JISE 11 (1) 2022 : 78–86

Journal of Innovative Science Education

http://journal.unnes.ac.id/sju/index.php/jise

The Development of Evaluation Tools to Measure Students' Science Literacy on the Solar System Theme

Nurul Afdhilla Asy'syakurni[1]✉, Ani Rusilowati[2], Sudarmin Sudarmin[2]

[1]SMP Negeri 4 Salatiga, Indonesia
[2]Pascasarjana, Universitas Negeri Semarang, Indonesia

Article Info

Article History :
Received
October 2021
Accepted
January 2022
Published
April 2022

Keywords:
Evaluation Tool, Item Response Theory (IRT), Scientific Literacy, Solar System

Abstract

Nowadays scientific literacy is a concern in the world of education. Various efforts have been made by the Government to improve students' scientific literacy. Efforts are made by changing the current curriculum. However, this has not been followed by the availability of assessment instruments or evaluation tools used to measure students' scientific literacy. The results of observations carried out show that currently the evaluation tool used is only to measure students' knowledge. This study aims to develop an evaluation tool based on aspects of scientific literacy competence on the theme of the solar system. This study also aims to determine the scientific literacy profile of students through Item Response Theory (IRT) analysis. The research was conducted with a sample of 256 students of class VIII SMP Negeri 2 Salatiga. IRT analysis using the BILOG MG computer program with a 2-parameter logistic model (2-PL). The results of the item analysis showed that of the 30 items, there was one item (item 25) that had a minus biserial correlation value, so item 25 that no calculations for the next phase. From the analysis of the difficulty level index, 86% of the items have a good difficulty index and all items have a good differentiation of problem index. The results of the students' abilities showed that the average obtained was -0.0143. The mean ability with a minus value indicates that most of the students have low abilities.

✉ correspondence :
Jl. Patimura, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tengah, Indonesia 50712
E-mail: afdhillaasysya@gmail.com

p-ISSN 2252-6412
e-ISSN 2502-4523

INTRODUCTION

Science discipline basically consists of four elements, attitudes, processes, products and applications. These four elements can be obtained by students in learning science. The science learning process in the 2013 Curriculum is carried out with a scientific approach and the assessment is carried out authentically. The success of the learning process can be known by conducting an assessment or evaluation (Heitink et al., 2016). Assessment or evaluation is a process of collecting and processing information to measure the achievement of student learning outcomes. The assessment was conducted to determine the learning outcomes of students both in terms of attitudes, knowledge and skills.

Based on the results of observations carried out at SMP Negeri 2 Salatiga and SMP Negeri 4 Salatiga, the evaluation carried out to determine student learning outcomes has covered three aspects there are cognitive, affective and psychomotor aspects. However, evaluation only emphasizes the cognitive aspects which are usually carried out with test and non-test techniques. The test technique is carried out in the form of daily tests, mid-semester tests, end-of-semester. Non-test techniques are usually carried out in the form of assignments. Affective aspects or attitudes are assessed based on observations made by science teachers. Psychomotor aspects or skills are observed through practical learning activities and learning in the classroom.

Indonesian education is still at a low level, especially in science education. The quality of education, especially science education in Indonesia is still low compared to other developing countries (Rusilowati et al., 2016). The low level of science education can be indicated by the acquisition of Indonesia's scientific literacy ranking in its participation in PISA (Program for International Student Assessment). Scientific literacy is the ability to use scientific knowledge to identify problems and draw conclusions based on evidence in order to understand and make decisions about nature and the changes made to nature through human activities (OECD, 2016). In 2012, Indonesia was ranked 64th out of 65 participating countries. In 2015 the results of the ranking on the subject of science, Indonesia was ranked 62 out of 70 participating countries.

Nowadays, every effort has been made by the Government to improve the scientific literacy skills of students. One of the efforts made is to implement the Curriculum 2013 which demands integrated learning with a scientific approach. However, it has not been followed by the availability of literacy-based teaching materials and evaluation tools used to measure scientific literacy. When students are not accustomed to working on questions that require literacy skills in understanding a discourse/reading that accompanies the questions, students will not succeed in answering scientific literacy questions (Rusilowati et al., 2016; Park, 2013).

Based on the background that has been described, an evaluation tool is needed to measure scientific literacy. The evaluation tool developed in this study was based on aspects of scientific literacy and was in the form of multiple choice questions with four answer choices. The objectives of this study are (1) to describe the characteristics of the evaluation tool developed to measure scientific literacy (2) to describe the characteristics of the evaluation tool items to measure students' scientific literacy on the theme of the solar system, and (3) to describe the profile of students' scientific literacy at solar system theme.

METHODS

The research conducted is research and development. Research and development (R&D) is a research method used to produce certain products, and test the effectiveness of these products (Sugiyono, 2013). The research and development that has been carried out has produced a product in the form of an evaluation tool to measure scientific literacy on the theme of the solar system. The stages of developing an evaluation tool can be seen in Figure 1.

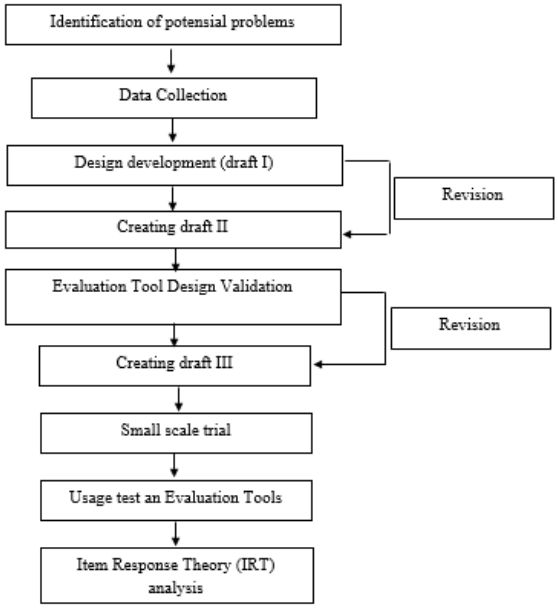

Figure 1. The stages of developing an evaluation tool

This research was conducted at SMP Negeri 2 Salatiga with 256 students sitting in class VIII. Research and development of evaluation tools to measure students' scientific literacy uses research and development procedures according to Sugiyono (2013). The evaluation tool developed is in the form of multiple choice questions with four answer choices. The evaluation tool developed is 50 numbers, the preparation of the evaluation tool developed is based on the indicators contained in the 2015 PISA. The results of the development of the evaluation tool design are in the form of a grid of questions and questions to measure students' scientific literacy on the theme of the solar system. The evaluation tool that has been developed is then validated by the content by the validator. Content validity is a validity that is estimated through testing the test content with rational analysis or through personal judgment (Rusilowati, 2014).

After the content validation was carried out, the evaluation tool was revised and a small-scale trial was conducted. The analysis carried out on a small-scale trial was classical test theory and obtained 30 questions which were then used to measure the scientific literacy of students. The evaluation tool was tested on 256 students, then the Item Response Theory (IRT) analysis was carried out. The analysis was carried out using the BILOG MG computer program with a 2-parameter logistic model (2-PL). The analysis of the two-parameter logistic model (2-PL) focuses on the level of difficulty and differentiation of problem. From the IRT analysis carried out, information on the ability of participants was obtained which was then transformed into a scale to determine the scientific literacy profile.

RESULTS AND DISCUSSION

Characteristics of Evaluation Tool for Measuring Students' Scientific Literacy

The characteristics of the evaluation tools that have been developed include each item developed with the theme of scientific literacy aspects. There are four aspects of scientific literacy, namely aspects of knowledge, aspects of context, aspects of competence and aspects of scientific attitudes. In this study, the aspect used is the aspect of competence. Each item developed is based on indicators of scientific literacy competence and grouped in scientific literacy knowledge. Items are a type of multiple choice questions, consisting of one correct answer key and 3 distracting items. The items developed consist of 10 sub-topics of the solar system. The sub-topic of the solar system consists of planetary orbits and gravitational forces, the solar system, the sun, layers of the sun, the composition of the solar system, the motion of the moon, the motion of the earth and the moon, the rotation and revolution of the earth, and the change of seasons.

The development of this evaluation tool is very necessary to measure the scientific literacy of students. There is a tendency for students to memorize knowledge, but lack the skills to apply knowledge in everyday life (Maatturdiyah & Rusilowati, 2015). The preparation of this evaluation tool is based on a grid that has been designed according to core competence (KI), basic competencies (KD) and learning indicators. There are 10 learning indicators which are then formulated into 50 question indicators. The format of the questions developed is in the form of multiple choice questions with one correct answer choice and 3 distracting answer choices.

The evaluation tool developed was then validated by experts. The validation carried out is content validation which consists of 3 assessment criteria, namely material, construction and language. Based on the results of content validation carried out by the validator, an average of 855 was obtained with a very good category. The assessments

made by the three validators were very good, meaning that the instrument was valid to be used without revision. However, suggestions and input are still given by each validator for instrument improvement. Furthermore, after validating the content of the developed evaluation tool, it was then tested on a small scale. A small-scale trial was conducted by 30 students of SMP Negeri 2 Salatiga. This trial was conducted to determine the reliability, level of difficulty and differentiation of problem from the developed evaluation tool.

Reliability analysis was carried out using the SPSS application, the results of the analysis showed that the evaluation tool for measuring students' scientific literacy on the theme of the solar system was reliable. Reliable in this case means that the instrument has a high level of confidence. If the instrument is given to the same subject but at different times, it gives the same results (Arikunto, 2012). According to Matondang (2019), reliability states the extent to which the results of a measurement can be trusted. This means that the evaluation tool developed is consistent in measuring scientific literacy on the theme of the solar system.

The results of the analysis of the level of difficulty showed that the overall questions included 18% easy questions, 60% moderate questions and 22% difficult questions. Most of the questions developed are included in the medium criteria. The next item analysis is the differentiation of problem of the questions. Questions that have good and sufficient discriminating power mean that the items can distinguish between students who master and do not master the material (Arikunto, 2012). The results of the analysis show that from 50 items there are 16 items that have poor discriminating power. Items that have poor differentiation of problem are not used for field tests. The results of the content validation analysis, reliability, level of difficulty and discriminating power of items on the evaluation tool are used to determine which items can be used for field tests. Items used to measure students' scientific literacy. The result of 50 test items, which can be used are 30 items. The thirty questions used already include material indicators contained in the solar system theme.

Student Scientific Literacy Profile

The evaluation tool that had been developed was then tested on 256 Class 8th students of SMP Negeri 2 Salatiga. After being tested, the Item Response Theory (IRT) analysis was carried out. Modern item analysis or Item Response Theory (IRT) is a theory that uses mathematical functions to link the probability of answering a question correctly with students' abilities. The analysis was carried out using a two-parameter (2-PL) logistic model, where the probability of students correctly answering an item was determined by two item characteristics, namely the difficulty index and the discriminating power index.

The results of the IRT analysis using the BILOG MG application consist of three outputs for each phase. The output in the first phase informs about the number of students who answered correctly (Right), the proportion of correct answers (PCT) divided by the probability of answering incorrectly and the biserial correlation coefficient (Biserial Correlation). The biserial correlation value is a representation of the item's distinguishing power. According to Ebel & Fresbie (1986), biserial correlation describes the relationship between scores on test items and scores on total tests for each testee. A biserial with a high positive value describes the tendency of test takers with a high score to answer correctly. The results of the phase 1 analysis provide information that item 25 has a biserial of less than –0.15 so that the BILOG program is not analyzed or calibrated in the next phase.

After item 25 is not calibrated, the BILOG MG program is run again by analyzing only the remaining 29 items. The second phase informs about 2 parameters used to measure the ability of students, namely the item difficulty index (bi) and the differentiation of problem index (ai). Someone who has high ability will find it easy to work on the items, on the contrary those who have low ability will find it difficult to answer the items. According to Hambleton & Swaminathan (1991) the level of difficulty of the item (bi) has a scale of $-2 \leq b \leq 2$, items that have a level of difficulty close to or below the scale of -2 indicate the easy category. While items that have a level of difficulty (bi) close to or above a scale of 2 indicate that the item is in the difficult category.

The level of difficulty (bi) at the output of the BILOG-MG application can be seen at the threshold value. Based on the analysis that has been done, the difficulty level of all items moves from -2.864 to 4.825. 86% of all items have a good level of

difficulty, so that all test items are able to describe the function of a person's ability. Where students with high abilities will find it easy to work on questions, while students with low abilities will answer questions. Meanwhile, the results of the analysis of the differentiation of problem index (ai) parameters can be seen from the slope value at the output of phase 2. According to Hambleton & Swaminathan (1991), the differentiation of problem index is good if it is on a scale of $0.0 \leq ai \leq 2.0$. Based on the results of the analysis in phase 2, it shows that the differentiation of problem index for all items is in the range of values from 0.219 to 0.803. It is known that all items have good differentiation of problem. This proves that all test items have the ability to emphasize the differences between participants who can answer correctly and answer incorrectly. The analysis of the difficulty level index (bi) and the differentiation of problem index (ai) can be seen in Table 1.

Table 1. Analysis of Difficulty Index (bi) and Differentiation of Problem Index (ai)

Item	Slope	Threshold	Criteria
ITEM0001	0.365	-0.254	Good
ITEM0002	0.287	-0.379	Good
ITEM0003	0.412	-1.006	Good
ITEM0004	0.298	-1.357	Good
ITEM0005	0.399	-1.230	Good
ITEM0006	0.352	-1.270	Good
ITEM0007	0.533	-1.916	Good
ITEM0008	0.713	-0.064	Good
ITEM0009	0.37	-0.75	Good
ITEM0010	0.243	-0.96	Good
ITEM0011	0.307	-1.878	Good
ITEM0012	0.625	-1.572	Good
ITEM0013	0.236	2.130	Not Good
ITEM0014	0.803	-0.077	Good
ITEM0015	0.298	-1.179	Good
ITEM0016	0.629	-1.223	Good
ITEM0017	0.257	2.057	Not Good
ITEM0018	0.305	1.957	Good
ITEM0019	0.332	0.875	Good
ITEM0020	0.276	-0.004	Good
ITEM0021	0.472	-2.864	Not Good
ITEM0022	0.219	4.825	Not Good
ITEM0023	0.241	-0.723	Good
ITEM0024	0.4	-0.159	Good
ITEM0025	Not analyzed		Bad
ITEM0026	0.299	0.824	Good
ITEM0027	0.673	-0.579	Good
ITEM0028	0.442	-1.834	Good
ITEM0029	0.348	-0.73	Good
ITEM0030	0.461	-0.797	Good

Based on 30 items of evaluation tool that were tested on 256 students, 1 item was not used because it had bad bisserial correlation. The main characteristic of IRT is that the test taker's response

to a question being tested will form a characteristic curve of the question (Rusilowati, 2014). According to Hambleton et.al quoted in Rusilowati (2014), this curve is a mathematical function formula that states the relationship between the probability of answering correctly P (θ) and ability (θ). This curve is in the form of an ogive, which is a cumulative frequency curve. The combination of the item characteristic curves of all items with the 2-parameter logistic model can be seen in Figure 2.

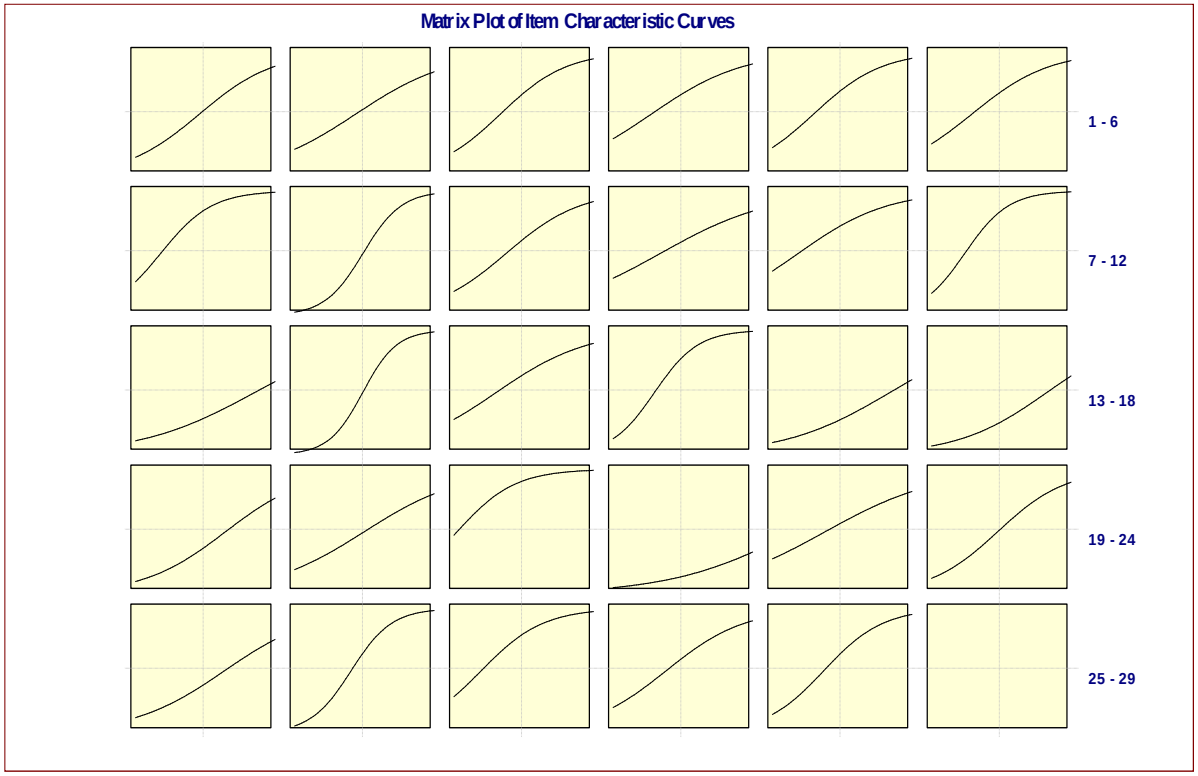

Figure 2. Grain Characteristic Curve of All Items

The characteristic curve describes the characteristics of 29 items using a 2-parameter logistic model. In sequence, the curves are the characteristic curves of items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30. Each item has different characteristics and forms depending on the magnitude of the index of difficulty level and the differentiation of problem index. The steepness of the curve of each item also varies depending on the magnitude of the slope or index of differential power. The sloping curve shows that the grain has low discriminating power.

IRT analysis using BILOG-MG in phase 3 informs the participants' ability parameters. Participants' ability has a scale of $-3 < \theta < 3$. Based on the output of the 3 BILOG programs, information is obtained that the average ability of students is -0.0143 with empirical reliability of 0.6914. The mean ability that has a minus value indicates that most students tend to have low abilities. The results of the analysis of students' abilities are presented in the form of graphs as shown in Figure 3.

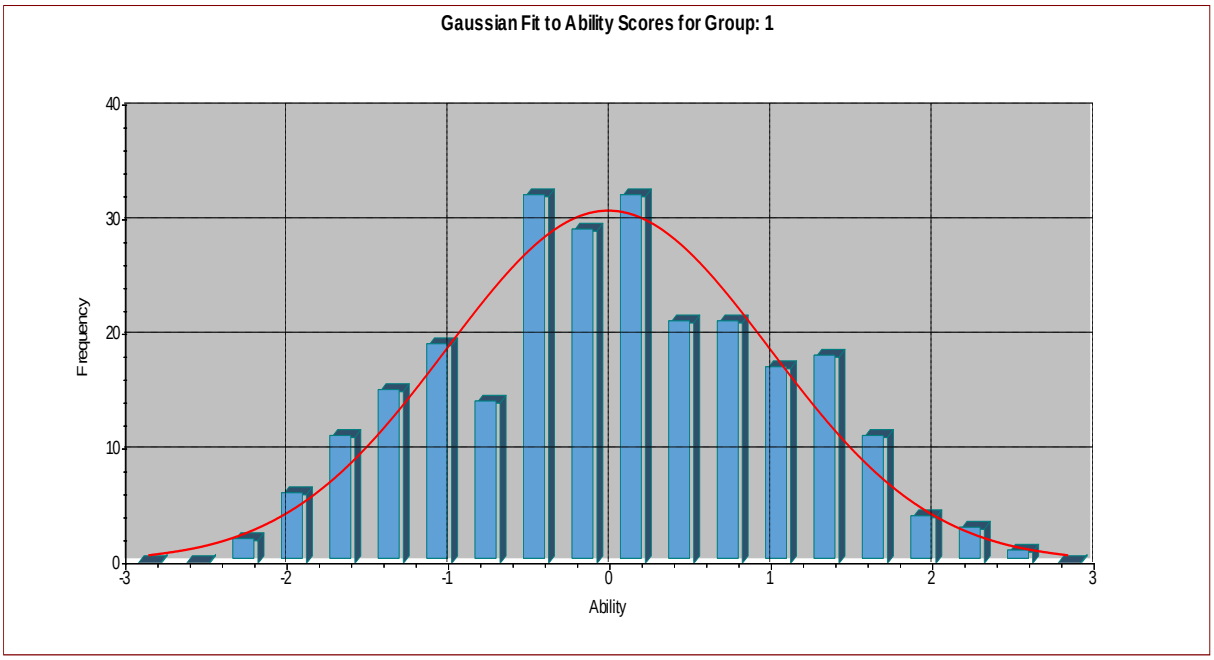

Figure 3. Graph of Student Ability Estimation

Based on the results of the IRT analysis in phase 3, it shows the magnitude of the participants' abilities, the results of the participants' abilities are still in the form of a logit scale. The form of this scale cannot be easily understood by the public. Then the logit scale needs to be changed to another form. In this study, the transformation model used is WITs. The WITs scale transformation equation is as follows:

$$W_\theta = 500 + 45{,}5\ \theta$$

After the scale is made, it is interpreted by making criteria in the scientific literacy of students. The criteria for the ability of participants are made in 3 criteria, namely high, medium and low. The ability of participants in this case is the scientific literacy of students on the theme of the solar system. The criteria made with the highest score of the participant's ability is 636.5 for the participant's ability scale (θ) of 3. While the lowest value of the participant's ability is 364 for the participant's ability scale (θ) of -3. After the criteria for the scientific literacy ability of students are made, then the classification of the ability of participants in scientific literacy on the theme of the solar system is carried out. The results of the participants' ability in scientific literacy on the theme of the solar system were tested on 256 students, the results are shown in Table 2.

Table 2. Classification of Students' Scientific Literacy Ability Criteria

Criteria	Number of Students
High	26
Currently	197
Low	33

Based on the classification results obtained information as much as 10.15% of students have high abilities, 76.95% have moderate abilities and 12.90% have low abilities in scientific literacy on the theme of the solar system. From the results of IRT analysis obtained data that the average ability of students is moderate and tends to be low, this is because the average value of students is minus (- 0.0143). The scientific literacy ability of students obtained from the results of measurements using the evaluation tool developed in this study is the same as the results of measurements carried out by other studies.

Likewise with the results of the latest PISA study, namely in 2018 based on OECD data (2018), Indonesia is ranked 70th out of 78 countries with an average score of 396. These results indicate that students' scientific literacy is low because it is below the average PISA scores (Hwang et al., 2018; Kastberg et al., 2016; Udompong & Wongwanich, 2014). The low scientific literacy ability of students can be caused by several things. There are several factors of the low ability of students including the selection of textbooks, misconceptions, non-contextual learning, low reading skills and the learning environment and climate (Fuadi, 2020); Agustina, 2017); Edge et al., 2011).

CONCLUSION

Based on the results of research on the development of evaluation tools to measure students' scientific literacy on the theme of the solar system, among others, from the development of evaluation tools, 30 items that met the criteria of content validation, reliability, level of difficulty and distinguishing power were obtained. Modern test theory analysis or Item Response Theory (IRT) was carried out using a 2-parameter logistic model (2-PL), obtained 29 questions that met the criteria of biserial correlation, level of difficulty and differentiation of problem. The results of output 3 of the Item Response Theory (IRT) analysis obtained information that the average ability of students was -0.0143 with empirical reliability of 0.6914. The average ability with a minus value indicates that most of the students' scientific literacy profiles on the theme of the solar system tend to have low abilities.

REFERENCE

Agustina, D. (2017). A complex system of teachers' beliefs and practices in developing learner autonomy in Indonesian junior high school contexts: A mixed-methods study.

Aji, S., Wiyanto, & Nugroho, S. E. 2015. Pengembangan Asesmen untuk Mengukur Kemampuan Memecahkan Masalah, Bekerjasama dan Berkomunikasi Calon Guru Fisika. *Journal of Innovative Science Education,* 4(2): 24-33.

Arikunto, S. 2012. *Dasar–dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.

Edge, D., Searle, E., Chiu, K., Zhao, J., & Landay, J. A. (2011, May). MicroMandarin: mobile language learning in context. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3169-3178).

Fuadi, H., Robbia, A.Z., Jamaludiadin, & Jufri A.W. 2020. Analisis Faktor Penyebab Rendahnya Kemampuan Literasi Sains Peserta Didik. *Jurnal Ilmiah Profesi Pendidikan,* 5(2): 108-116.

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational research review*, *17*, 50-62.

Hwang, J., Choi, K. M., Bae, Y., & Shin, D. H. (2018). Do teachers' instructional practices moderate equity in mathematical and scientific literacy?: An investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*, *16*(1), 25-45.

Jufri, W. 2013. *Belajar dan Pembelajaran Sains.* Bandung: Pustaka Reka Cipta.

Kastberg, D., Chan, J. Y., & Murray, G. (2016). Performance of US 15-Year-Old Students in Science, Reading, and Mathematics Literacy in an International Context: First

Look at PISA 2015. NCES 2017-048. *National Center for Education Statistics*.

OECD. 2016. *PISA 2015 Science Framework.*

OECD. 2016. *PISA 2015 Results in Focus.*

Park, J. Y. (2013). All the ways of reading literature: Preservice English teachers' perspectives on disciplinary literacy. *English Education*, *45*(4), 361-384.

Rusilowati, A., Kurniawati, L., Nugroho, S.E., & Widyatmoko, A. 2016. Developing an Instrument of Scientifi Literacy Assesment on the Cycle Theme. *International Journal of Environmental & Science Education*, 11(12): 5718-5727.

Rusilowati, A., 2014. *Pengembangan Instrumen Penilaian*. Semarang: Unnes Press.

Sugiyono, 2013. *Metode Penelitian Pendidikan.* Bandung: Alfabeta.

Udompong, L., & Wongwanich, S. (2014). Diagnosis of the scientific literacy characteristics of primary students. *Procedia-Social and Behavioral Sciences*, *116*, 5091-5096.