*IOSR Journal of Research & Method in Education (IOSR-JRME)*
*e-ISSN: 2320–7388, p- ISSN: 2320-737x Volume 11, Issue 1 Ser. VII (Jan. – Feb. 2021), PP 41-47*
*www.iosrjournals.org*

# Validity & Reliability: Developing Instrument of Assessment to Measure Critical Thinking Skills in High School Physics

Alexander Andi Kurnianto[1*], Mundilarto[1]
[1](*Department of Physics Education, Yogyakarta State University, Indonesia)*

***Abstract:***
***Background****:Critical thinking skills were one of the things that must be developed in students learning. The instrument for assessing this skill should be good and feasible. The purposes of this study are: (1) determine the test quality of critical thinking skills based on content validity, (2) determine the test quality based on empirical testing.*
***Materials and Methods****: The research method used was the combination of R & D and 4D model. In this research, there are 12 items were designed to measure critical thinking skills. The validity of instruments had been validated by experts, practitioners, and peers. 6 items have been chosen for empirical testing which each item measures one of the critical thinking skills indicators. The validation results are analyzed by Aiken's V for content validity, then by QUEST and PARSCALE software to analyze the data of empirical testing.*
***Results****: Based on results of the analysis, known that (1) all item of critical thinking test meet content validity with Aiken's index between 0.88 to 1, (2) by QUEST, all item that has been tested were fit to partial credit model (PCM) and have reliability score 0.83; based on PARSCALE analysis, item difficulty index was on a good category which ranged from -1,234 to 1,853.*
***Conclusion:*** *Hence, the assessment instrument which developed is valid and reliable to measure critical thinking skill.*
***Key Word****: Critical thinking skills, assessment, validity, reliability*

---

Date of Submission: 16-02-2021 Date of acceptance: 02-03-2021

---

## I. Introduction

Critical thinking (CT) skill development is the main goal of education[1,2,3]. CT is a style of thinking about a problem, where the thinker capable to improve the thinking skill[4]. This thing can be built in the act of obtaining, develop, and train the ability to understand the statements used in the conclusions process[5]. CT depends on the ability, motivation, and opportunities that we owned[6]. Someone with good CT skills should be interested in arguing and give a good reason for an action and their belief about a thing[7]. The activity of CT consisting of three main issues, namely: analysis, evaluation, and arguing further[8]. CT can be defined as the ability to think logically and reflectively related to what has been done and decided, based on depth analysis. The used reasons for supporting a decision must make sense and be based on trusted references.

Most the teachers agree that students should develop their CT skill, but unfortunately, a lot of students have less developed on this skill[9], nevertheless, student's CT skill is needed to develop in school. Based on the research by Whiley, Witt, Colvin, Sapiains & Kotir, the new students at Queensland University, have been given the matriculation program for a year about their critical thinking skills[10]. The same thing could be done in Indonesia, but depends on the principal and teachers who have responsibility for the learning process[11]. The goal is to fulfill the needed to evaluate something critically, either in future work or in their social environment[12]. This means if at the pre-university level, that is high school, the students already have CT skills, it would be useful for their future lives.

CT is the most important skill in physics learning. Many previous researchers thought that CT was the base for science learning[13]. Ct skill includes: analysis; interpretation; evaluation; inferences based on the evidence and write the explanations; supposition and coherence; self-organized; and further explanation[14,15]. The indicators were used in this research are interpretation; analysis; inference; evaluation; explanation; and further argument.

The phase of developing CT skills cannot be done instantly. It needs to pay attention to many things, includes collecting all information about the student. Assessment can be used to obtain the information of students related to the process, progress, and their learning outcomes[16]. One of the challenges in learning is how to make an assessment for the process and/or result of learning. The assessment is needed to know how successful the learning process has been implemented. Without good assessments, we cannot be sure if the student has learned what is expected[17,18]. The biggest challenge is the moreyounger the student, the more

difficult to assess, related to the assessment accuracy and effectiveness[19]. Students in formal operational age are the target in this research because they were able to think dynamically and imagine abstract things[20]. This fits in with physics learning characteristic that abstract and needs a deep analysis. Besides using complex mathematical equations, physics demands the student to able to analyze the physics phenomenon, then they can solve the faced physics problems[21,22,23].

An assessment can be used to make decisions because it was the process of collecting information about student learning outcomes and the result[24]. If the instrument of assessment does not meet the quality standards, and then the result of the assessment cannot be used for evaluation. This is important for developing CT skills. The assessment is needed to determine how successful a process of learning has been done by the teachers and students. Many assessment types are often used on learning, such as written and oral assessment[25]. Written assessment, which good to be used in physics learning, is a test with questions and demands students to write the answers in detail. One type of written assessment is the essay test. The essay test is a test that demands responses from students who use to measure or obtain information about the student's knowledge of factual information and their ability[26].

Based on the explanation above, the essay test is a good alternative instrument to measure critical thinking skills. So, it is needed development of physics essay test to measure critical thinking skills in senior high school. Therefore, the aims of this research are: (1) to develop a physics essay test to measure critical thinking skills, and (2) to determine the test quality by the goodness of fit, reliability, and item's difficulty index.

## II. Material and Methods

This development research was carried out on SMAN 6 Yogyakarta, Special Region of Yogyakarta, Indonesia from August 2019 to August 2020. A total 206 students of 12th grade of science was the subject for this study.

**Study Design:** This research is development research. The research method used was the combination of R & D and 4D model. R & D model consists of research and development[27]. The development stage of this model uses the 4D model. The 4-D model consists of define, design, develop, and disseminate[28]. The combination of these two models resulted in a development stage consists of research, define, design, develop, and disseminate. The diagram combination of both models can be seen in Figure 1.

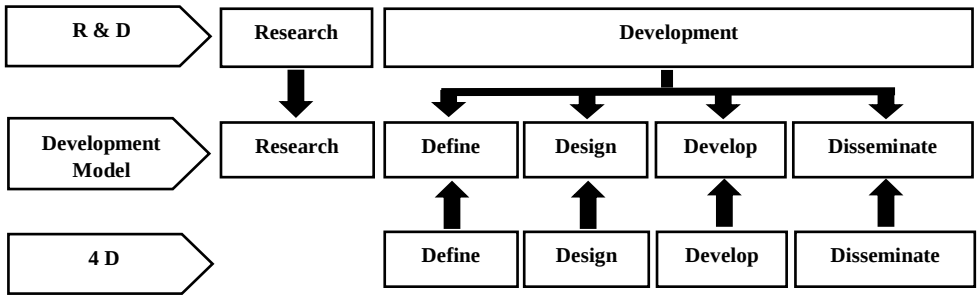


**Figure 1**. Diagram of combination between R & D and 4-D

**Study Location**: The study was done in SMAN 6 Yogyakarta, Special Region of Yogyakarta, Indonesia.
**Study Duration:** August 2019 to August 2020.
**Sample size:** 206 students.
**Sample size calculation:** The sample distribution was shown in table 1.

**Table 1.** *Sample Distribution*

| *Class* | *Students* | *Class* | *Students* |
|---|---|---|---|
| 1st 12th Science | 26 | 5th 12th Science | 20 |
| 2nd 12th Science | 29 | 6th 12th Science | 21 |
| 3th 12th Science | 30 | 7th 12th Science | 26 |
| 4th 12th Science | 27 | 8th 12th Science | 27 |

**Subjects & selection method**: The subject used in this research is 12th grade of science students of senior high school in Yogyakarta.The sample is determined by purposive sampling which consists of 206 students of SMAN 6 Yogyakarta in the 2019/2020 school year. Purposive sampling was used to obtain samples consisting of the high, medium, and low-class groups to obtain a logistical curve.

**Procedure methodology**

The procedures for each stage can be seen in Figure 2.

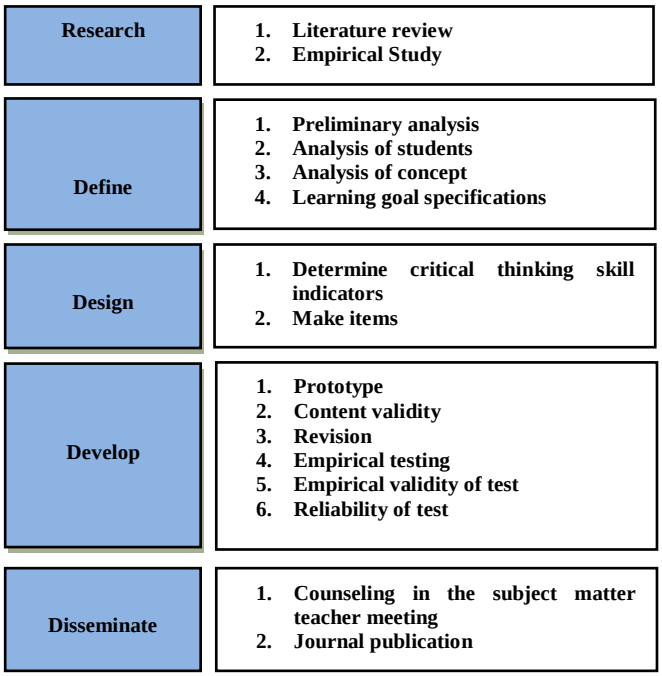


**Figure 2**. Diagram of procedure development model

**Statistical analysis**

The validation results are analyzed by Aiken's V for content validity. The polytomous data with four categories was analyzed using Item Response Theory (IRT) according to the Partial Credit Model (PCM) using QUEST and PARSCALE program. The QUEST program is used to determine the goodness of fit and reliability. PARSCALE program is used to determine an item's difficulty index, information function, and standard error measurement (SEM).

## III. Result and Discussion

**Planning the assessment instrument**

The physics essay test consists of 12 items of CT skill. The indicator CT skill includes interpretation, analysis, inference, explanation, evaluation, and further argue. The development of this test is aimed at the physics matter of sound waves. The indicator description of CT skill was shown in Table 2.

**Table 2.** Indicator of Critical Thinking

| Topic | Indicator | Indicator Description |
|---|---|---|
| Critical Thinking | Interpretation | Students able to understand mean, which in this case relate to the delivery of impression, opinion, or theoretical views about physics phenomenon are presented |
| | Analysis | Students able to identify relatedness or connection between the one concept and the other |
| | Inference | Students able to conclude related physics phenomenon that has been observed, understood, and analyzed |
| | Explanation | Students able to explain regarding conclusion has been taken through the arguments based on facts and based on a trusted |

| | | |
|---|---|---|
| | Evaluation | Students able to determine how well an argument and/or reason that supports a conclusion; even how strong the facts which form the basis of the argument |
| | Further argument | Students able to provide opinions and other explanation (additional) to support of his belief decision has been taken, or to oppose a decision or claims that have been made |

The item distribution is presented in Table 3. The items that have been made are validated by instrument experts, practitioners and peers amount 8 raters.

**Table 3.** Items Distribution for Validation

| *Topic* | *Indicator* | *Physics Matter (sound wave)* | | | | |
|---|---|---|---|---|---|---|
| | | The characteristics and velocity of sound wave | Doppler Effect | String phenomenon and *Organa* pipe | Sound intensity | Sound intensity level |
| Critical Thinking | Interpretation | 6 | | 8 | | |
| | Analysis | | | 5 & 7 | | |
| | Inference | | | | 9 & 12 | |
| | Explanation | 2 | 3 | | | |
| | Evaluation | 1 | | | | 11 |
| | Argue further | | 4 | | | 10 |

**Constructing the assessment instrument**

Items with 3 rating categories and 8 raters are declared valid if the Aiken's validity coefficient is $V \geq 0{,}88$[29]. The result of all CT skill items is valid with Aiken's *V* index between 0.88 to 1. But, according to the experts, four items are valid with revision. The revision details were shown in Table 4.

**Table 4.**Aikens'Validation Result for Critical Thinking Item

| **Item** | **Aikens' V** | **Criteria** | **Revision** |
|---|---|---|---|
| 1 | 1 | Valid | No revision |
| 2 | 1 | Valid | No revision |
| 3 | 0,94 | Valid with revision | Music instrument which called *sape'* should be replaced by another instrument that known better by the student, except the student already familiar with that thing (*sape'*) |
| 4 | 0,88 | Valid with revision | The concept of the Doppler effect should be used properly. Replace the "approaching sound source" with "away from sound source". |
| 5 | 1 | Valid | No revision |
| 6 | 0,94 | Valid with revision | In the test of sound wave velocity, replace the "$F$" by "$\sqrt{F}$" |
| 7 | 1 | Valid | No revision |
| 8 | 1 | Valid | No revision |
| 9 | 1 | Valid | No revision |
| 10 | 1 | Valid | No revision |
| 11 | 1 | valid | No revision |
| 12 | 0,88 | Valid with revision | In the test of sound intensity, replace the "$r$" by "$r^2$" |

Based on rater suggestion, the items were revised and assessment rubrics to be adjusted to the item's difficulty index. In the next steps, 6 items of 12 items were tested to 206 high school students in SMAN 6 Yogyakarta to see the characteristics of items about the goodness of fit, reliability, difficulty index, information function, and SEM. Each item of 6 items is used to measure each CT skill's indicator. The item distribution for the empirical test was shown in Table 5.

**Table 5.**Items Distribution for Empirical Test

| Item | Item number | CT skill indicators |
|---|---|---|
| 1 | 1 | Evaluation |
| 2 | 3 | Explanation |
| 3 | 5 | Analysis |
| 4 | 6 | Interpretation |
| 5 | 10 | Further Argue |
| 6 | 12 | Inference |

**Testing the assessment instrument**

Item or testee is fit to the model if INFIT MNSQ value is between 0.77 to 1.30[30].Based on Figure 3, 6 items are fit to the model with INFIT MNSQ values of items between 0.83 to 1.26.

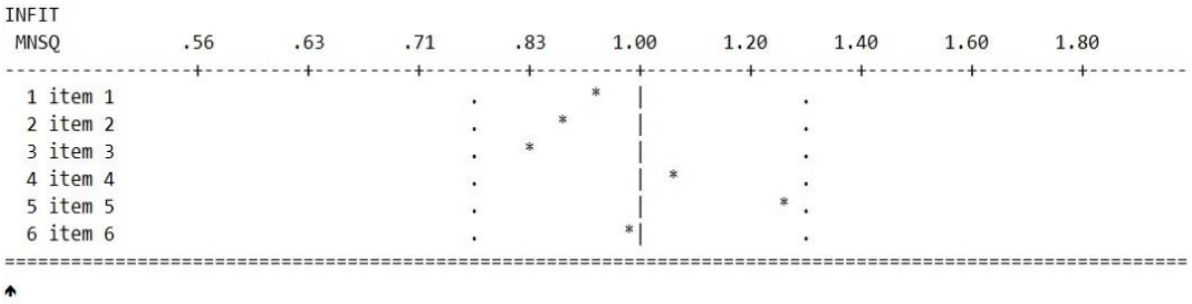


**Figure3.**Goodness of Fit of Item

The detail of items INFIT MNSQ values is shown in Table 6.

**Table 6.**Items INFIT MNSQ value

| Item | INFIT MNSQ | Criteria |
|---|---|---|
| 1 | 0.92 | Fit with model |
| 2 | 0.88 | Fit with model |
| 3 | 0,83 | Fit with model |
| 4 | 1.05 | Fit with model |
| 5 | 1.26 | Fit with model |
| 6 | 0.98 | Fit with model |

**Reliability**

Reliability of item estimates indicates the reliability of items or samples. The analysis result shows the reliability of case estimates of 0.61. It means that measurements provide consistent results when the instrument is used to measure at different times. Another result shows the reliability of the item estimate of 0.83, which means the test sample fit by item tested, or the sample provides information as expected.

The item is good if it has the item's difficulty index between -2.00 to 2.00[31]. Based on Figure 4, the overall item is good because it has an item's difficulty index range of -1.234 to 1.853. Item 1 whose difficulty index of -1.234 indicates that this item is very easy. And, item 4 whose difficulty index of 1.853 indicates that this item is very difficult. The result of the items difficulty index was shown in Figure 4. Pay attention to the *location* column.

| ITEM | BLOCK | SLOPE | S.E. | LOCATION | S.E. | GUESSING | S.E. |
|---|---|---|---|---|---|---|---|
| 0001 | 1 | 0.425 | 0.009 | -1.234 | 0.053 | 0.000 | 0.000 |
| 0002 | 2 | 0.425 | 0.009 | 0.097 | 0.058 | 0.000 | 0.000 |
| 0003 | 3 | 0.425 | 0.009 | -0.088 | 0.050 | 0.000 | 0.000 |
| 0004 | 4 | 0.425 | 0.009 | 1.853 | 0.069 | 0.000 | 0.000 |
| 0005 | 5 | 0.425 | 0.009 | 1.823 | 0.058 | 0.000 | 0.000 |
| 0006 | 6 | 0.425 | 0.009 | 0.404 | 0.051 | 0.000 | 0.000 |

**Figure 4.** Difficulty Index of Item by PARCSALE

Based on the result of analysis using the PARSCALE program, it was obtained information function and standard error of measurement (SEM). Figure 5 shows that the test is suitable for the students who have the ability (θ) between -2 to + 3.2.

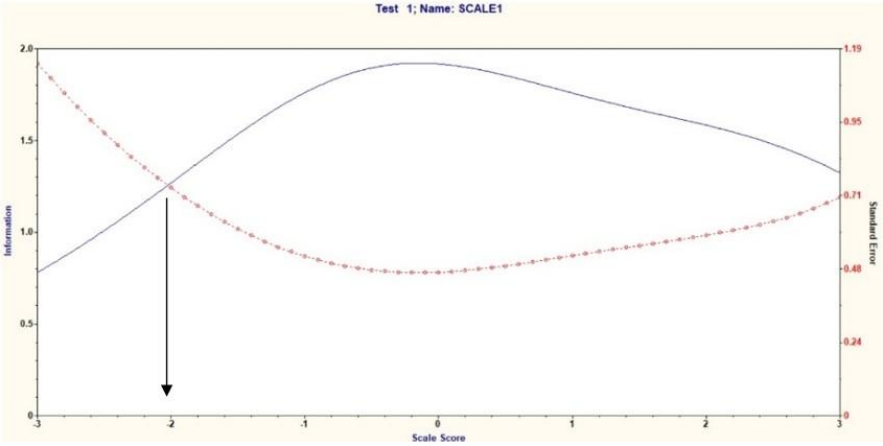


**Figure 5.** Test Information Curve

Figure 6 shows that 30 of 206 students have the ability score (θ) of + 0.25. It is the highest frequency obtained.

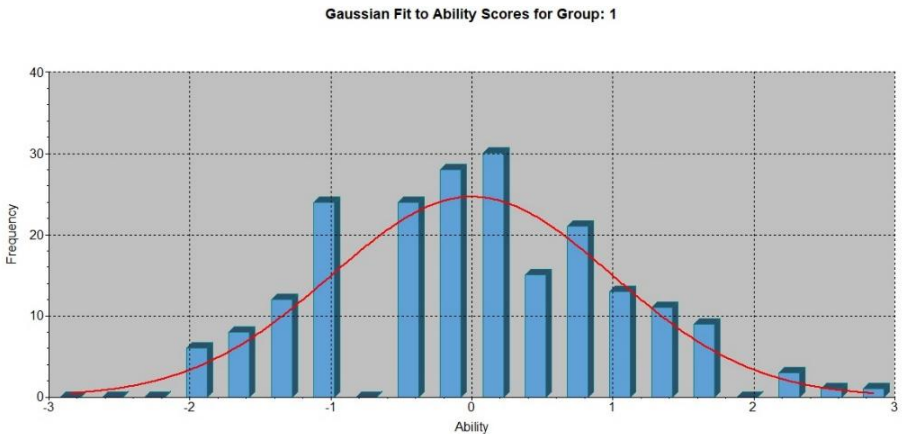


**Figure 6.** Ability Histogram

## IV. Conclusion

The physics essay test was successfully developed to measure the CT skills of students in high school. The CT skill indicators include interpretation; analysis; inference; evaluation; explanation; and further argument. The physics essay test consists of 12 items validated by experts, practitioners, and peers. 4 items are known valid with revision. Based on that validation, the critical thinking test meets content validity with Aiken's index between 0.88 to 1. For empirical testing, 6 items have been chosen which each item measures one of the critical thinking skill indicators. 6 items of physics essay test were fit with the Partial Credit Model

(PCM) based on the QUEST program's analysis. The physics essay test is qualified based on the reliability of item estimates of 0.61 and reliability of case estimates of 0.83. Item's difficulty index range of -1.234 to 1.853 based on PARSCALE program's analysis. Based on information function and SEM, teachers can use this physics essay test to measure the CT skill on students who have low ability to high ability of -2.00 to +3.20.

## References


[1]. Sanders, M., & Moulenbelt, J. Defining critical thinking. Inquiry: Critical Thinking Across the Disciplines. 2011; 26(1), 38–46. https://doi.org/10.5840/inquiryctnews20112616

[2]. Toy, B. Y., & Ok, A. Incorporating critical thinking in the pedagogical content of a teacher education programme: does it make a difference? European Journal of Teacher Education. 2012; 35(1), 39–56. https://doi.org/10.1080/02619768.2011.634902

[3]. Larsson, K. Understanding and teaching critical thinking—A new approach. International Journal of Educational Research. 2017; 84, 32-42. https://doi.org/10.1016/j.ijer.2017.05.004

[4]. Paul, R., & Elder, L. & F. for C. T. The miniature guide to critical thinking: Concepts and tools. 27th International Conference on Critical Thinking. 2012. https://cpb-us-e1.wpmucdn.com/blogs.uoregon.edu/dist/d/14812/files/2017/07/2006_paul_elder_concepts_tools-2frmy26.pdf

[5]. Mulnix, J. W. Thinking Critically about Critical Thinking. Educational Philosophy and Theory. 2012; 44(5), 464–479. https://doi.org/10.1111/j.1469-5812.2010.00673.x

[6]. Howard, L. W., Tang, T. L. P., & Jill Austin, M. Teaching Critical Thinking Skills: Ability, Motivation, Intervention, and the Pygmalion Effect. Journal of Business Ethics. 2014, 1–15. https://doi.org/10.1007/s10551-014-2084-0

[7]. Bowel, T., & Kemp, G. Critical Thinking A Concise Guide. New York: Routledge. 2015. https://doi.org/10.1017/CBO9781107415324.004

[8]. Butterworth, J., & Thwaites, G. Thinking Skills: Critical Thinking and Problem Solving. Cambridge: University of Lethbridge. 2013. https://books.google.com/books?id=o_NjrgmT_loC&pgis=1

[9]. Sanders, M., & Moulenbelt, J. Defining Critical Thinking. Inquiry: Critical Thinking Across the Disciplines. 2011; 26(1), 38–46. https://doi.org/10.5840/inquiryctnews20112616

[10]. Whiley, D., Witt, B., Colvin, R. M., Sapiains Arrue, R., & Kotir, J. Enhancing critical thinking skills in first year environmental management students: a tale of curriculum design, application and reflection. Journal of Geography in Higher Education. 2017. https://doi.org/10.1080/03098265.2017.1290590

[11]. Wartono, W., Hudha, M. N., & Batlolona, J. R. How are the physics critical thinking skills of the students taught by using inquiry-discovery through empirical and theorethical overview? Eurasia Journal of Mathematics, Science and Technology Education. 2018. https://doi.org/10.12973/ejmste/80632

[12]. Whiley, D., Witt, B., Colvin, R. M., Sapiains Arrue, R., & Kotir, J. Enhancing critical thinking skills in first year environmental management students: a tale of curriculum design, application and reflection. Journal of Geography in Higher Education. 2017. https://doi.org/10.1080/03098265.2017.1290590

[13]. Rodrigues, A., & Oliveira, M. The Role of Critical Thinking in Physics Learning. Core. 2018; 1–8. https://core.ac.uk/download/pdf/38301359.pdf

[14]. Oliver, R. W. W., & Engle. Handbook of Understanding and Measuring Intelligence. California, USA: SAGE Publications. 2005.

[15]. Ennis, R. H. The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions. University of Illinois. 2011; 1–8.

[16]. Pratama, A. C., Supahar, Warsono, & Jumadi. The development physics essay test to measure vector and mathematics representation ability in senior high school. Journal of Physics: Conference Series. 2018. https://doi.org/10.1088/1742-6596/1097/1/012013

[17]. Golnik, K. Assessment principles and tools. Middle East African Journal of Ophthalmology. 2014. https://doi.org/10.4103/0974-9233.129746

[18]. Mastel, K., Morris-Knower, J., & Marsalis, S. Locating tests and measurement instruments for assessment. Journal of Extension. 2016; 54(4).

[19]. Wakabayashi, T., Claxton, J., & Smith, E. V. Validation of a Revised Observation-Based Assessment Tool for Children Birth Through Kindergarten: The COR Advantage. Journal of Psychoeducational Assessment. 2019; 37(1), 69–90. https://doi.org/10.1177/0734282917732491

[20]. Ahmad, S., Hussain, A., Batool, A., Sittar, K., & Malik, M. Play and Cognitive Development: Formal Operational Perspective of Piaget's Theory. Journal of Education and Practice. 2016; 7(28), 72–79. https://doi.org/10.1017/CBO9781107415324.004

[21]. Li, S. L., & Demaree, D. Physics learning identity of a successful student: A plot twist. {AIP} Conference Proceedings. 2013; 1513(1), 242–245. https://doi.org/http://dx.doi.org/10.1063/1.4789697

[22]. Saleh, S. Malaysian students' motivation towards Physics learning. Malaysian Students' Motivation towards Physics Learning. 2014; 2(4), 223–232.

[23]. Karam, R. Framing the structural role of mathematics in physics lectures: A case study on electromagnetism. Physical Review Special Topics - Physics Education Research. 2014; 10(1). https://doi.org/10.1103/PhysRevSTPER.10.010119

[24]. Nitko, A. J., & Brookhart, S. M. Educational assessment of students. In Pearson Education Inc.: Vol. 6th ed. Boston:Pearson Education, Inc. 2011.https://doi.org/10.1016/j.humov.2005.02.001

[25]. Wragg, E. C. Assessment and Learning in the Secondary School. London: Routledge. 2001.https://doi.org/10.4324/9780203164211

[26]. Miller, M. D., Linn, R. L., & Gronlund, N. E. Measurement and assessment in teaching (8th edition). In Pearson Education Ltd.: Vol. 8th ed. Boston: Pearson Education Ltd. 2009.https://doi.org/10.1007/bf03025364

[27]. Borg, W. R., Gall, M. D., & Gall, J. P. Educational research: an introduction (8th ed.). Boston: Pearson Education, Inc. 2007.

[28]. Thiagarajan, S., Semmel, D. S., & Semmel, M. I. Instructional development for training teachers of exceptional children: A sourcebook. Center for Innovation in Teaching the Handicap. 1974; 194. https://doi.org/10.1016/0022-4405(76)90066-2

[29]. Aiken, L. R.Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement. 1985.https://doi.org/10.1177/0013164485451012

[30]. Adam, R. J., & Khoo, S. T. QUEST: The interactive test analysis system. In Australian Council for Educational Research. Camberwell, Victoria: The Australian Council of Educational. 1996.

[31]. Paidi, Djukri, Yulaikah, S., & Alfindasari, D. Development of Instrument to Assess Cognitive Process and Product in Biology Senior High School. International Journal of Environmental & Science Education. 2017; 12(8), 1719–1735. http://staffnew.uny.ac.id/upload/132048519/penelitian/Development of Instrument to Assess Cognitive Process.pdf