of this study indicate that the items are valid in content and then empirical. The empirical results showed that the MNSQ Outfit averaged 1.01 ± 0.22, ZSTD Outfit 0.1 ± 1.2. Referring to the lowest criteria of 0.5 <MNSQ <1.5 and 2.0 <ZSTD 2.0, 50 test items proved suitable. Test reliability shows good consistency at the trial and measurement stages, namely 0.92 and 0.95. Difficulty level of test items in the range of -1.50 to 1.04 which means that the level of difficulty of the test items is good. With that in mind, this instrument is used to measure the ability of high-level students in high school.**

Keywords—measurement; physics

I. INTRODUCTION

Human Resources (HR) in the 21st century is required to have quality that is able to compete globally. Anticipating these demands, Indonesia prepares education that will produce high-quality human resources. One method used is to improve the quality of learning. Formal learning activities in Indonesia occur in schools. The teacher provides learning activities that support the school to achieve the desired learning goals. One given by teachers in schools to improve is to teach students about ways of thinking.

The Teacher Training and Education Faculty (FKIP), University of Bengkulu has a vision to produce professional teachers who are able to compete nationally and internationally. Teachers are individuals who will develop scientific skills interactively, inspirational, fun, challenging, motivating students, creative and independent according to the physical and psychological development of students through learning activities. To achieve this goal, various efforts were made to improve the learning process in FKIP, including the physics education study program. One of the efforts made is to improve learning activities to improve graduates. One of the goals of this improvement is to create learning that is more fun, motivating, and improves the quality and results of learning. Based on the observations made by the researcher while teaching physics courses shows indicators, namely: (1) prominent students in the class are the same individual at each meeting, (2) skilled students solve questions in each meeting do not care about other students, (3) students have difficulty translating a problem, (4) students are not able to describe and analyze problems, and (5) students cannot solve problems without examples. This shows that the ability of the learning process and results is factually low.

Based on observations in classroom learning, student science learning outcomes achieved still need to be improved. Test scores obtained in the 2016/2017 school year.

TABLE I. VALUE DISTRIBUTION OF 2016/2017 SCIENCE EDUCATION STUDENTS

No	Range Value	Number of Student
1	50-59	6
2	60-69	7
3	70-79	20
4	80-89	3
5	90-99	1

Based on the results obtained by students, if it is converted, it can be seen that the IPA value that gets an A is less than 50%. In addition, another problem that arises is the ability of students to achieve thinking. Students should have high-level thinking skills. The ability to think is one of them functions as a selective activity in receiving information. Based on these problems, science learning needs to be improved.

Low order thinking habits taught to students cause students not to have high order ranking skills (HOTS). HOTS is needed to solve problems creatively and innovatively. To overcome this problem, students need to be directed to improve HOTS abilities. In the context of developing physics, students need to be accustomed to using HOTS.

The most important element in learning physics is supporting and motivating students to learn actively [1]. Thinking can connect from one concept to another with a series of thinking, speaking, reading, writing, seeing, listening, and counting [2].

HOTS as the transfer process of a problem then the problem is to find a solution using critical thinking [3]. Anderson and Krathwohl revise the cognitive level into two, namely [4]; Low order thinking method is at the level of remembering (C1), understanding (C2), and applying (C3), while HOTS thinking is at the level of analyzing (C4), evaluating (C5), and creating (C6). HOTS is divided into four groups, namely solving problems, making decisions, thinking critically, and thinking creatively [5]. HOTS will occur when individuals receive foreign information and "call" old information stored in memory.

Analyzing is a process that involves the part and structure of thought as a whole to solve problems. Analyzing includes cognitive processes attributing, distinguishing, and organizing [6]. Distinguishing is one of the abilities to choose the relevant part of a particular structure. When a process of differentiating occurs in students there will be attention and focusing between relevant and irrelevant material. Distinguishing ability can be seen from the test questions which are made in the form of short answers or choices. In a short answer, students are given a mathematical sentence and asked to show which parts are the most important or relevant [7]. Organizing can also be said as a form of integrating or structuring. When there is a process of organizing students will construct a systematic sequence of each piece of information. The format of assessment in organizing can be either a choice or a short answer. In the choice questions, students are asked to choose one truth from the four organizational structures that best suit the organization presented. With this process, students are easier to identify and the introduction of structures will become easier [8].

HOTS profiles of students can be known by testing students in terms of solving problems presented in the form of tests [9]. For this reason, questions are needed including the Higher-Level Question (HLQ). "If one persistently teaches for enhancing higher-order thinking skills, there are chances for success", so the assessment can provide stimulation to students in developing HOTS abilities [10]. An assessment is an activity to obtain information needed to make decisions about students, curriculum systems, programs in schools, and certain policies [11].

The test provides stimulation to students which will then produce a response. If the assessment only measures low-level thinking skills, then students only work on low-level thinking skills. Very much different if the assessment applied requires high-level thinking students, so students will be motivated to develop high-level thinking skills. Therefore, an assessment instrument for the ability of high-level thinking physics is needed. Assessment instruments in high-level physics thinking skills can be done in various forms, but to better understand the flow of thinking of students, reasoned multiple choice tests can be used. Therefore, in this study using reasoned multiple choice to find out and describe the high-level thinking skills of physics in Bengkulu Tengah district. Finally, from the results of this study, teachers, especially physics teachers, can develop tests that measure students' high-level thinking abilities.

In order for a high-level thinking assessment instrument to be used properly, it must meet the criteria of the instrument, including: validity and reliability. Content validity and construct validity are criteria that this instrument must fulfill.

II. METHOD

A. Types of Research

This research is a development research. Research that reveals and describes an event that occurs in the present condition. The description revealed was the students' high-level thinking ability regarding high-level thinking skills in physics in Central Bengkulu Regency. The approach that will be used in this descriptive research is a quantitative approach.

B. Research Subject

The subjects of this study were high school students in Bengkulu Tengah Regency, class XI IPA 2014/2015 academic year. The total number of high schools in Bengkulu Tengah is eight schools in total. Of the eight schools, five sample schools were taken by purposive sampling. The number of respondents for the empirical trial in this study were four classes with the number of students 95. The number of respondents for measurement was 148 students.

C. Procedure

The stages carried out in this study adapted Wallace's procedure which consisted of five stages: (1) theory study, (2) research questions, (3) making test items, (4) measurement, and (5) conclusions.

Making test items is done because there is no instrument that is in accordance with the needs of the researcher. The design of making test items to be carried out is (1) designing test instruments, (2) validating test instrument experts, and (3) empirical validation of test instruments. The design of the test consists of: designing test kits and content validation tests by experts, if there are items that have not fulfilled the requirements, revisions must be made before the test is valid in content. After the test design is complete, the test instrument is empirically validated.

D. Techniques and Data Collection Instruments

This research data is data. The data obtained in the form of the results of the validation of material experts and expert evaluations, and data response responses of students. Data from expert validation results were obtained through a validation sheet, the response data of students' answers was obtained when students answered items about the ability to think at a high level of physics. The instrument used in this study was a test. The test instrument is in the form of test questions as a measure of students' abilities regarding the ability to think at a high level of physics.

E. Data Analysis Technique

The collected data will be inputted into Excel, then changed in the notepad format. Data that has been collected along with answers are given scores with four categories, namely 1, 2, 3, or 4. In order to easily do the analysis, for example the data is arranged as follows.

AAAAAAAA …… 4424244444 …….
BBBBBBBBB …… .4442443333 …….
CCCCCCCC …… ..1121144444 …….
EEEEEEEEEE …… ..4241324311 ……
FFFFFFFFFF ... ... 2442443223 ……

Based on the format of notepad, letters in the form of "A, B, C, D, and so on" are the code of students. Signs in the form of numbers are the results of scores obtained by students. Determination of scores of students refers to polytomous scoring with four categories, namely: Score 4, if students are able to answer the choice correctly and be able to give the right choice of reasons, Score 3, if students are unable to answer the choice correctly and able give the right choice of reasons, Score 2, if students are able to answer the answer choices correctly and are unable to give the right choice of reasons, and Score 1, if students are able to answer the answer choices correctly and are unable to give the right choice of reasons .

Data analysis in this study sought validation, reliability, level of difficulty, and normality in the empirical trial. To facilitate the analysis, analyzed using the help of the Microsoft Excel Office 2010 computer program, SPSS, and winstep3.73.

Qualitative data analysis is carried out by expert validators regarding test items that are viewed in terms of material, construction, and language. Content validation is done before the test item is used. Corrected student answer sheets will be analyzed quantitatively.

The validity of the instrument in this study was carried out by construct validation of content validation, language validation, and empirical validation. Construct validation is done by making an instrument grid and then used the expert judgment technique. Validation of test items was carried out by two experts in the field of physics education. Improper test items will be corrected, after being corrected, empirical validation will be carried out.

The reliability of a test shows the determination or accuracy of an evaluation tool. The purpose of measuring the reliability of the test is to find out the degree of stability, consistency and stability of the data or findings on the size of the measured [12]. Reliability in match winsteps output seen from reliability items with criteria: (1) classified as weak if reliability items <0.67, classified as sufficient if the reliability items are 0.67-0.80, are good if the reliability items are 0.81- 0.90, classified as excellent if the reliability item is 0.91-0.94, and is classified as special if the item is reliable> 0.94.

The level of difficulty is a characteristic that will be used in determining the index of difficulty. By utilizing the analysis using the Winsteps program, the index of difficulty is obtained by means of a measure on the appearance of Winsteps. The conclusion in looking at the level of difficulty refers to the question is considered good, that is, the moderate questions are questions that have a difficulty index of 0.30 to 0.07 ".

The analysis of the research data obtained the average ability of the students as a whole, the maximum score, and the minimum score. The average value is used to calculate the average score obtained by students. The average calculation of students uses the following equation:

$$\mu = \frac{\sum x}{N}$$

Based on these equations, it can be explained that is the average, is the number of scores obtained by students, and N is the number of students. Data analysis measuring the ability of students was analyzed using the Microsoft Excel program which will later be divided into five categories of criteria. Determination of graduation standards or standards for achieving competency using standards adapted from tanwey is 60%, then the reference can be used 90%% x then "A (very high)", 75% ≤ x <90 then "B (high)", 60 % ≤ x <75% then "C (medium)", 40% ≤ x <60% then "D (low)", and if x <40% then "E (very low)".

III. Result and Discussion

The results of expert validation analysis on test items show that the test items used have met the standards of competence and basic competency. Material criteria, constructively, and in language on this instrument have been fulfilled. This shows that the test used fulfills good content validity. By specification, there are 14 items that do not meet criteria because the correlation between the answers to the reasons for choosing an illogical answer. Seven test items were not yet based on construction criteria, and seven test items were not yet based on language. Factors that cause poor test items are errors in checking concepts, and typing errors.

Overall, it can be concluded that errors caused by test items made less provide information to find accurate answer keys, thus making students wrong in reading when working on test items. In addition, it will cause confusion for students if they read the test items. Test products for assessment of learning outcomes, a writer of test items must provide clear information including graphics, maps, images, and information in the form of writing [13]. Mardapi explained that in the study of writing multiple choice tests must use communicative language and use standard grammar [14].

In the Winstep application, a match analysis is carried out for each logistic model which is intended to determine whether the respondent of all test participants is against a fit item (fit) with the chosen model. Based on the results of the analysis, the two-parameter logistic model is a model that matches the item can be seen in Table 1. The estimated reliability obtained in the trial phase is 0.92. Based on these estimation parameters, this test instrument is reliable and has a high enough stability and can be used to measure high-level thinking skills of high school students.

TABLE II. SUMMARY OF ESTIMATES IN THE VALIDATION AND MEASUREMENT PHASE

No	Description	Estimated Validation
1	Difficulty Level	1,04 – (-1,05)
2	The average score and standard deviation	323, and 1 ± 36,5
3	Reliability	0,92
4	Average value and standard deviation of MNSQ Infit	1,04, and ± 0,18
5	The average value and standard deviation of the MNSQ Outfit	1,01, and± 0,22

Empirical validity proves that all test items are on average fit. This is caused by several factors, among others: (1) test items that follow procedures, (2) test items developed in accordance with indicators derived from aspects of high-level thinking abilities and physics material, (3) test items consisting of the 50 test items have been tested through content validity, and (4) data retrieval and measurement are carried out in earnest so that students really do the work [15]. "The social validity criterion using empirically validated procedures". This means that social criteria validation uses empirical validation procedures [16]. Respondents' seriousness in working on test items when retrieving data will affect test results.

The normality test in this study was used to test the normality of the test instrument validation data. Normality is one of the methods used to show that the distribution of data taken in research samples represents the population [17]. If the sample taken is not normally distributed then the conclusion will be wrong. The significant level used for nominations is 0.05, if the significance obtained is more than or equal to 0.05 then the data is normally distributed [18].

IV. CONCLUSION

Based on the results of the analysis, it was found that the instruments matched the test items tested. The estimated validity and reliability obtained in the trial phase is 0.92. Based on these estimation parameters, this test instrument is valid and reliable. This instrument has high stability and can be used to measure high-level thinking skills of high school students.

REFERENCES

[1] H. Gardner, Multiple Intelligences (terjemahanYelvi Andri Zaimul). New York: Basic Books, 2013.

[2] E.L. Arwood, Language function an introduction to pragmatic assessment and intervention for higher Order thinking and better literacy. Philadelphia: Jessica Kingsley Publishers, 2011.

[3] S.M. Brookhart, Assess higher-order thingking skills in your classroom. Alexandria: ASCD, 2010.

[4] L.W. Anderson and D.R. Krathwohl, Kerangka landasan untuk pembelajaran, pengajaran, dan asesmen (Translated by Agung Prihantoro). New York: Addition Wesley Longman, 2010.

[5] A.L. Costa, Developing minds: programs for teaching thinking (Rev.Ed). Volume 2. Alexandria: ASCD, 1991.

[6] Y.K.H. Bogan and R.C. Porter, "On the Ball with Higher-Order Thinking," ProQuest Research Library, vol. 36, pp. 46-47, 2005.

[7] T.R. Koballa Jr. and E.L. Chiappetta, Science intruction in the middle and secondary schools: developing fundamental knowledge and skill(Rev.Ed). Boston, USA: Pearson, 2010.

[8] E. Etkina, Pedagogical Content Knowledge and Preparation of High School Physics Teachers. Physics Education Research, vol. 6, pp. 21-22, 2010.

[9] E. Istiyono, Pengukuran Kemampuan Berpikir Tingkat Tinggi Fisika Peserta Didik SMA di DIY [Desertasi]. UNY: Program Pascasarjana, 2013.

[10] R.K. Hambleton, H. Swaminatha, and T.H. Roger, Fundamental of item Response Theory. Newburg park, LA. Sage publication ICC, 1991.

[11] I. Rianawati, Berfikir Tingkat Tinggi (Higher Order Thinking) [Online]. Retrieved from: http://idarianawaty.blogspot.com/2011/08/berpikir-tingkat-tinggi-higher-order.html , accessed on 14 Februari 2014.

[12] B. Sumintono and W. Widhiarso, Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial. Cimahi: Trim Komunikata Publishing House, 2013.

[13] G. Van der Berg, "The use of Assessment in the Development of Higher-order Thinking Skills," Africa education Reviewer, vol. 1, no. 2, pp. 279-294, 2008.

[14] D. Mardapi, Teknik penyusunan instrument tes dan nontes. Yogyakarta: Mitra Cendikia, 2008.

[15] D. Colton and R.W. Covert, Designing and Constructing Instruments for Social Research and Evaluation. San Francisco: John Wiley & Sons Inc., 2008.

[16] R.L. Ebel and D.A. Frisbie, Essentials of educational measurement (5rd ed.).. New Dehli. Prentice-Hall. Inc., 1979.

[17] S. Arikunto, Prosedur Penelitian. Suatu Pendekatan Praktik (edisi Revisi VI). Jakarta: PT Rineka Cipta, 2006.

[18] W.J. Popham, Test better, Teach Better: The Instructional Role of Assessment. Virginia: ASCD, 2003.