

## **Developing Assessment Instrument of Data, Technology, and Human Literacy in Physics Learning**

**Candra Dewi<sup>1</sup> ✉, Ani Rusilowati<sup>2</sup>, Fianti<sup>2</sup>**

<sup>1</sup>. SMK Texmaco Semarang, Indonesia

<sup>2</sup>. Universitas Negeri Semarang, Indonesia

### **Article Info**

Article history:  
Received 25 January  
2018  
Approved 12 March  
2019  
Published 25 August  
2019

Keywords:  
Data literacy,  
technology literacy,  
human literacy,  
validity

### **Abstract**

This study aims to determine the validity of the assessment instrument data, technology, and human literacy in physics learning. The research method uses research and development (RnD) with a modified 4-D model until the development stage. The instrument of validation was assessed by 6 validators. The instrument was tested by 36 students of class X SMK Texmaco Semarang. The form of instruments is a multiple choice test with 21 items to measure the cognitive domain. Data analysis in this research are tests of content validity, reliability, internal consistency, discrimination power, and level of difficulty. The results of the content validity by using Aiken's coefficient get an average score of 0.84 with a valid category at a significance level of 0.05. Cronbach's Alpha and ICC values from the reliability test are 0.77 and 0.806 with reliable categories. The results of internal consistency showed that 80.95% of items had good internal consistency. There are 4 items with categories not used from the discrimination power test. Furthermore from the results of the difficulty level, 7 items are easy, 13 items are medium, and 1 item is difficult. Based on the results of the instrument analysis can be concluded that the instrument developed was valid and reliable than can be used to measure the ability of data, technology, and humans literacy in physics learning.

✉ Correspondence Address:  
Jalan Raya Mangkang Km.16, Kota Semarang, Jawa Tengah 50155,  
Indonesia  
E-mail: candradewi875@gmail.com

## INTRODUCTION

Education has entered the 21st century in the era of the industrial revolution 4.0 accompanied by rapid and sophisticated development of science and technology. As many as 75% of work involves the ability of science, technology, engineering and mathematics, the internet of things, and lifelong learning in the era of the industrial revolution 4.0 (Zimmerman, 2018). One way to face up these challenges by developing new literacy skills including data literacy, technology literacy, and human literacy (Aoun, 2017).

Education 4.0 aligns people with technology to respond to the needs of the industrial revolution 4.0. Education must move from old literacy (reading, writing, and counting) to new literacy (data, technology, and human). Data literacy, technology literacy, and human literacy are translations of digital literacy. The conceptualization of digital literacy has been characterized by the integration of basic technical skills, media analysis as an object, critical of content and technology, and learning strategies by using technology (Bjørger & Erstad, 2015). The study of Ainley et al. (2016) reviewed the definitions of digital literacy and ICTs (Technology, Information, and Communication) that have been adopted in cross-national studies, investigated approaches to digital literacy and ICT assessment and articulated criteria that should guide the development of global measures of digital literacy and ICT skills.

Kerpati (2011) suggested that an important component in digital literacy skills in using computers and get the better of ICT. ICT skills can provide authentic learning opportunities to prepare students to compete in an increasingly digital world globally (USDOE, 2011; P21, 2011; NRC, 2011). Therefore important for educators to have access to actions in presenting valid and reliable evidence of the acquisition of ICT skills (Huggins et al., 2014). Determination of the success rate these skills to support learning in the 4.0 education era by conducting assessments and measurements use evaluation tools.

The results of observation at SMK Texmaco Semarang state that the student evaluation system since 2016 has been digitally used computer-based tests and now starting in 2019 has switched by using

smartphone assisted tests. However, the results of the analysis found that the questions tested did not contain analytical questions from digital information processing to practice digital thinking skills. This skill is that students not only can operate ICT devices but can also manage digital data, develop technology, and communicate. Thus, it can be said that the implementation of digital evaluation at SMK Texmaco Semarang is still at the stage of training students' skills to use ICT tools.

This research was conducted to develop instruments for assessment of the data, technology, and humans literacy in learning physics for oscillation, wave, and sound material. The focus of this study is to develop instruments and analyze validity and reliability. This is to present information about the validity of the instrument before applied to evaluate learning.

## METHODS

Research and development (R & D) uses a 3-D model (an adaptation of 4-D Models), consists of define, design, and develop. The trial of the assessment instrument was conducted at the SMK Texmaco Semarang in the academic year 2019/2020, with 36 students as the test subjects.

The form of learning assessment instruments developed was a multiple-choice test with 21 items. This instrument was tried out after implementing physics learning by using PhET simulation media. The stage of the trial instrument includes validation testing by judgment expert and empirical testing. Analysis of the trial results using Aiken's formula for content validity, Interclass Correlation Coefficient (ICC), and Alfa Cronbach's values for reliability, internal consistency, difficulty level, and discrimination power.

Content validity examines test items to determine the relevance of instrument content (Kowsalya et al., 2012). Aiken's standard value is determined through the number of validator questions, there are 6 validators with a rating of the scale of 1 – 4. Based on Aiken's table, the minimum value for an instrument to be valid with the significance level of 0.05 is 0.78. Aiken's value is determined through equation (1),

$$V = \frac{\sum r - l_0}{[n(c - l_0)]}$$

*V*: Aiken's validity  
*r*: the number given by rater  
*l*<sub>0</sub>: low validity rating  
*n*: the number of rater  
*s*  
*c*: high validity rating

Reliability analysis using IBM Statistic 22 software. Then the level of difficulty of the items is analyzed to find the categories of items, (2) including difficult, medium, or easy. The level of difficulty analysis is determined through equations (2),

$$TK = \frac{\text{Number of students who answer the right}}{\text{number of students who doing the test}}$$

**Table 1.** The Level of Difficulty Classification according to Rusilowati (2014: 35)

Interval	Criteria
0.00 ≤ TK < 0.30	Difficult
0.30 ≤ TK < 0.70	Medium
0.70 ≤ TK ≤ 1.00	Easy

Arikunto (2013: 226) said that the discrimination power is the ability to discriminate between smart and less clever students. According to Rusilowati (2014: 38), the calculation of discrimination power can be done using the following calculation (3).

$$DP = \frac{2(BA - BB)}{N}$$

where *DP* is discrimination power, *BA*: the number of the correct answer in the upper group, *BB*: the number of the correct answer in the lower group, *N*: the number of students taking the test.

**Table 2.** The Discrimination Power According to Rusilowati (2014: 38)

Interval	Criteria
0.40 < DP ≤ 1.00	Accepted
0.30 < DP ≤ 0.40	Accepted with few corrections
0.20 < DP ≤ 0.30	Corrected
0.00 ≤ DP ≤ 0.20	Not used

**RESULTS AND DISCUSSION**

**Instrument Development**

In learning the industrial revolution 4.0, digital literacy skills are clarified by the terms data literacy, technological literacy, and human literacy. Data literacy includes managing and analyzing information, technological literacy consists of work


principles and tool operations, and human literacy is design and communication (Aoun, 2017). The development of data, technology, and human literacy assessment instruments contain cognitive levels of applying (C3), analyzing (C4), evaluating (C5), and creating (C6). The distribution of each aspect of the instrument is seen in Table 3.

**Table 3.** The Aspect of Literacy in the Assessment Instrument

Literacy	Indicator of Literacy	Cognitive Level	Number of Items
Data Literacy	Organizing data	C3	2
	Interpreting data	C4	2
	Making a conclusion	C5	3
Technology Literacy	Applying concepts in technology	C3	3
	Determining formula	C3	1
	Evaluating technology	C5	3
Human Literacy	Rearranging concept	C6	2
	Combining concepts	C6	3
	Designing concepts	C6	2
Total			21

The examples of item instruments from the literacy data aspect with the indicator are concluding can be seen in Figure 1.

7. The results of the pendulum experiment are presented in the following figure!



The pendulum masses 1 and 2 are 0.5 kg and 0.25 kg, while the length of the pendulum ropes 1 and 2 are 0.7 m and 1 m. Experiments carried out by giving a deviation of  $5^\circ$  on both pendulum and let go together. The results of experiments on these images can be interpreted that ...

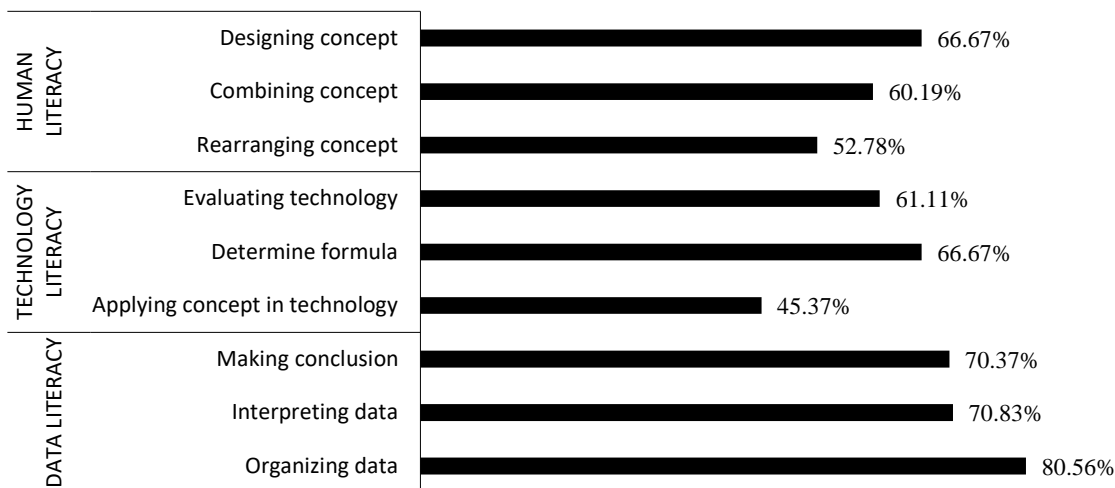
A. pendulum motion 1 is faster than pendulum motion 2  
 B. pendulum motion is slower than pendulum motion 2  
 C. pendulum 1 and 2 move together hand in hand  
 D. pendulum mass affects the frequency of the pendulum harmonic motion  
 E. the pendulum rope affects the frequency of the pendulum harmonic motion

**Figure 1.** The Example of Literacy Data Item

Figure 1 is an example of an item instrument with a cognitive level evaluating (C5). This item contains aspects of data literacy that are making conclusions from digital data analysis. Achievement indicators from this item are students able to prove the concept of vibration based on the results of

practicum data. Practicum activities doing by using digital labs based on Android with Physics Education Technology (PhET) simulations.

The results of the trial test obtained that the average score from each indicator of the literacy aspect is shown in Figure 2.



**Figure 2.** The Average Score of Indicator Items

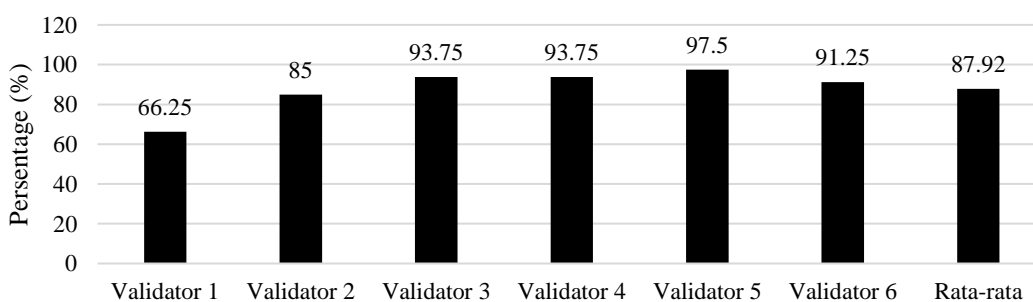
Figure 2 shows the highest average score is in the aspect of data literacy with an average percentage of 73.92%. The average percentage in the aspect of technological and human literacy was 57.72% and 59.88%. These results indicate that the ability of technology and human literacy is still low. The application of science in low technology shows low science knowledge (Rusilowati et al., 2016; Ibrahim & Aspar, 2011). The results of this score can also be used as a basis for evaluating the learning methods used in physics.

**Instrument Validity**

The validity of the instrument is the initial trial stage before the instrument being tested for research subjects. Data, technology, and human literacy assessment instruments that have been

developed are validated by 6 validators consisting of 3 lecturers and 3 teachers who have obtained a Masters's degree. Analysis of instrument validity includes the feasibility and content validity using the Aiken coefficient.

The percentage of feasibility shows the results of the judgment expert from the validator. The results of the percentage of feasibility are presented in Figure 3. Information from Figure 3 is the average percentage of the feasibility of 6 validators is 87.92% with very valid criteria and can be used without correction. These results indicate that the instrument is feasible for use in evaluating physics learning. However, even though the average results indicate a high level of feasibility, improvements are still needed according to the validators' recommendations.



**Figure 3.** The Percentage of Validity Results by Validator

Then the validation results are analyzed to determine the value of content validity based on Aiken's coefficient. There are 3 aspects of assessment that is material, construction, and language aspects. In the material aspect, there are 6

items, the construction aspect has 10 items, and in the language aspect, there are 4 items. The specified value interval is 1–4 with the lowest criteria which is not good, good enough, good, and very good. The results of Aiken's coefficient can be seen in Table 4.

**Table 4.** The Results of the Aiken's Analysis

Aspect	Aiken's Value	Criteria
Material	0.66	Invalid
Construction	0.90	Valid
Language	0.96	Valid
Average	0.84	Valid

Table 5 present the Aiken's value in the material aspect having the lowest value which is below 0.78 so that has invalid criteria. Unlike the case in the construction and language aspects which have Aiken's value of more than 0.78 with valid criteria. However, if the average overall value of Aiken's is 0.84 with valid criteria.

The content validity results show that in terms of material, the assessment instruments developed by following assessment indicators. Thus, it is necessary to correct the content before the item is tested on further research subjects. Content validity tests to ensure that the assessment instruments are relevant to the objectives to be achieved (Hendryadi, 2014).

The results of expert judgment are used as a basis for making improvements to the instrument,

especially for material aspects. Then the results of improvement were trialed to the research subjects. These results were analyzed to determine reliability, internal consistency, difficulty level, and discrimination power. This analysis for after the instrument is corrected according to the suggestions from validators.

### Instrument Reliability

The instrument reliability test to determine the level of confidence of the test. There are two stages in the reliability test, to determine the validation results reliability from the validator and the reliability from the students' results. The reliability test of the assessment instrument uses IBM Statistic 22 software and the results are shown in Table 6.

**Table 6.** The Results of Reliability Instrument

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	ICC	N of Items
0.770	0.767	0.806	21

The results in Table 6 show that the overall reliability value of the items gets a Cronbach Alpha value of 0.770 in a good category. This is suitable for the standard value that a good Cronbach Alpha value for a measuring instrument is between 0.6 to 0.8 (Sufren & Natanel, 2012). Another result is the ICC value from 6 rater gets a score of 0.806 with a satisfactory category. This means that the validation results from 6 validators are consistent and reliable.

Furthermore, the reliability test is complemented by an internal consistency analysis of each item. The internal consistency value of each item is shown in Figure 4.

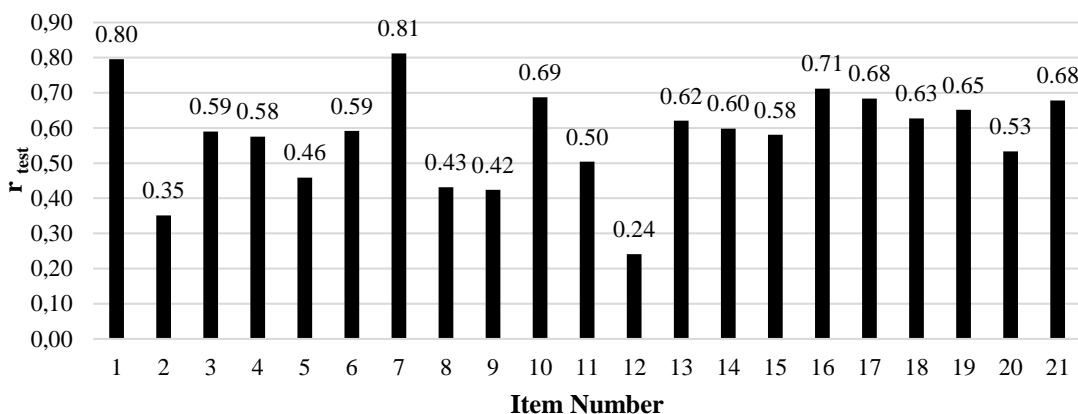


Figure 4. The Results of Internal Consistency

An instrument is consistent and reliable based on an internal consistency test if it has a  $r_{test}$  greater than  $r_{table}$  ( $r_{table} = 0.44$ ). Total of 4 items which is number 2, 8, 9, and 12 have a value of  $r_{test}$  smaller than  $r_{table}$ . This means that the 4 items are not correlated and are not equal (Khumaedi, 2012).

Quantitative evidence of the internal consistency results shows that 80.95% of items have good and reliable consistency. Internal consistency tests are used in managing the test with the same material and connecting the measurement results at

any time to obtain the level of instrument reliability (Hajjar, 2018).

Analysis results of the difficulty level found that 1 item was difficult, 7 items were easy, and 13 items were medium. One difficult item is question number 9 with the category C3 to measure technological literacy abilities. This item is difficult because the elaboration of answer choices is too long and difficult to understand. The results of the measurement of the difficulty level are presented in Table 7.

Table 7. Results of the Difficulty Level

Interval Value	Criteria	Item Number	Total Item	Percentage (%)
0.00 – 0.29	Difficult	9	1	4.76
0.30 – 0.69	Medium	4,5,7,8,10,11,13,14,15, 16,17, 18,19,21	14	66.67
0.70 – 1.00	Easy	1,2,3,6,12,20	6	28.57
	Total		21	100

Good questions are not too difficult and not too easy to do. The trial results show that the development of the instrument is good based on the level of difficulty. This is because of the 21 items there was only 1 item difficult for students. Besides, the comparison of the number of medium items is also higher than easy items.

Discrimination power analysis to determine whether the items can discriminate between students who have understood the concept and students who have not understood the concept. The results of the items discrimination power analysis are presented in Table 8.

**Table 8.** The Results of the Discrimination Power

Interval Value	Criteria	Item Number	Total Item	Percentage (%)
$0.40 < DP \leq 1.00$	Accepted	7,14,18,19,20,21	6	28.57
$0.30 < DP \leq 0.40$	Accepted with few corrections	1,3,10,17	4	19.05
$0.20 < DP \leq 0.30$	Corrected	4,5,6,8,13,15,16	7	33.33
$0.00 \leq DP \leq 0.20$	Not used	2,9,11,12	4	19.05
Total			21	100

Based on the analysis results in Table 8, it was found that 6 items were accepted, 4 items were accepted with a few corrections, 7 items were corrected, and 4 items were not used. A total of 4 items that are not used are numbers 2, 9, 11, and 12. This item is not used as a discrimination value of

less than 0.2. These items are categories not good so they cannot be used for further research.

Then the results of the internal consistency, the level of difficulty, and the discrimination items are compared and analyzed more. The results of the comparison are shown in Table 9.

**Table 9.** Comparison of Internal Consistency, Difficulty and Discrimination Power

Item Number	Internal Consistency	Level of Difficulty	Discrimination Power
2	Inconsistent	Easy	Not used
8	Inconsistent	Medium	Corrected
9	Inconsistent	Difficult	Not used
11	Consistent	Medium	Not used
12	Inconsistent	Easy	Not used

Based on the results in Table 9, it was found that 5 items were not used are numbers 2, 8, 9, 11, and 12. So the number of items instruments can be used to assess data, technology, and human literacy for material oscillations, waves, and sound are 16 items.

A valid test is needed as a tool to evaluate the effectiveness of the learning process (Perdana, 2019). This research presents the validity results of physics learning instruments to assess the ability of data literacy, technology, and humans. Learning assessment strategies can be done through providing learning success information, conducting formal and informal assessments, and then making feedback to improve the quality of learning (Cauley & McMillan, 2010; Ruiz-Primo, 2011; O'Leary et al., 2013; William & Leahy, 2015; Kippers et al., 2018).

The assessment strategies in this study include these three strategies by using digital learning media. The use of instructional media namely PhET can encourage increased understanding of the material so that the results can

provide information on learning success. The selection of PhET learning media to develop data, technology, and digital literacy skills is also determined through surveys of student conditions and the classroom environment. This is following with Rachmawati (2017) that effective learning takes place taking into account the conditions of students and the school environment.

## CONCLUSION

The form of the instrument developed was a multiple-choice test with option 5, to measure data, technology, and human literacy capabilities. This instrument has 3 aspects with 9 indicators are organizing data, interpreting data, making conclusions, applying concepts in technology, determining formulas, evaluating technology, rearranging concepts, combining concepts, and designing concepts.

The validity results show that the assessment instrument has been declared feasible and can consistently measure the literacy capabilities of



data, technology, and human. The results of the instrument feasibility level of 6 validators were 87.92% with very feasible criteria. The average value of Aiken's aspects is 0.84 with valid criteria. The Cronbach Alpha value of the reliability test results is 0.770 in the good category. This result is supported by the existence of an ICC value of 6 validators is 0.806. Furthermore, the correlation test results show that 5 items are invalid because they have a correlation value below 0.20. Of the 5 items, 4 of them should not be used based on the results of discrimination power. The difficulty level test results stated that of the 21 questions, 1 item was difficult, 7 items were easy, and 13 items were medium. So, the instrument that can be used to measure the literacy of the data, technology, and human consists of 16 questions, and all indicators are represented.

## SUGGESTIONS

Suggestions from this research are to be able to increase the number of research subjects in the trial of the instrument so that the reliability can be higher. The ability of data literacy in particular to apply concepts and human literacy indicators rearranges concepts not yet well mastered by students. Therefore, when studying physics material needs to emphasize the ability to apply and rearrange the concepts.

## REFERENCES

- Ainley, J., Schulz, W., & Fraillon, J. (2016). A global measure of digital and ICT literacy skills. *Physical Review C - Nuclear Physics*, *134*, 1–21.
- Akbar, S. 2013. *Instrumen Perangkat Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Aoun, J. E. 2017. *Robot-Proof Education in the Age of Artificial Intelligence*. US: MT Press.
- Arikunto, S. 2013. *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka Cipta, Jakarta.
- Björger, A. M., & Erstad, O. (2015). The connected child: tracing digital literacy from school to leisure. *Pedagogies*, *10*(2), 113–127. <https://doi.org/10.1080/1554480X.2014.977290>.
- Cauley, K. M., & McMillan, J. H. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, *83*(1), 1–6. <https://doi.org/10.1080/00098650903267784>.
- Hajjar, S. T. EL. (2018). Statistical Analysis: Internal-Consistency Reliability and Construct Validity. *International Journal of Quantitative and Qualitative Research Methods*, *6*(1), 27–38. <https://doi.org/ISSN 2056-3639>.
- Hendryadi. (2014). Content Validity (Validitas Isi). *Teorionline Personal Paper*, *01*(01), 1–5.
- Huggins, A. C., Ritzhaupt, A. D., & Dawson, K. (2014). Measuring Information and Communication Technology Literacy using a performance assessment: Validation of the Student Tool for Technology Literacy (ST2L). *Computers and Education*, *77*, 1–12. <https://doi.org/10.1016/j.compedu.2014.04.005>
- Kerpati, A. (2011). *Digital Literacy in Education*. Moscow: UNESCO Institute.
- Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin Unnes*, Vol. 12, p. 129541.
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>
- Kowsalya, D. N., Lakshmi, V., & Suresh, K. P. (2012). Development and Validation of a Scale to assess Self-Concept in Mild Intellectually Disabled Children. *International J. Soc. Sci. & Education*, *2*(4), 2223–4934.
- National Research Council. (2011). *Assessing 21st Century Skills: Summary of a workshop*. Washington, DC: The National Academies Press.
- O'leary, M., Lysaght, Z., & Ludlow, L. (2013). A Measurement Instrument to Evaluate Teachers' Assessment for Learning Classroom Practices. *The International Journal*

- of Educational and Psychological*, 14(2), 40–60.  
<https://doi.org/10.13140/2.1.1289.0241>
- Partnership for 21st Century Skills. (2011). *Framework for 21st Century Learning*. Washington, DC. Retrieved from <http://www.p21.org/tools-and-resources/publications/1017-educators#defining> (retrieved on 29 November 2019).
- Perdana, R., Riwayani, R., Jumadi, J., & Rosana, D. (2019). Development, Reliability, and Validity of the open-ended Test to Measure Student's Digital Literacy Skill. *International Journal of Educational Research Review*, 4(4), 504–516.  
<https://doi.org/10.24331/ijere.628309>
- Rachmawati, S., & Susanto, H. (2017). Penggunaan Metode CRI (Certainty Of Response Index) Berbantuan Soal PISA (Programme Of International Student Assessment) untuk Mengidentifikasi Miskonsepsi IPA Materi Tata Surya. *Unnes Physics Education Journal*, 6(3), 26–31.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15–24.  
<https://doi.org/10.1016/j.stueduc.2011.04.003>.
- Rusilowati, A. (2014). *Pengembangan Instrumen Penilaian*. Semarang: Unnes Press.
- Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an instrument of scientific literacy assessment on the cycle theme. *International Journal of Environmental and Science Education*, 11(12), 5718–5727.
- Sudijono, A. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Pers.
- Sufren & Natanael, Y. (2014). *Belajar Otodidak SPSS Pasti Bisa*. Jakarta: PT Elex Media Komputindo.
- U.S. Department of Education. (2010). *National Educational Technology Plan 2010*. Retrieved from <http://www.ed.gov/technology/netp-2010> (retrieved on 29 November 2019).
- Wiliam, D. & Leahy, S. 2015. *Embedding Formative Assessment. Practical Techniques for K-12 Classroom*. Florida, USA: Learning Sciences International.
- Zimmermann, T. 2018. *Industry 4.0: Nothing is More Steady than Change. In Smart Grid Analytics for Sustainability and Urbanization*. Hershey: IGI Global.