



Received: 07.10.2018
Received in revised form: 10.12.2018
Accepted: 19.12.2018

Lutasari, S., & Kartowagiran, B. (2019). Developing instruments for student performance assessment in physics practicum: A case study of state senior high school of Magelang. *International Online Journal of Education and Teaching (IOJET)*, 6(1). 104-114. <http://www.iojet.org/index.php/IOJET/article/view/542>

DEVELOPING INSTRUMENTS FOR STUDENT PERFORMANCE ASSESSMENT IN PHYSICS PRACTICUM: A CASE STUDY OF STATE SENIOR HIGH SCHOOL OF MAGELANG

Research Article

Silvia Lutasari 

Universitas Negeri Yogyakarta

silvia.luta11@gmail.com

Badrun Kartowagiran 

Universitas Negeri Yogyakarta

badrunkw@yahoo.com

Silvia Lutasari graduated her educational degree of Education physics at the State University of Yogyakarta. Currently, he is continuing his master's degree in research and evaluation of the education of state universities in Yogyakarta.

Prof. Dr. Badrun Kartowagiran, M.Pd, majoring in mechanical engineering. He currently serves as chairman of the research and evaluation study program for the postgraduate education at the state university of Yogyakarta.

Copyright by Informascope. Material published and so copyrighted may not be published elsewhere without the written permission of IOJET.

DEVELOPING INSTRUMENTS FOR STUDENT PERFORMANCE ASSESSMENT IN PHYSICS PRACTICUM: A CASE STUDY OF STATE SENIOR HIGH SCHOOL OF MAGELANG

Silvia Lutasari
silvia.luta11@gmail.com

Badrun Kartowagiran
badrunkw@yahoo.com

Abstract

Assessment of practicum activities has not been performed maximally in which it is limited to unstructured observations without applying assessment instruments and it only covers some skills oriented aspects. For that reason, this research aimed to develop instruments for student performance assessment in physics practicum in State Senior High School of Magelang. This research development study was based on the Borg & Gall model. There were three basic steps followed; Phase 1 (problematization), Phase 2 (product creation), and Phase 3 (product testing) that consists of a limited experiment and a large group experiment. In the study, Aiken formula was used for the validation purposes and the results showed that the instrument had a high degree of content validity with the Alpha coefficient value of 0.948. Based on the Total Variance Explained table, the component column showed that there were six factors representing the variables. Accordingly, the instrument of this research was found feasible to be used for high school student performance assessment in physics practicum.

Keywords: Instrument, performance assessment, physics practicum

1. Introduction

Learning is inextricable with assessment since both aspects are important efforts for education management. Efforts to improve education quality can be done by increasing the quality of learning and assessment system. The quality of learning can be examined from the assessment results. Furthermore, a good assessment system will encourage educators to structure good learning and teaching strategies. At the same time, improving education quality requires assessment system improvement.

Regulation of the Minister of Education and Culture Number 54 Year 2013 on Graduates Competency Standards (SKL) (Nomor, 2013) explains the expected aspects of high school graduates' competency, that is, students should have a balance between soft skills and hard skills covering aspects of attitude, skills, and knowledge. In an attempt to achieve these objectives, the curriculum requires the learning activities in each level of education - especially high school- to implement a scientific approach. This is to support the students' competencies such as attitudes, skills, and knowledge. The scientific approach applied in the learning activities involves observing, asking, reasoning, willing to try, and building networks in all subjects, including physics.

According to Abu Hamid (2011), the nature of physics as a part of natural science is in the realm of process and product. Process and product have an equivalent importance level in physics education, both in learning and assessing the results of the learning activities. Having said that, examination and assessment need to be executed in both process and product. The process of learning physics is often related to the skills in performing the tasks of observation, measurement, experiment or practicum, data analysis, etc. Assessing the learning activities require an appropriate type of assessment, that is, a performance assessment which can examine students' skills.

Assessment is a systematic activity for collecting, analyzing, and presenting information in an attempt to accurately interpret students' learning success (Fitzpatrick, Sanders, & Worthen, 2004; Kartowagiran, 2014). Arikunto (2004) suggests that assessment is an act of giving value in educational or school activities. Teachers and other teaching staffs conduct assessments in order to see whether their efforts have reached the goal. Whereas, Angelo (1991) argues that class assessment is a simple method to collect feedback at the beginning and after learning process and to observe how well students have absorbed the learning materials. Examining the aforementioned explanations, it can be said that assessment is a systematic process to determine the values of objectives, activities, decisions, performance, processes, people, objects, and others. A good assessment tool measures the success of the educational process in a precise and accurate way.

Assessment activities in the process of learning physics have not applied standard guidelines. The assessment is solely based on estimations rather than evaluations, and it tends to be subjective. A subjective assessment creates a particular difficulty for teachers to set up an appropriate follow-up action. To overcome this, an instrument with precise and clear criteria is needed to anticipate subjectivity in the assessment. It can be said that using a valid instrument may lead the results of the assessment reliable, at the same time, inform the actual conditions of the students' ability.

Performance assessment is an appropriate way to assess skills-related aspects (Hibbard, 1996; Nitko, 1996). Performance assessment is a distinctive assessment aiming to obtain data about students' ability in carrying out their tasks for each learning topics. To achieve these goals, the 2013 Curriculum requires the learning activities in each level of education, especially high school, to implement a scientific approach that supports students' competencies covering attitudes, skills, and knowledge. Referring to the 2013 Curriculum, a performance assessment instrument with a scientific approach is an urgent call. This is to facilitate teachers in measuring the learning activities and outcomes of the learners.

Performance assessment is an appropriate way to assess skills (Marzano, Pickering, & McTighe, 1993; Van Der Vleuten & Schuwirth, 2005; Wass, Van der Vleuten, Shatzer, & Jones, 2001). Performance assessment is not only measuring the learning outcomes but also providing clearer information about the learning activities. The assessment is based on the performance in completing a given task or a case problem such as presenting knowledge, using reasoning, demonstrating skill or product, and attitude/affection (Mehrens, 1992). Learners are provided with a task to show their ability in completing it.

According to Badrun Kartowagiran (2009), a research instrument is a tool used to collect research data in both qualitative and quantitative data. Qualitative data can be images, words, and/or other objects that are non-numerical. Whereas, quantitative data is a numerical-related data. In qualitative research, the main instrument is the researcher. Having said that, what is meant by research instrument in this study is the quantitative research instrument.

A good instrument needs to be valid and reliable. The instrument validity is about how far the instrument can measure what it should be measured. The instrument validity is seen from a specific purpose, that is, the validity of an instrument to measure attribute 'A' does not necessarily applies to attribute 'B'. Following this, an instrument must be reliable in term of the consistency of its measurement. For example, the test score or other valuation results remain unchanged from one measurement to another.

A closer look at the factual conditions in the field, the assessment of practicum activities has not been performed optimally. In the preliminary stage, the results of interviewing one of the teachers at the State Senior High School 1 Magelang and State Senior High School 2 Magelang show that their laboratory facilities are complete enough to support practicum activities. However, the assessment of practicum activities is only limited to unstructured observations without applying assessment instruments. It only covers a few skill aspect. Moreover, some teachers only use test scores to measure students' ability without providing a fair and open disclosure of the grading procedure. The grade is heavily based on the teacher's judgment without a valid assessment guideline. Studying this, it is evident that the lack of facilities lies on the absence of a valid instrument in assessing the students' performance in learning physics practicum, which later aims to develop their potentials in the long run.

Studying the aforementioned reasonings, the researchers are interested in researching "Developing Instruments for Student Performance Assessment in Physics Practicum: A Case Study of Magelang Senior High School". Specifically, the research questions of this study are (1) How are the structures of the performance assessment instruments used for assessing freshmen students in physics practicum at the State Senior High School of Magelang? (2) How are the characteristics of the performance assessment instruments used for assessing freshmen students in physics practicum at the State Senior High School of Magelang? and (3) How are the students' responses to the performance assessment instruments used for assessing freshmen students in physics practicum at the State Senior High School of Magelang?

2. Methodology

This study is a research development using Borg and Gall model. There were three basic steps that should be carried out by the researchers. These were: (1) Phase 1 (problematization) consisted of compiling instrument specification; (2) Phase 2 (product creation) consisted of creating product/instrument followed by supervisor consultation and assessment instrument review; and (3) Phase 3 (product testing) consisted of a limited experiment and a large group experiment in an attempt to check the readability, practicality, usage, response of the students and teachers as well as the interpretation of the measurement results.

The subject of this research was the performance assessment instrument for high school students. Whilst the object of this research was the students of the State Senior High School of Magelang. The obtained data from the students and teachers are presented in the following table.

Table 1. Total of the experiment subject

Assessment Stage	Total		
	Total of High Schools Involved	Assessor (teachers)	Assessor (students)
Limited	1	3	90
Extended	3	3	500

Data collection was conducted through validation, questionnaire, observation, and documentation. This study applied data analysis techniques of product feasibility analysis and questionnaire analysis about the teacher's response; which included questionnaires of readability, practicality, and student's response. The results of data collection were analyzed by using the qualitative method, while the instrument development data was analyzed quantitatively by finding its validity and reliability. Based on the instrument used, the data analysis applied factor analysis. It began with Exploratory Factor Analysis with the KMO criteria of ≥ 0.5 , and the loading factor was in accordance with the Exploratory Factor Analysis (EFA) criteria.

3. Results & Discussion

3.1 Preliminary Results

During the preliminary stage, the researchers interviewed physics teachers at 3 different State Senior High School of Magelang. All of the three physics teachers provided similar answers. Accordingly, it can be concluded that:

The school system still applies a simple assessment in which the teachers tend to grade practicum activities based on reports only.

The practicum facilities in all three schools are good enough. However, the facilities are rarely used, only when students have a practicum class.

Assessment of the practicum activities is inadequate. An up-to-date practicum assessment is an urgent call.

Based on the results of interviewing the teachers and analyzing the assessment instruments, the implementation of physics practicum is not supported by the implementation of student performance assessment in an effective way. The limitation of teachers in observing a large number of students has led the teachers to rely on their memory to determine the student's performance. In other words, the assessment is not accurate because it is not conducted when the students directly show their performance. The results of the interview and observation of the assessment instruments indicate that it is necessary to develop student performance assessment instruments, specifically on the course of physics practicum.

3.2 Stage 1

In the early stages of creating the instruments, the researchers first formulate a draft to be further consulted with the supervisor. After making the draft, the researchers then develop an assessment rubric by setting indicators and scoring system. Subsequently, the researchers create an assessment sheet for physics practicum.

3.3 Stage 2

At the early stage, researchers have finished formulating the performance assessment instrument by making drafts and assessment sheets which have been approved by the supervisor. The subsequent process is Phase II in which the researchers validate the instrument together with 3 experts as approved by the supervisor. This is to see to what extent the instrument content represents the conceptual frameworks. To do this, Aiken formula is applied. The researchers appoint 3 physicists. Referring to the assessment data from the 3 experts, the researchers then perform data analysis by using Aiken formula (Aiken, 1980).

$$V = \frac{\sum s}{[n(c-1)]} \quad (1)$$

$$V = \frac{54}{[7(3(4-1))]} = \frac{54}{63} = 0.857 \quad (2)$$

Analysis results of the sub-indicator accuracy score on indicators

$$V = \frac{\sum s}{[n(c-1)]} \quad (3)$$

$$V = \frac{233}{[28(3(4-1))]} = \frac{233}{252} = 0.924 \quad (4)$$

A closer look at the analysis results of the experts using Aiken formula, it can be concluded that the analysis result coefficient of the indicator accuracy score on variables and dimensions is 0.857. Whilst, the analysis result coefficient of the sub-indicators accuracy score on indicators is 0.924. Therefore, the formulated instrument is quite good with an adequate content validity. However, the experts suggest to do some revisions, especially the use of language in the instrument.

3.4 Stage 3

3.4.1 Experiment of the Preliminary Product

The experiment of the preliminary product has been validated by the experts under a supervision of the supervisor. Subsequently, the researchers perform an initial test to 90 freshmen students with a representative of one class of each high school. Also, an instrument assessment was done by 3 teachers in an effort to provide suggestions about the upcoming instrument.

In the initial product analysis phase, the assessment analysis of the 3 teachers is conducted by using Ebel formula (Ebel, 1951). This is also applied to instrument validity with factor analysis and instrument reliability, using Cronbach Alpha (Cronbach, 1951) with SPSS.

The following is the result of the instrument analysis and the average of reliability coefficient of the three raters using Ebel formula (Ebel, 1951):

$$\rho_{xx} = \frac{(0,524-0,143)}{0,524} = 0.727 \quad (5)$$

The calculation results indicate that the results of reliability analysis, assessed by the 3 teachers using Ebel formula, is 0.727. These results show that the reliability of the inter-rater is considered to be in a high category.

Instrument validity uses factor analysis. Results of the KMO analysis more than 0.5 is 0.780, therefore, the instrument can be further processed with factor analysis. The anti-image correlation explains that all variables from the first to the seventh variable is more than 0.5 (> 0.5), that are, 0.741, 0.829, 0.725, 0.746, 0.793, 0.878, 0.710. For that, none of the items is aborted.

The instrument reliability using the Cronbach Alpha is 0.849 > 0.7. Thus, it can be said that the variable is reliable. Based on the analysis of the preliminary product experiment and consultations with the supervisors, the initial instrument which the researchers formulate is considered to be reliable. However, the experts and supervisors suggest to add some aspects to make a better improvement. After finalizing the analysis under the supervision of the supervisor about the preliminary product as well as considering experts' suggestion, the next step is product revision.

3.4.2 Experiment of the extended products

After the revision, the researchers conducted a large group experiment with 500 freshmen students from 3 different high schools. The following is an explanation of the analysis results of the assessment instrument. The instrument validity with factor analysis and the instrument reliability applies Cronbach Alpha by using SPSS. The following is a chart of the instrument analysis results.

3.4.2.1 Instrument validity

Table 2. KMO and Bartlett Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.747
Bartlett's Test of Sphericity	Approx. Chi-Square	3704.498
	Df	136
	Sig.	0.000

Table 2 indicates that the results of the KMO more than 0.5 (> 0.5) is 0.747. Therefore, the instrument can proceed with factor analysis.

Table 3. Anti-Image correlation

		Anti-image Matrices																
		VAR 01	VAR 02	VAR 03	VAR 04	VAR 05	VAR 06	VAR 07	VAR 08	VAR 09	VAR 10	VAR 11	VAR 12	VAR 13	VAR 14	VAR 15	VAR 16	VAR 17
Anti-image Covariance	VAR01	.416	-.189	-.113	-.031	.002	-.071	.028	.048	.009	-.036	.083	.013	.001	.002	-.056	.039	-.040
	VAR02	-.189	.390	-.138	-.056	.023	-.035	-.005	-.006	.018	-.007	-.054	-.057	.006	-.015	.031	-.061	-.007
	VAR03	-.113	-.138	.507	.017	.010	.005	-.072	-.049	-.030	-.005	-.012	-.002	-.011	-.027	.040	.036	-.056
	VAR04	-.031	-.056	.017	.355	-.282	.002	.010	-.033	-.017	.007	.014	.010	-.039	.022	.013	.004	.009
	VAR05	.002	.023	.010	-.282	.386	-.014	.005	.004	-.009	.012	-.003	-.001	.025	-.026	.003	.019	-.015
	VAR06	-.071	-.035	.005	.002	-.014	.437	-.225	-.189	.022	.025	-.016	-.076	-.031	.026	.031	.008	.010
	VAR07	.028	-.005	-.072	.010	.005	-.225	.537	-.100	-.073	-.001	.015	.051	-.003	-.002	-.054	.050	.029
	VAR08	.048	-.006	-.049	-.033	.004	-.189	-.100	.616	.014	.018	-.025	.078	-.040	.014	-.017	-.013	-.019
	VAR09	.009	.018	-.030	-.017	-.009	.022	-.073	.014	.592	-.134	-.071	-.017	.007	.019	.040	-.133	-.083
	VAR10	-.036	-.007	-.005	.007	.012	.025	-.001	.018	-.134	.381	-.219	.024	-.030	-.030	.039	.073	
	VAR11	.083	-.054	-.012	.014	-.003	-.016	.015	-.025	-.071	-.219	.377	-.041	.026	-.023	-.021	.092	-.081
	VAR12	.013	-.057	-.002	.010	-.001	-.076	.051	.078	-.017	.024	-.041	.724	-.232	-.011	.000	.040	.009
	VAR13	.001	.006	-.011	-.039	.025	-.031	-.003	-.040	.007	-.030	.026	-.232	.532	-.289	-.047	.015	-.006
	VAR14	.002	-.015	-.027	.022	-.026	.026	-.002	.014	.019	-.012	-.023	-.011	-.289	.673	.005	-.052	.020
	VAR15	-.056	.031	.040	.013	.003	.031	-.054	-.017	.040	-.081	-.021	.000	-.047	.005	.330	-.180	-.198
	VAR16	.039	-.061	.036	.004	.019	.008	.050	-.013	-.133	.039	.092	.040	.015	-.052	-.180	.621	-.033
	VAR17	-.040	-.007	-.056	.009	-.015	.010	.029	-.019	-.083	.073	-.081	.009	-.006	.020	-.198	-.033	.385

Studying the explanation of anti-correlation images, it is evident that none of the items is aborted. The table shows all variables, from the first to the seventh variable, are more than 0.5 ($> 0,5$). These are: (1) the first variable is 0.788; (2) the second variable is 0.813; (3) the third variable is 0.872; (4) the fourth variable is 0.588; (5) the fifth variable is 0.553; (6) the sixth variable is 0.761; (7) the seventh variable is 0.768; (8) the eighth variable is 0.813; (9) the ninth variable is 0.842; (10) the tenth variable is 0.729; (11) the eleventh variable is 0.733; (12) the twelfth variable is 0.728; (13) the thirteenth variable is 0.713; (14) the fourteenth variable is 0.705; (15) the fifteenth variable is 0.735; (16) the sixteenth variable is 0.673; and (17) the seventeenth variable is 0.781. Therefore, it can be concluded that none of the items is aborted

Table 4. *Commonalities*

	Initial	Extraction
VAR01	1.000	.599
VAR02	1.000	.802
VAR03	1.000	.814
VAR04	1.000	.797
VAR05	1.000	.792
VAR06	1.000	.707
VAR07	1.000	.767
VAR08	1.000	.720
VAR09	1.000	.694
VAR10	1.000	.791
VAR11	1.000	.749
VAR12	1.000	.717
VAR13	1.000	.548
VAR14	1.000	.772
VAR15	1.000	.628
VAR16	1.000	.889
VAR17	1.000	.890

The value of commonalities indicates to what extent a variable explains a factor. The result shows that the highest variable value is in variable 17 with a value of 0.890; meaning that variable 17 can explain a factor of 89,0%. While the lowest variable value is in variable 1 with a value of 0.548; meaning that variable 1 can explain a factor of 54,8%. The result of the aforementioned 17 variables shows a value of more than 50% ($> 50\%$). To sum up, all variables can explain the factors.

Table 5. Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,489	26,405	26,405	4,489	26,405	26,405	2,419	14,227	14,227
2	2,648	15,578	41,983	2,648	15,578	41,983	2,337	13,750	27,977
3	1,582	9,307	51,290	1,582	9,307	51,290	2,157	12,690	40,667
4	1,545	9,089	60,379	1,545	9,089	60,379	2,036	11,979	52,646
5	1,264	7,437	67,816	1,264	7,437	67,816	1,921	11,297	63,944
6	1,147	6,745	74,561	1,147	6,745	74,561	1,805	10,617	74,561
7	.723	4,255	78,816						
8	.613	3,605	82,421						
9	.507	2,984	85,405						
10	.475	2,794	88,200						
11	.406	2,386	90,585						
12	.379	2,227	92,812						
13	.314	1,848	94,661						
14	.298	1,754	96,415						
15	.215	1,264	97,678						
16	.204	1,199	98,878						
17	.191	1,122	100,000						

The Total Variance Explained Table is the obtained results that determine the number of possible factors to be established. Based on the Total Variance Explained Table, it can be seen that the component column shows six factors representing the variables. The next step is specifying the matrix component, followed by a rotation of the matrix component. The results are presented in the following table.

Table 6. Rotated component matrix

	Component					
	1	2	3	4	5	6
VAR01	.708	.060	.080	.292	.008	.056
VAR02	.871	.115	-.004	.092	.125	-.075
VAR03	.882	.096	.063	.070	.118	-.063
VAR04	.023	.849	.129	.176	.089	.138
VAR05	.134	.832	.179	.080	.152	.143
VAR06	.164	.776	.256	.032	.104	-.014
VAR07	-.029	.302	.803	-.078	.131	.085
VAR08	.110	.171	.823	.009	.026	-.019
VAR09	.042	.073	.816	.096	.032	.106
VAR10	.387	.136	.063	.775	.122	-.066
VAR11	-.025	-.008	-.040	.863	.027	-.036
VAR12	.384	.272	.062	.699	.054	.003
VAR13	.095	.189	.007	-.111	.700	-.005
VAR14	.084	.089	.156	.135	.843	.052
VAR15	.056	.022	.016	.130	.779	.021
VAR16	-.049	.174	.091	-.038	.048	.919
VAR17	-.032	.048	.058	-.041	.015	.939

Based on the above table, the 17 variables are grouped into 6 factors which described in the following:

Factor 1: variable 1 (students are capable to choose the practicum instruments), variable 2 (students are capable to distinguish the required procedure in accordance with the practicum guideline), variable 3 (students are capable to identify the required instruments in accordance with the practicum guideline).

Factor 2: variable 4 (students get prepared for the practicum by reading the practicum guideline), variable 5 (students are capable to understand every step of the procedure), variable 6 (students are enthusiastic in carrying out the practicum activities).

Factor 3: variable 7 (students are willing to engage in practicum activities without depending on other group's assistance), variable 8 (students are capable to follow the instruction in accordance with the practicum guideline): variable 9 (students are capable to quickly respond every instruction in accordance with the practicum guideline).

Factor 4: variable 10 (students are capable to assemble the practicum instruments in accordance with the practicum guideline), variable 11 (students are capable to do observation during the practicum), variable 12 (students are capable to do measurement activities during the practicum).

Factor 5: variable 13 (students are capable to record the practicum data), variable 14 (students are capable to participate in a group setting), variable 15 (students are capable to observe their friend's skill during the practicum).

Factor 6: variable 16 (student are capable to analyze data), variable 17 (students are capable to draw a conclusion).

3.4.2.2 Instrument reliability

The reliability value of Cronbach's Alpha is shown in the following table:

Table 7. *Instrument reliability*

Cronbach's Alpha	N of Items
0.948	17

Table 7 indicates the value of the Alpha coefficient of $0.948 > 0.6$. It can be concluded that the variable is very reliable. The value of Cronbach's Alpha reliability is presented in the following table:

Table 8. *Reliability of the Cronbach's Alpha*

Value of Cronbach's Alpha	Reliability Level
0.0 - 0.20	Less Reliable
>0.20 - 0.40	Slightly Reliable
>0.40 - 0.60	Quite Reliable
>0.60 - 0.80	Reliable
>0.80 - 1.00	Very Reliable

Source: *Hair et al. (Hair, 2010)*.

What distinguishes this study from other previous studies lies in the application of the scientific approach in accordance with the 2013 curriculum and the Education and Culture Ministerial Decree No.104 year 2014 (Penyusun, 2014). Thus, this study may be useful in all high school physics courses due to its suitability for the student performance assessment.

4. Conclusion

The formulation of the assessment instruments is carried out through several stages: preliminary studies, limited experiments, and product testing. The structures of the performance assessment instrument with a performance assessment indicator include several aspects namely perception, preparation, action response, complex mechanism, communication, and creativity. A closer look at the first experimental test about student's performance in the practicum activities, the result is considered to be good enough yet it still needs a little improvement. The revisions are based on the input of the experts, teachers, and supervisors; this contributes to making the assessment instrument valid and reliable. The instrument can be used to assess the performance of high school students.

This study suggests the future research to develop instruments for other aspects in the course of physics. This instrument should be further applicable in other schools to see whether teachers have already met the assessment standards or they still use an outdated assessment.

References

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, 40(4), 955-959.
- Angelo, T. A. (1991). *Ten easy pieces: Assessing higher learning in four dimensions*. New Directions for Teaching and Learning, 1991(46), 17-31.
- Arikunto, S., & Jabar, C. S. A. (2004). *Evaluasi program pendidikan*. Jakarta: Bumi Aksara, 1-2.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, 16(4), 407-424.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). Program evaluation: Alternative approaches and practical guidelines.
- Hair, J. (2010). *Multivariate data analysis, a global perspective*. New Jersey. Pearson. Ed, 7, 816.
- Hamid, A. A. (2011). *Pembelajaran fisika di sekolah*. Yogyakarta: Universitas Negeri Yogyakarta.
- Hibbard, K. M. (1996). Performance-Based Learning and Assessment. A Teacher's Guide: ERIC.
- Kartowagiran, B. (2009). *Penyusunan Instrumen Kinerja SMK-SBI*. Paper presented at the Makalah dalam Workshop Evaluasi Kinerja SMK-SBI P4TK Matematika.
- Kartowagiran, B. (2014). Penilaian Berbasis Kurikulum 2013.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model: ERIC.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3-9.
- Nitko, A. J. (1996). Educational assessment of students: ERIC.
- Nomor, P. (2013). Tahun 2013 tentang Standar Kompetensi Lulusan. Jakarta: Departemen Pendidikan Nasional.
- Penyusun, T. (2014). Lampiran Permendikbud No. 104 Tahun 2014 tentang Penilaian Hasil Belajar oleh Pendidik. Jakarta: Kemdikbud.
- Van Der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Medical education*, 39(3), 309-317.
- Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The lancet*, 357(9260), 945-949.