



Measurement Instrument of Scientific Reasoning Test for Biology Education Students

Bagus Endri Yanto

Student of Post-Graduate Educational Sciences, the State University of Yogyakarta, Indonesia, bagus.endri2015@student.uny.ac.id

Bambang Subali

Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Indonesia, b_subali@yahoo.com

Slamet Suyanto

Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Indonesia, slametsuyanto@yahoo.com

The purpose of this research is to develop a measurement instrument of scientific reasoning tests for biology education students. The research method used the development stage of the instrument based on the test development procedure according to Oriondo & Dallo-Antonio including: (1) design of test, (2) trial test, (3) determination of validity, and (4) determination of reliability. Test instrument sheet consisting of 13 items and then validated by promotor and measurement experts. The validated instruments were then tried out on 100 students in Biology Education Program of UIN Sunan Kalijaga Yogyakarta. The polytomous data were analyzed using Partial Credit Model (PCM-1 PL). The results showed that 13 items of scientific reasoning instruments were proved fit with the PCM model. By using the lowest limit criteria, INFIT MNSQ was 0.77 and the upper limit was 1.30, the instrument reliability was 0.74 that based on the estimate of case estimate was high. The index of difficulty of items ranged between -2.25 and 0.97. Thus, scientific reasoning instruments which were qualified to measure scientific reasoning skills should be prioritized in learning for biology education students.

Keywords: instrument development, scientific reasoning skills, analyzing, evaluating, creating, polytomus scale, PCM-1PL

INTRODUCTION

Science education today has great challenges for science teachers, not only about how to teach, but also how to do it in a more relevant way, innovation so that students can

Citation: Yanto, B. E., Subali, B., & Suyanto, S. (2019). Measurement Instrument of Scientific Reasoning Test for Biology Education Students. *International Journal of Instruction*, 12(1), 1383-1398. <https://doi.org/10.29333/iji.2019.12188a>

significantly develop their scientific skills and attitudes (Martin-Gamez, Prieto-Ruz & Jimenez-Lopez, 2016). Science teaching should improve the reasoning process by using experimental stages in inquiry methods involving cognitive and metacognitive activity. the effectiveness of science teaching process of the student has an effect on science understanding (Artayasa at al., 2018). The learning experience forms a mental structure that provides the storage and transformation of existing knowledge (Makarova at al., 2017). In addition, science education plays a role to develop a pattern of scientific reasoning that supports intellectual thought (Kind & Osborne, 2017). Biology learning, especially in college as the development of integrated skills of scientific work can be further enhanced through scientific reasoning ability. The results of the teaching should be able to develop process skills by using logical, creative and analytical logical reasoning to develop investigative and problem-solving skills (Thummathong & Thathong, 2016). Therefore, students are required to be able to use scientific reasoning in making decisions and solving problems (Ding, 2014).

Reasoning is a logical process of thinking to draw conclusions from a previous reality (Partanto & Al Barry, 1994) by linking evidence, facts to construct and evaluate arguments resulting in the achievement of logical conclusions (Shaw, 2010). Reasoning is one aspect of intelligence possessed by humans (human intelligence). According to Bruner (Lohman & Lakin, 2010) reasoning is a process of drawing conclusions or inferences based on available information. Based on some of these definitions it can be concluded that reasoning is a process or an activity of thinking in connecting the facts, by building and evaluating arguments so as to produce relevant new conclusions. Then, the conclusion is the thought processes based on reason or facts previously. In learning the need to strive for and empower the potential of student reasoning. Based on this description reasoning is an important thought process in learning, such as solving problems and making critical and creative decisions that are expected to be achieved through science education. It is supported by research results Steinberg and Cormier (2013) showed that prospective science teachers with the training provided had knowledge of scientific reasoning skills in certain contexts through training and reflection in accordance with the curriculum. Furthermore, teachers gained perspective on how this should have an impact on their teaching.

According to Opitz at al., (2017) education systems increasingly emphasize the importance of scientific reasoning skills such as generating hypotheses and evaluating evidence. The results of the study conclude that there are several variables of scientific reasoning tests namely the skill of generating hypotheses, producing evidence, evaluating evidence, and drawing conclusions. Novkovic-Cvetkovic and Stanojevic (2017) revealed that the success of the quality of the teaching process can be identified in the ability of students to transfer practically to apply the same knowledge in new situations. The importance of teachers' innovation is in order to accelerate the progress of learning because it contributes to the efficiency of the education process. Without innovative development in the assessment of learning then no pedagogical quality improvement is achieved, as it becomes an integral part of the teaching process.

There are several assumptions why scientific reasoning is important to develop such as to be able to link and use the concept of science in everyday life to solve problems with logical decisions, without using intuitive reasoning, and be able to conduct sustainable scientific research. To develop scientific reasoning in the learning process, we must integrate the subject materials with scientific methods through a scientific approach. A scientific approach plays a role as a method of acquiring knowledge by which researchers use inductive reasoning from their observations to create hypotheses and then deductively make a logical conclusion (Ary at al., 2010). Scientific reasoning skills is one of the important goals that must be achieved in learning that is useful for students by involving the ability to think in the process of investigation, experiment, analysis, evaluation of evidence, drawing conclusions. From the description above, it can be explained about the importance of developing scientific reasoning through the scheme below.

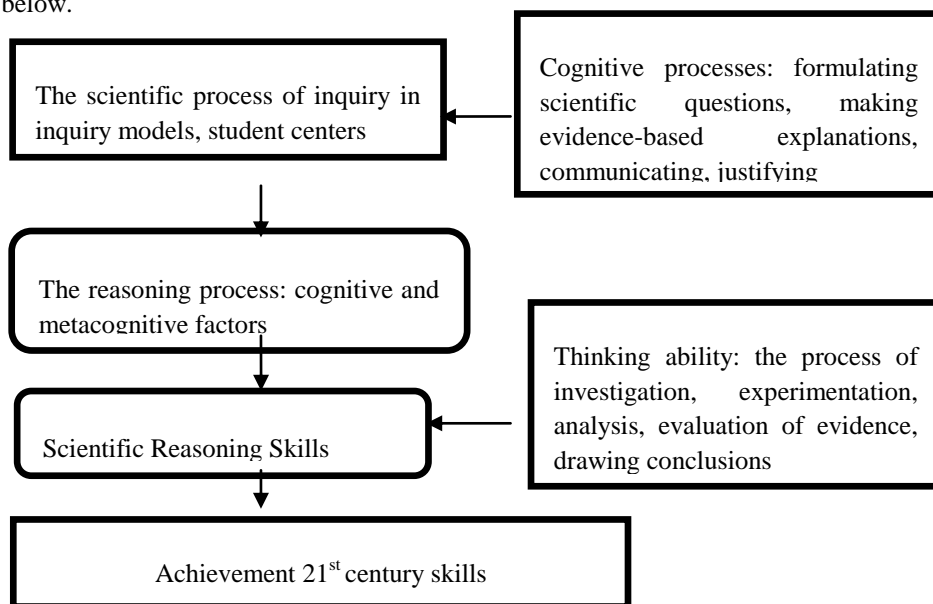


Figure 1
Development of Scientific Reasoning Through Inquiry Learning

Many high-level cognitive processes and thinking skills can be assessed well using assessment procedures. Lecturers should consider a number of principles when making test items to measure improvements in scientific reasoning ability in the learning process. According to Subali (2016), the gain of learning achievement was obtained after applying a program of learning, through evaluation of assessment as indicator to know the improvement of the learners' learning achievement. The assessment procedure should always be tailored to the objectives to be achieved and it is important for the lecturer to collect reliable and valid information of evaluation. When assessing students' understanding by showing what they know and their way of thinking using a

variety of performance measurements and reflective techniques (Arends, 2012) so it can be concluded that in the assessment of learning outcomes it is important to conduct evaluation of learning by providing a test. The results of the assessment are collected through measurement and observation by lecturers, so that the data were obtained in the form of progress and various learning obstacles that need to be improved further in terms of methods, syntax of learning and assessment instruments.

Based on observations on biology learning at Ahmad Dahlan University, Sanata Dharma University and the State Islamic University of Sunan Kalijaga Yogyakarta, Indonesia. observation using the instrument sheet, observations of lecturer learning as many as four lecturers of biology education from each university as respondents based on the aspects of the learning methods used, the form of the test, and the thinking skills were developed. The results of observations can be identified on several things, namely lack of trained scientific reasoning skills that lead learners to solve a problem, developed competencies still focus on skills achievement low level cognitive form of memory, understanding, or application, whereas scientific reasoning ability is important for learners to develop. Taking into account some of these observations, it becomes a challenge for current lecturers to optimize reasoning ability through innovation of scientific reasoning instrument development in biology subject so that the expected learning achievement is achieved. To achieve some of the objectives of the learning objectives, it is important for lecturers to use a variety of assessment techniques used in the form of test. For example, in relation to developing thinking and reasoning skills, it is necessary to understand what aspects and indicators which are related to that ability.

The basic principle for conducting an assessment of scientific reasoning ability should know the characteristics of the instrument which is capable of measuring the competencies achieved. These efforts are supported by McNeill at al., (2012) who stated that to assess the ability of HOTS learners, an instrument that involves critical thinking skills, problem solving and creativity is needed so that instruments based on competence related in learning are required. Teachers should plan well and involve learners in learning activities that can encourage and develop high-level thinking skills (Istiyono at al., 2014). Nitko and Brookhart (2011) explained that the basic provision of high-level thinking is to use problems that require the use of knowledge and skills in new situations by providing a description statement. Based on some experts' opinions it supports the development of scientific reasoning instruments that have the same characteristics as high-order thinking skills.

Many experts have examined the importance of scientific reasoning based on some of their current research findings that support to be applied as a variable to be achieved in learning. For example, in the research of Moore and Rubbo (2012), it explains Scientific reasoning has an important role in the problem-solving process. The attempt to trace scientific reasoning will have more high problem-solving skills; it can have an impact on the achievement of more effective students' learning outcomes (Nieminen at al., 2012; Stephans & Clement, 2010). Competence of scientific reasoning is a part of high-level thinking skills and a feature of intellectual maturity so it is important to do training in higher education, especially for science education students (Marusic at al., 2012). Each

learner is increasingly encouraged to participate in making decisions about a variety of science issues, so it needs to develop reasoning abilities and logical abilities based on scientific knowledge (Jeong et al., 2014). In addition, the significant increase of scientific reasoning ability has positive impacts on the learning practices (Steinberg & Cormier, 2013).

Thus, based on some of these experts' opinions supporting the development of scientific reasoning, then it is important for lecturers to design assessment instruments by determining the appropriate indicators to apply. Based on these matters, it is necessary to develop an instrument of measuring students' scientific reasoning abilities in biology learning. In this study, scientific reasoning ability is a part of higher order thinking skill which is in the top level of thinking ability in Bloom's revised cognitive taxonomy that includes analytical, evaluation and creative skills in a cognitive process. The instrument grid is structured on the basis of aspects, descriptions and scientific reasoning indicators which are then used to compile the items. Table 1 presents a lattice of scientific reasoning abilities as follows.

Table 1
Lattice of Scientific Reasoning Abilities

Aspek	Description	Indicator
Analyzes	Thinking to simplify information into parts of its components	a. Distinguishing relevant sections of factual objects, b. Analyzing the relationship between variables c. Describing the causal relationship of a phenomena
Evaluation	Making a consideration of a condition, value, an argument given based on facts, principles or guidelines.	a. Reviewing critically factual statements b. Testing the validity of the procedure based on data collection.
Creativity	Rational thinking by developing new ideas through imagination	a. Formulating a conclusion b. Formulating hypothesis c. Designing scientific procedures

It is expected that the instruments are useful for lecturers and researchers at any institution. The general aim of this study is to develop an instrument for measuring students' scientific reasoning skills in learning biology, and to obtain the characteristic of scientific reasoning skills assessment which includes aspects of analyzing, evaluating, and creating. Based on the theoretical review and relevant research, then research questions are formulated such as (1) How are the results of the construct validity of scientific reasoning instruments that have been developed? (2) Does the scientific reasoning assessment instrument consist of fit items based on the polytomous IRT model?

METHOD

Development Model

This research used a development research. This research is designed to obtain the product that is the measurement instrument of scientific reasoning ability. The steps of

this development study refer to the development stage of the instrument referring to the test development procedure according to Oriondo and Dallo-Antonio (1984) stages in the development of tests include: (1) designing tests, (2) test trials, (3) determining validity, and (4) determining reliability. After designing instruments scientific reasoning tests, then-reviewed by five experts assessment of education namely Prof. Dr. Zuhdan K. Prasetyo, Prof. Dr. Djukri, Dr. Paidi, Dr. Agung W. Subianto and Dr. Widodo. All test items that have been prepared are reviewed to meet the relevance requirements. The items are arranged in the form of descriptions for all indicators. Determination of the reliability criteria according to Arikunto (2009). Classification of the reliability test is as follows:

Table 2
Test Reliability Criteria

Range	Criteria
$0.80 < r \leq 1.00$	Very high
$0.60 < r \leq 0.80$	High
$0.40 < r \leq 0.60$	Enough
$0.20 < r \leq 0.40$	Low
$0.00 < r \leq 0.20$	Very low

The reliability of the test was 0.74, including high category. Reliability of test instruments with high category that confirms the results of measurement with this instrument is reliable and has fulfilled the requirements. Thus, this instrument can be used to explore information about students' scientific reasoning abilities appropriately. Students who have high abilities can solve science problems based on understanding scientific concepts (Ramos at al., 2013). Test instruments that have been designed are tested and analyzed to get a valid and reliable test. The conclusion from the results of the expert validation is a test of scientific reasoning ability is suitable for use.

Measurement of item difficulty index was based on value delta. According to Hambleton at al., (1991) the parameter of difficulty level in each items will be good if $-2 \leq b \leq +2$. The level of difficulty for good items varies between -2.00 to 2.00. Items with a difficulty level of -2.00 indicate that the item is very easy; while the difficulty level of 2.00 means that the item is very difficult. Thus, in terms of the difficulty level of the item, this instrument includes good categories. Index item difficulty criteria refer to Hambleton at al., (1991) is as follows:

Table 3
Index Item Difficulty Criteria

Range	Criteria
$b > 2$	Very difficult
$1 < b \leq 2$	Difficult
$-1 \leq b \leq 1$	Medium
$-1 > b \geq -2$	Easy
$b < -2$	Very easy

Participants and Research Sample

This study involved five experts for studying the instrument of scientific reasoning test. The experts are consisting of two measurement experts and three biology education experts from university state of Yogyakarta Indonesia. This study also involves 100 students concentration of biology education. The population in this study were all students of biological education study program UIN Sunan Kalijaga. The sample of this research consist of 50 students from the second semester and 50 students from the fourth semester in the academic year of 2017/2018 in biology education program of UIN Sunan Kalijaga Yogyakarta Indonesia. Sampling in this study uses cluster sampling technique to determine samples that are not based on individuals but on class groups.

Data Collection and Analysis

Data collected from test scores. Number of scientific reasoning skills test instruments are 13 items essay. The instruments test were an essay-type test with answer criteria consisting of three categories. The score of the essay answer was between 0, 1 and 2. The criteria of answers of scientific thinking is by giving score with the description as follows: score (2) if the answer is complete and correct with clarity of the argument, score (1) if the answer is only partially complete and less relevant, score (0) if the answer does not mention completely and true with clarity of argument (Brookhart, 2010). The instrument grille is structured on the basis of aspects, descriptions and scientific reasoning indicators which are then used to compile the items. Table 4 presents a grille of scientific reasoning abilities as follows.

Table 4
The Grille for Scientific Reasoning Skills Test Istrument

Aspekct	Indicator	Question Number
Analyzes	a. Distinguishing relevant sections of factual objects,	1, 2
	b. Analyzing the relationship between variables	3
	c. Describing the causal relationship of a phenomenon	4
Evaluasion	c. Reviewing critically factual statements	5,6, 7
	d. esting the validity of the procedure based on data collection.	8, 9
Creativity	d. Formulating a conclusion	12, 13
	e. Formulating hypothesis	10
	f. Designing scientific procedures	11

Expert assessment results were analyzed by using content validity, measurement instruments of scientific reasoning skills to the accuracy of each test item with the indicators formulated, to evaluate and assess the quality of the instruments that have been made so that the instruments that have been made are valid. While reliability testing used inter-rater reliability level measurement on test instruments using the Cohen's Kappa coefficient involving several experts as assessors. The construct validity of scientific reasoning tests is then tested using confirmatory factor analysis (CFA) method to test the dimensionality as a reference to test the unidimensionality assumption

of the measuring instrument. Analysis of this research data used Partial Credit Model 1 PL (PCM1-PL) using quest program. If the average value of INFIT MNSQ is close to 1.0 and the standard deviation is close to 0.0 then the entire items fit test item with the model. An item is declared Fit with the model if the MNSQ INFIT value is in the range 0.77 to 1.30 (Adams & Seik-Tom, 1996). In this test, the results of fit items and difficulty index are obtained.

FINDINGS

Content Validation of Scientific Reasoning Testing Instruments

The test of the validity of the test instrument of scientific reasoning ability was done through expert judgment of the grid aspects of scientific reasoning and items. All test items that have been compiled were reviewed to meet relevancy requirements. Items were arranged in the form of a description for all indicators. The items that have been compiled were then assessed and validated by five experts. An expert judgment was required to evaluate the relevant constructs of the research instruments developed. Data from content validity analysis by calculating the Aiken index of each item from the expert agreement, regarding the validity is presented in the following table 5.

Table 5
Calculation Results of Aiken Index Test Instruments

Item	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Value	Information
1	1	1	1	1	1	1.0	Valid
2	1	1	1	1	1	1.0	Valid
3	1	1	1	0	1	0.8	Valid
4	1	1	1	1	1	1.0	Valid
5	0	1	1	1	0	0.6	Valid
6	1	1	1	1	1	1.0	Valid
7	1	1	1	1	1	1.0	Valid
8	1	1	1	1	1	1.0	Valid
9	1	1	1	1	1	1.0	Valid
10	1	1	1	0	1	0.8	Valid
11	1	1	0	1	1	0.8	Valid
12	0	1	1	1	0	0.6	Valid
13	1	1	1	1	1	1.0	Valid

Information:

Score 1: if the statement is relevant to the criteria of various aspects

Score 0: if the statement is irrelevant to the criteria of various aspects

Based on the results presented in the table, the results of all items show a valid category, because the lowest index is 0.6 and the highest is 1. Then the results are interpreted, if the agreement index is less than 0.4 then the validity is low and if it is more than 0.8 is said to be very high (Guilford, 1956). In conclusion, from the results of the validation of experts in general, tests of scientific reasoning ability were feasible to use after revising several things.

Some of the main points of input from the four assessors were: 1) scoring on the assessment of this aspect needs to be emphasized. For example the answer "relevant" is given a score of 1 and the answer "irrelevant" is given a score of 0, 2) synthesis is made from variable from the results of the theoretical study developed into a grid for the instrument development; 3) the grid needs a less clear variation of sentences, 4) grammatical restructuring to make the instrument is needed in order that the sentences are more communicative and the focus of the substance is more clear, 5) the instrument indicators need to be reviewed to be relevant. Data of feasibility test results by validator on each item of measurement instrument indicated that all items of instrument were eligible to be used as measurement instrument. Based on the level of agreement (reliability) between the five assessors can be explained by calculating the inter-rater reliability coefficient using the Cohen's Kappa (κ) coefficient. The results of the calculation are presented in the table 6 below.

Table 6
The Coefficient (κ) Inter-rater Assessment Test Instruments

	Rater				
	1	2	3	4	5
Rater	1				
	2	1.00			
	3	1.00	1.00		
	4	0.60	0.60	0.60	
	5	1.00	1.00	1.00	0.60

Description::

$\kappa < 0.40$: less

$0.40 < \kappa < 0.75$: good

$\kappa > 0.75$: very good

Based on the table above, overall inter-rater reliability can be determined by determining the average reliability of the five assessor pairs, which is equal to 0.74 in the good category. Reliability coefficient value of the scientific reasoning ability test instrument obtained in accordance with the minimum criteria used, is 0.70 (Linn, 1989), so that the instrument meets reliable requirements.

Construct Validation of the Scientific Reasoning Test Instrument

The CFA first order test was performed by making a measurement model to prove how well the indicators in the measuring instrument could be used as an instrument of measuring latent variables in terms of scientific reasoning, analytical, evaluation and creative aspects. In this test, a model of three latent variables of scientific reasoning was made by 8 points as directly measured indicators. This test would prove that all grain measurements measured three variables. At this stage, unidimensional assumption test using confirmatory factor analysis (CFA) must be fulfilled in the use of item response theory. The test results on CFA first order measurement model with 8 indicators yielded p -value = 0.09180 ($p > 0.05$) and RMSEA = 0.073 (RMSEA < 0.8). Based on these data,

it can be concluded that this model fit with the data. It means that the relationship between variables hypothesized was supported by empirical data, as table 7 as follows.

Table 7

Compatibility Index of 8 Items of Confirmatory Factor Analysis (CFA) Model		
Aspect	Compatibility index	Description
Chi-Square	25.13	Model fit
p-value	0.09180	Model fit
RMSEA	0.073	Model fit

Based on the test results, these 8 indicators were valid indicators for the measurement of the constructs of scientific reasoning because the model fit with the data based on the criteria of the CFA measurement model. Thus, this measuring instrument could be said to have fulfilled the unidimension assumption so that the application of the model with IRT approach could be done.

Items Description of Scientific Reasoning Test Instrument through Item Response Theory (IRT) Approach

Based on the measurement results using IRT analysis, describing the characteristics of measurement instruments was done by analyzing the results of the quest program. In this test, the results of item fit, and parameter estimation of each item were obtained. The data was considered to fit the model when the mean square value was almost 1.0 and the value of the test t is close to 0. Adams & Seik-Tom (1996) suggest the use of fit mean square because it is more useful to see the relevance between the model and the data. The criterion of item fit testing in testing the scientific reasoning test instrument is the infit mean square value. An item is said to be fit if it has a mean square infit value in the range 0.77 to 1.30 (Adams & Seik-Tom, 1996).

The analysis results showed that all items were in the acceptable infit mean square range. A total of 13 test items fit criteria with the model. According to Wright & Master (1999) stated the reliability of the test calculated based on the error measurement calculated based on estimation by test (case estimate). The reliability of the test reached 0.74 so that reliability was high. If the calculation was based on internal consistency using the classical approach as alpha Cronbach value, it was medium, as 0.63. A test instrument categorized as medium if the alpha value is 0.6-0.7 (Taber, 2017). The complete results are presented in table 8 to the test of the fit items as follows.

Table 8
Item Fit Testing

Item	Infit mean square	Description
Item 1	1.21	Item fit
Item 2	0.91	Item fit
Item 3	0.95	Item fit
Item 4	1.14	Item fit
Item 5	0.79	item fit
Item 6	1.14	Item fit
Item 7	1.06	Item fit
Item 8	0.91	Item fit
Item 9	0.93	item fit
Item 10	1.25	Item fit
Item 11	0.83	Item fit
Item 12	1.21	Item fit
Item 13	0.78	Item fit
Mean	1.01	
SD	0.17	

Based on the above table, all items of scientific reasoning tests were in the acceptable infit mean square range so that all of the fit items are used in measuring scientific reasoning abilities. The result of the analysis description of item difficulty index using the quest showed that the device of scientific reasoning test had a range of difficulty index value between -2.25 up to 0.97. Item 2 up to 13 on the instrument test were in good criteria because of the level of difficulty in the range between -2.00 up to 2.00. While item 1 with criteria is not good. The mean value of the difficulty index of the scientific reasoning tests was $0,00 \pm 0,89$. The complete results are presented in table 9 of the item difficulty index.

Table 9
Item difficulty index

Item	Ability means	Description
Item 1	-2.25	Very easy
Item 2	0.14	Medium
Item 3	0.97	Medium
Item 4	0.40	Medium
Item 5	-0.72	Medium
Item 6	-0,34	Medium
Item 7	0.67	Medium
Item 8	0.53	Medium
Item 9	-0.10	Medium
Item 10	0.82	Medium
Item 11	-0.22	Medium
Item 12	0.82	Medium
Item 13	-0.72	Medium
Mean	0.00	
SD	0.89	

The ability of test participants was based on the number of items that could be answered correctly, the estimated ability of all test takers was -0.36, below the average of the item difficulty index of 0.00.

DISCUSSION

After starting various stages in the development of the instrument starting from the expert judgment and then revising the draft of the instrument of assessment of scientific reasoning as a whole, the final result of the validation of experts shows that the assessment instrument has met the valid category and was ready to be used in data retrieval. Assessment tools were based on strong and relevant supporting theories. Draft instruments of scientific reasoning ability after being validated and worthy of use, it was then used to test the construct validity of scientific reasoning test using confirmatory factor analysis (CFA) method. The use of CFA was to test dimensionality as a reference to test the unidimensionality assumption of measurement instrument. At this stage, the analysis was performed using Lisrel program version 8.8. Analysis of construct validity with CFA was done by first order confirmatory analysis. In this study, scientific reasoning consisted of three aspects, namely the ability of analysis, evaluation and creativity for it was necessary to test suitability model of scientific reasoning which fit with the data using CFA

At this stage, unidimensional assumption test was done using confirmatory factor analysis (CFA) which must be met in the use of item response theory. Test results on CFA first order measurement model with 8 items yielded p-value = 0.09180 ($p > 0.05$) and RMSEA = 0.073 (RMSEA < 0.8). Based on these data it could be concluded that this model was fit with the data means that the relationship between hypothesized variables has been supported by empirical data and was a valid indicator for the measurement of scientific reasoning skills. In addition, the results of the first order CFA could be used to identify construct reliability (CR) based on the value of the loading items. A CR value of 0.74 indicated the acceptable range. Thus, that it could be concluded that the instrument of scientific reasoning test was a valid and reliable instrument.

Therefore, this measurement instrument can be said has met the unidimensional assumption so that the application of the model with item response theory (IRT) approach can be done. In the analysis of empirical validity results are validity test and act as the concept being measured (Hair et al., 2010). Based on the analysis of person fit tests, of all participants, the mean infit mean square value in the person estimation showed the model of all persons fit with the data of 0.99 is in the acceptable range from 0.77 to 1.30. While in the test item fit based on infit mean square value showed all items were in the range of infit mean square in the range 0.77 to 1.30 which as acceptable. A total of 13 test items met the fit criteria with the model. Thus, all items could serve as a scientific reasoning measurement item. The item difficulty level indicated that the easiest item was item 1 of the test device of -2.25 while the other item indicated the degree of difficulty of the moderate category. Based on the overall item analysis the

items in the test form were considered fit with the model as they met the requirements of the fit statistics required in the quest program. Thus, the overall item analyzed was fit according to PCM 1-PL.

CONCLUSION AND IMPLICATION

Based on the results of data analysis, the conclusion is the scientific reasoning test instrument was ready to be used as a measurement instrument of scientific reasoning ability aspects of analysis, evaluation and creativity should be prioritized in biology learning. Data of validity test results by validator on each item of measurement instrument indicated that all items of instrument were eligible to be used as measurement instrument. The measurement results showed all items of scientific reasoning assessment instrument in biology learning were fit with PCM 1-PL model.

The result of this research can provide implication to the curriculum development on the biology lecturing and used as an assessment model that can facilitate lecturers as a basis for improving learning outcomes expected. Therefore, lecturers must have the skills to develop assessment instruments and students are trained to have scientific reasoning skills in the learning process. This ongoing assessment process has implications for the demands of curriculum achievement. Suggestions of this study based on the findings of the characteristics of the instruments developed, it is recommended that teachers be able to train students through inquiry learning that is suitable for developing scientific reasoning skills based on analytic, evaluative and creative aspects. Lecturers should apply the test of scientific reasoning skills in science learning, which is still not optimal. Thus, there will be development scientific reasoning skills.

REFERENCES

- Adams, R.J. Kho, & Seik-Tom. (1996). *Acer quest version 2.1*. Camberwell, Victoria: The Australian Council for Educational Research.
- Arends, R I. (2012). *Learning to teach (9rd ed)*. New York : McGraw-Hill.
- Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Artayasa, I P. Susilo, H., Lestari, U., & Indriwati, S. E. (2018). The Effect of Three Levels of Inquiry on the Improvement of Science Concept Understanding of Elementary School Teacher Candidates. *International Journal of Instruction*, 11(2), 235-248. <https://doi.org/10.12973/iji.2018.11216a>.
- Ary, D., Jacobs, L. C., Sorensen, C. K., Walker, D. A., & Razavieh, A.(2010). *Introduction to research in education. Measurement* (Vol.4, p. 669). <https://doi.org/10.1017/CBO9781107415324.004>.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom. Ascd*. Retrieved from <http://www.ascd.org/Publications/Books/Overview/How-to-Assess-Higher-Order-Thinking-Skills-in-YourClassroom.aspx>.

- Ding, L. (2014). Verification of Causal Influences of Reasoning Skills and Epistemology on Physics Conceptual Learning, *Physical Review Special Topics - Physics Education Research*, 10(2). <https://doi.org/10.1103/PhysRevSTPER.10.023101>.
- Guilford, J. P. (1956). *Fundamental Statistics in Psychology and Education*. New York: Mc Graw-Hill Book Co. Inc
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. Basim, Pearson Education Inc.
- Hambleton, R.K.; Swaminathan, H. & Rongers, H.J. (1991). *Fundamental of item respons theory*. Newbury Park, CA: Sage Publication Inc
- Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (pysthots) peserta didik sma. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 1–12. <https://doi.org/10.21831/PEP.V18I1.2120>.
- Jeong, J., Kim, H., Chae, D. hyun, & Kim, E. (2014). The Effect of a Case-based Reasoning Instructional Model on Korean High School Students' Awareness in Climate Change Unit, *Eurasia Journal of Mathematics, Science and Technology Education*, 10(5), 427–435. <https://doi.org/10.12973/eurasia.2014.1105a>.
- Kind, P., & Osborne, J. (2017). Styles of scientific reasoning: a cultural rationale for science education. *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>.
- Linn, R.L. (1989). *Educational measurement*. (3rd ed.). New York: Macmillan Publishing Company
- Lohman, D.F., & Lakin, J.M. (2010). *Reasoning and Intelligence*. Retrieved from http://www.faculty.education.uioa.edu/docs/dloman/reasoning-and-intell_Lohman-LAkin-102709.
- Makarova, A., Lvovna, M., & Mikhailovna, V. (2017). Education Process Visualization in Metacognition Development and Sustainability, *International Journal of Cognitive Research in Science, Engineering and Education*, 5(2), 65–74. <https://doi.org/10.5937/ijcrsee1702065A>
- Martin-Gamez, C., Prieto-Ruz, T., & Jimenez-Lopez, M. A. (2016). Developing Preservice Science Teachers' Beliefs about New Approaches to Science Education, *Journal of Turkish Science Education*, 13(4). <https://doi.org/10.12973/tused.10181a>
- Marušić, M., Mišurac Zorica, I., & Pivac, S. (2012). Influence of learning physics by reading and learning physics by doing on the shift in level of scientific reasoning. *Journal of Turkish Science Education*. Volume 9, Issue 1, 146-161.
- McNeill, M., Gosper, M., & Xu, J. (2012). Assessment Choices to Target Higher Order Learning Outcomes: The Power of Academic Empowerment, *Research in Learning Technology*, 20(3), 283–296. <https://doi.org/10.3402/rlt.v20i0.17595>

- Moore, J. C., & Rubbo, L. J. (2012). Scientific reasoning abilities of nonscience majors in physics-based courses. *Physical Review Special Topics - Physics Education Research*, 8(1). <https://doi.org/10.1103/PhysRevSTPER.8.010106>.
- Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between Representational Consistency, Conceptual Understanding of the Force Concept, and Scientific Reasoning, *Physical Review Special Topics - Physics Education Research*, 8(1). <https://doi.org/10.1103/PhysRevSTPER.8.010123>
- Nitko, A. & Brookhart, S. M. (2011). *Educational assessment of students (6th ed)*. Boston: Pearson Education, Inc
- Novkovic-Cvetkovic, B., & Stanojevic, D. (2017). Educational Needs of Teacher for Introduction and Application of Innovative Models in Educational Work to Improve Teaching, *International Journal of Cognitive Research in Science, Engineering and Education*, 5(1), 49–56. <https://doi.org/10.5937/IJCRSEE1701049N>
- Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning – a review of test instruments. *Educational Research and Evaluation*, 1–24. <https://doi.org/10.1080/13803611.2017.1338586>.
- Oriondo, L.L. & Dallo-Antonio (1998). *Evaluating educational outcomes(test, measurement, and evaluation)*, 5th. Quezon City: REX Printing Company, Inc.
- Partanto, A.P., & Al Barry, D. M. (1994). *Kamus Ilmiah Populer*. Surabaya: Arkola.
- Ramos, J. L. S., Dolipas, B. B., & Villamor B. B. (2013). Higher Order Thinking Skills and Academic Performance in Physics of College Students: A Regression Analysis. *International Journal of Innovative Interdisciplinary Research*, ISSN 1839-90531 , p: 48-60. Retrived from <http://www.auamii.com/jiir/Vol-01/issue-04/5ramos.pdf>
- Shaw, V. F. (2010). *The cognitive processes in informal reasoning*. Thinking & Reasoning. Publisher: Routledge.
- Steinberg, R., & Cormier, S. (2013). Understanding and Affecting Science Teacher Candidates' Scientific Reasoning in Introductory Astrophysics, *Physical Review Special Topics -Physics Education Research*, 9(2), 1-10. <https://doi.org/10.1103>.
- Stephens, A. L., & Clement, J. J. (2010). Documenting the Use of Expert Scientific Reasoning Processes by High School Physics Students, *Physical Review Special Topics - Physics Education Research*, 6(2).1-15. <https://doi.org/10.1103>.
- Subali, B. (2016). *Prinsip asesmen dan evaluasi pembelajaran*, Edisi Kedua. Yogyakarta: UNY Press.
- Taber, K. S. (2017). The Use of Cronbach's alpha When Developing and Reporting Research Instruments in Science Education, *Research in Science Education*, 1–24. <http://doi.10.1007/s11165-016-9602-2>.

Thummathong, R., & Thathong, K. (2016). Construction of a Chemical Literacy Test for Engineering Students, *Journal of Turkish Science Education*, 13(3), 185-198. <https://doi.org/10.12973/tused.10179a>.

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