LOOK
ACADEMIC PUBLISHERS
OPEN ACCESS

INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
2017, VOL.12 , NO.5, 1269-1281

The Development of Performance Assessment of Stem-Based Critical Thinking Skill in the High School Physics Lessons

Fikroturrofiah Suwandi Putri[a] and Edi Istiyono[a]

[a]Physics Education Study Program, Graduate School, Yogyakarta State University, Yogyakarta, Indonesia

ABSTRACT

This research aims to produce an assessment instrument of STEM-based critical thinking skill which meets the feasibility criteria. This development research refers to the model developed by Borg & Gall and is modified using the development model instruments developed by Oriondo & Antonio. The research subjects were one senior high school Physics teacher, 129 tenth grade students, and 331 eleventh grade students of senior high school. The data gathering was carried out using self-evaluation sheets, observation rubric, students' worksheets (LKPD) and reportscoring rubrics, and a test instrument on critical thinking skill. The research results show that the developed performance assessment has fulfilled the content validity based on the evaluation by 3 experts and 3 practitioners. The reliability of all the rubrics in the performance assessment is categorized as very high. The Test on critical thinking skill consisting of 72 items was declared fit usingPCM and the level of difficulty of the items ranged between -0.69 and 1.14, which implies good category. The test also had a reliability coefficient of 0.81 and was categorized as very high and suitable to measure the students whose ability ranged from -1.60 to 1.70 in the logit scale. In addition, teachers and students gave positive responses to the application of the developed assessment. Therefore, the developed performance assessment of STEM-based critical thinking skill has fulfilled the feasibility criteria to be applied in the senior high school physics class.

KEYWORDS
Performance Assessment, critical thinking skill, STEM, physics subject

ARTICLE HISTORY
Received 21 August 2016
Revised 10 March 2017
Accepted 22 March 2017

Introduction

The development of education in Indonesia conforms with the international educational framework. Partnership for 21st Century Skills(2013) as one of the educational frameworks mentions that the 21st Century student outcomesincluded content knowledge, learning and innovation skills, information, media, and technology skills, life and career skills. Life skills

CORRESPONDENCE Fikroturrofiah Suwandi Putri fikrotur.rofiah@gmail.com
Edi Istiyono edi_istiyono@uny.ac.id

consist of hard skill and soft skill. Therefore, Permendikbud No. 81 A Tahun 2013 states that education as implemented in the school curriculum requires a balance between hard skill and soft skill.

The research conducted by Widarto, et al. (2012, p.411) shows that senior high schools focus more on the knowledge aspect and technical skill (hard skill) whereas the biggest contributing aspect in the work environments is self-management skill and interpersonal skill (soft skill). Thus, it is urgent to develop soft skills in education.

One of the important 21^{st} century soft skills according to Wagner (2008) is critical thinking skill. In line with this opinion, Permendikbud No.64 Tahun 2013 on the content standards states that one of the competences to be developed in the implementation of the 2013 Curriculum is the competence to think critically. Based on the regulation, it can be observed that education in Indonesia tries to develop critical thinking skill to face the challenges of the 21^{st} Century.

Ennis (1996, p.50) defines critical thinking skill as the ability to think reflectively focusing on the pattern of presenting decision on what is believed and what is to be done. Research conducted by Qing, et al. (2010) develops the critical thinking skill through experiment activities. In addition, research conducted by Ku (2009); and Blattner & Frazier (2012) found that critical thinking can be assessed through performance assessment. Another previous research conducted by Sari & Sugiarto (2015) shows that critical thinking skill can be developed by designing activities to involve learners to solve problems. Hence, critical thinking skill can be fosteredthrough a performance assessment designed based on critical thinking skill indicators to assess the problem solving activities by means of experiment methods.

Nitko & Brookhart (2011, p.246) assert that, "performance assessment requires student to create a product or demonstrate a process, or both, and uses clearly defined criteria to evaluate the qualities of student". Based on the explanation, performance assessment deals with the learning process experienced by the learners and the developed product at the end of the learning process. Therefore, the performance assessment can be used to assess the classroom learning as a whole.

The performance assessment adopts the classroom assessment model which has three objectives, namely assessment for learning, assessment of learning, and assessment as learning (Arends, 2012, p.230). Assessment for learning is used to improve the learning outcomes aligned with the assessment objectives. Assessment of learning is used to monitor the knowledge the learners have accumulated by considering the learner's self evaluation. Assessment as learning is used to evaluate the attainment of the learning objectives carried out from the beginning to the end of the learning. This is the base where a holistic assessment can improve the learner's critical thinking skill if the learning outcomes are adjusted to the indicators of critical thinking skills.

Developing a performance assessment takes a specific approach which can answer the challenging advancement of science, information, and technology which refers to the development of critical thinking skill. STEM is one approach which integrates science, technology, engineering, and mathematics in the learning process. The integration in the learning process can encourage the learners to develop their critical thinking skills. The research conducted by

Becker & Park (2011) showed that there was a significant difference in the learning outcomes between classes which applied STEM and classes which did not. Further research conducted by Petrie, et al. (2014:p.1) found that STEM-based learning exercised the learners' thinking skills. Thus, the STEM-based approach in the performance assessment is assumed to have been able to develop learners' critical thinking skill.

Based on the explanation, physics learning needs an operational assessment to measure learners' critical thinking skills in the classroom learning process. The indicators of the critical thinking skills are arranged in a systematic way to construct learners' knowledge and to exercise their critical thinking during the learning process. In addition, the learners integrate science, technology, engineering and mathematics in the physics problem solving process. Therefore, it is urgent to develop a performance assessment of the STEM-based critical thinking skill in the physics learning.

Research Methodology

Types of Research

This is a research and development according to the R&D model developed by Borg& Gall (1983) and is modified using the instrument development method developed by Oriondo & Antonio (1984). The development process is presented in Figure 1.

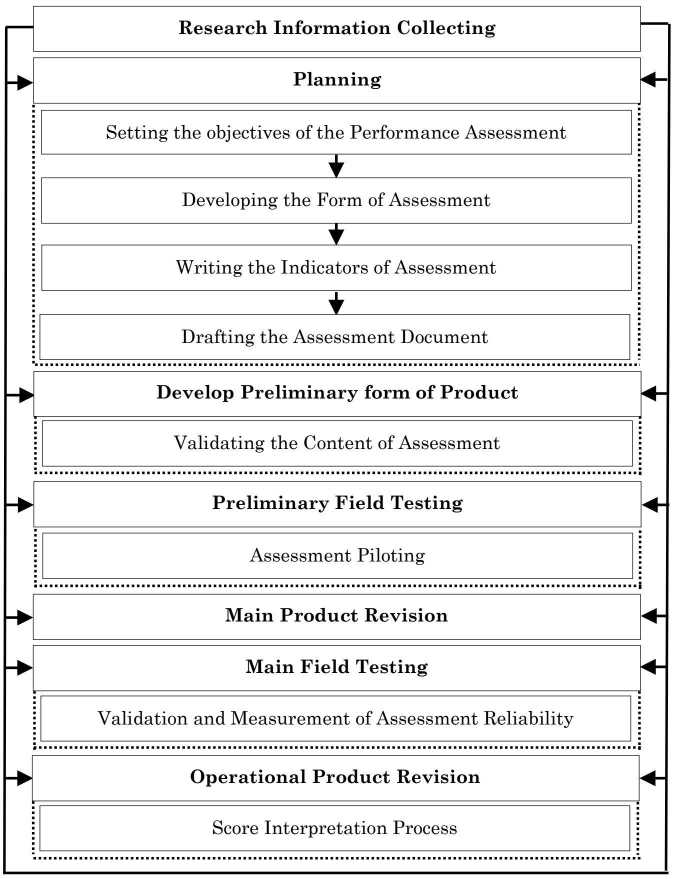

Figure 1. Flowchart of the Performance Assessment Development

Research Subject

The subjects of the preliminary field testing consisted of one Physics teacher and 32 tenth grade students. The subjects of the main field testing for assessment sheet were 97 tenth grade students whereas the subjects of the critical thinking skill test were 331 eleventh grade students. The number of research subjects in the empirical validity testing of the test instruments was more than 200 students. This is corroborated by Seon (2009, p.3) who states that the number of samples to analyze based on Item Response Theory was around 200 to 1000 people. The research subjects were students from several senior high schools in Yogyakarta whose grades were categorized as low, medium, and high based on the 2015 National Examination Results. The selection of samples was made so as to get the results which would show the learners' low, medium, and high degree of critical thinking skills.

Techniques and Instruments of Data Gathering

The data gathering technique used in this research was a questionnaire, observation, test and documentation. The instrument of data gathering included: (1) the evaluation sheet of the validation instrument and evaluation sheet of the product; (2) teacher's response sheet and learners' response sheet; (3) self-evaluation sheet; (4) observation rubric; (5) students' worksheet and reportscoring rubrics; and (6) test instrument of critical thinking skills.

Technique of Data Analysis

The product feasibility was analyzed based on the experts' and practitioners' judgment, namely by counting the theoretical mean of the criteria in each developed assessment aspect (Azwar, 2016, p.147-148) as presented in Table 1.

Table 1. Ordinal Categorization

Theoretical Mean Interval	Category
$\mu \le -1.5\,\sigma$	Very low
$-1.5\,\sigma < \mu \le -0.5\,\sigma$	Low
$-0.5\,\sigma < \mu \le +0.5\,\sigma$	Medium
$+0.5\,\sigma < \mu \le +1.5\,\sigma$	High
$+1.5\,\sigma < \mu$	Very High

with,
μ : theoretical *mean*
σ : standard deviation

The content validity of the test instrument was analyzed using the Aiken's V formula. According to Aiken (1985, p.139), the Aiken's V formula criterion to be fulfilled for 7 raters and 4 numbers of rating was 0.67. If the Aiken's V formula obtained from the content validity of the test instruments of the critical thinking skill was more than 0.67, the instruments were declared valid.

The reliability of the Observation Sheet, Students' Worksheet Scoring Sheet, and the Report Scoring Sheet was analyzed using Intraclass Correlation Coefficient. The intraclass correlation coefficient is related to the alpha reliability coefficient (Gliem & Gliem, 2003). The alpha reliability coefficient could be interpreted according to Table 2.

Table2. Category of the Alpha Coefficient

Alpha	Category
$\alpha \geq 0.9$	Excellent
$\alpha \geq 0.8$	Good
$\alpha \geq 0.7$	Acceptable
$\alpha \geq 0.6$	Questioned
$\alpha \geq 0.5$	Poor
$\alpha \leq 0.5$	Unacceptable

The empirical validity of the test instrument was counted using Partial Credit Model (PCM). PCM is a polytomous scoring model derived from theRasch model in the dichotomous data (Retnawati, 2016, p.49).PCM was used to analyze the test items which have several steps to solve them. The synchronization of the test item and the PCM model was interpreted based on the average means of INFIT Mean of Square (Mean INFIT MNSQ) and the standard deviation (Hambleton & Swaminathan, 1985, p.36). If the average mean of INFIT MNSQ was 1.0 and the standard deviation was 0.0 or the mean of INFIT t approached 0.0 and the standard deviation was 1.0, the entire test items were synchronized with the model. An item or testee/case/person is declared to be suitable to the model in the range of INFIT MNSQ of 0.77 to 1.30. In addition, the item is declared to be good when the index of difficulty was more than -2.0 or less than 2.0.

The reliability of the test instrument was interpreted based on the Cronbach's Alpha coefficient. The degree of the Cronbach's Alpha(α)reliability of the test item was divided into five-scale range (Sumintono & Widhiarso, 2014, p.112) as presented in Table 3.

Table 3. Degree of the Test Reliability

Alpha	Degree of Reliability
$\alpha < 0.5$	Unreliable
$0.5 < \alpha < 0.6$	Less Reliable
$0.6 < \alpha < 0.7$	Fairly Reliable
$0.7 < \alpha < 0.8$	Reliable
$\alpha > 0.8$	Very Reliable

Research Findings

Research Information Collecting Phase

In this phase, the data was gathered through a field study and literature review. The field study conducted in several senior high schools in Yogyakarta shows that teachers need an operational assessment in the Physics learning. In addition, the learning at schools had not integrated the four aspects, namely science, technology, engineering, and mathematics. Based on the Partnershipfor 21st Century Skillsand Permendikbud No. 64 of 2013, it was found that critical thinking skill is a competence that students need to master. Therefore, this phase provides a picture of the assessment design needed by the schools.

Planning Phase

This phase includes setting the objectives of assessment, development of the form of the assessment, drafting the assessment indicators, and writing the assessment document. The objectives of assessment were to construct learners' critical thinking skills through literature review on the aspects, subaspects, and indicators of critical thinking skills integrated with STEM. The forms of assessment were classroom assessment for (1) assessment for learning in the form of students' worksheet; (2) assessment of learning in the form of self-assessment; and (3) assessment as learning in the form of observation sheet, students' worksheet and reportscoring rubrics, and the test instrument for critical thinking skills in the form of two-tier multiple choice.

Further, items referring to the indicators of critical thinking skills had been adjusted to the basic competence, materials, evaluation technique, and assessment strategies. The selected competence was adjusted to the curriculum implemented by the research subject. The lesson materials selected were temperature, types of heat, melting heat, and the Black's principles.

After the indicators were made, the prototype of the performance assessment of the STEM-based critical thinking skill was designed. Before the drafting of the assessment instrument, validation instrument and product assessment evaluation sheets were drafted so that the designed product fulfilled the evaluation criteria and the development principles of performance assessment.

The Developing Preliminary Form of Product Phase

In this phase, there were two findings, namely the results of the validation instrument for product evaluation based on FGD and the results of the product evaluation based on experts and practitioners. The findings from the FGD stated that the evaluation instrument of the product assessment can be used after several revisions. The revised scoring sheets were then used to evaluate the product of performance assessment for STEM-based critical thinking being developed.

The assessors who assessed the assessment instrument were a Physics material expert, measurement and evaluation expert, physics education expert, practitioner or a physics teacher. The assessors evaluated and gave suggestions related to the developed product. The result of the recapitulated product evaluation can be seen in Table 4.

Table 4. The Result of Product Evaluation

Instrument of Performance Assessment	Theoretical *Mean*	Category
Students' Worksheets	90.43	High
Self-Evaluation Sheet	31	Very High
Observation Rubric	47	Very High
Students' Worksheets Scoring Rubric	47.43	Very High

Report Scoring Rubric	47.43	Very High

The content validation of the critical thinking skill test was determined by counting the Aiken's V formula coefficient from the experts' and practitioners' evaluation. Based on the analysis using Aiken's V formula Equation, it was found that each item developed was more than 0.67. Therefore, each item of the critical thinking skill test instrument was declared valid.

Preliminary Field Testing Phase

In this phase, the learners stated their opinion in terms of the language that they did not understand in the Students' Worksheets and the Self-Evaluation Sheet, whereas the teachers gave their opinion on the language that they did not understand in the observation sheet, Students' Worksheet scoring rubric, and the report scoring rubric.

Main Field Testing Phase

Based on the implementation simulation of the performance assessment in three classes, the data was found in the form of evaluation from three raters analyzed using ICC. The analysis result obtains the reliability of the evaluation sheet which is presented in Table 5. The three evaluation sheets are categorized as "very high."

Table 5. Reliability of the Evaluation Sheet

Performance Assessment	***Intraclass Correlation Coefficient* (ICC)**	**Alpha Reliability Coefficient**
Observation Sheet	0.814	0.929
Students' Worksheet Scoring Rubric	0.948	0.982
Report Scoring Rubric	0.971	0.990

During the implementation simulation of the performance assessment, teachers were given response sheets to find out what they thought about the developed assessment. The results of the teachers' and students' responses were analyzed using the descriptive statistics and are presented in Figure 1 and Figure 2.

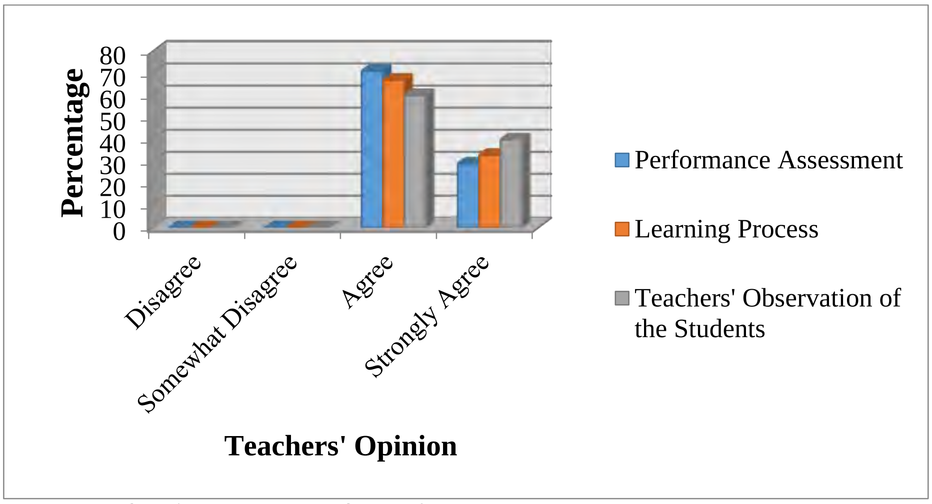

Figure 1. Teachers' Responses to the Performance Assessment

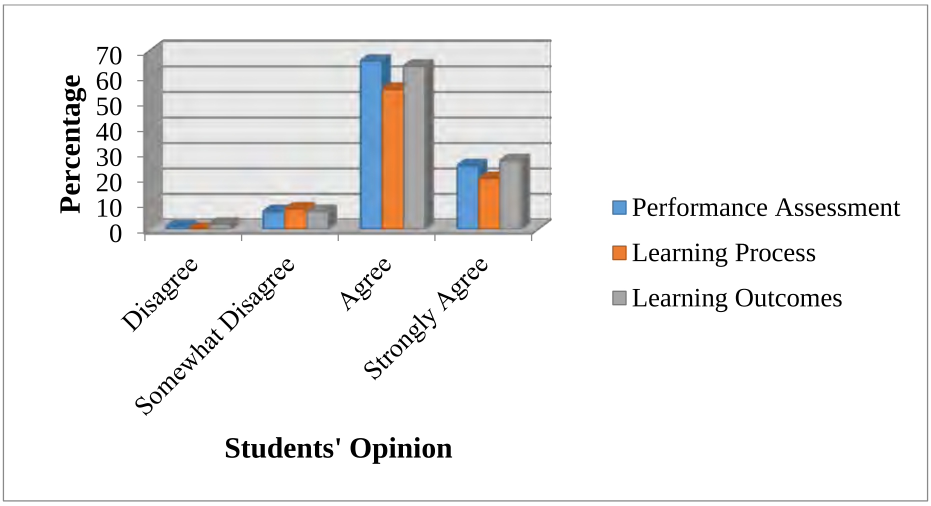

Figure 2. Students' Responses to Performance Assessment

In the main field testing phase, the test instrument on critical thinking skills was tried out and the result can be seen in Table 6.

Table 6. The Result of Test Trial on Critical Thinking Skills

Parameter	Estimated Items	Estimated Testee
INFIT MNSQ	1.01 ± 0.04	1.01 ± 0.07
OUTFIT MNSQ	1.01 ± 0.04	1.01 ± 0.08
Reliability of estimate	0.81	
AverageDifficulty	0.0 ± 0.33	

The level of difficulty of the items lies between -0.60 and 1.24 with the average of 0.0 and the standard deviation of 0.33. The level of difficulty of the item in each sub-aspect can also be seen in Figure 3, which shows the items based on the order of difficulty, namely problem identification, data presentation, discussion, offering solutions, designing procedures, identifying errors, formulating hypothesis, and drawing conclusions.

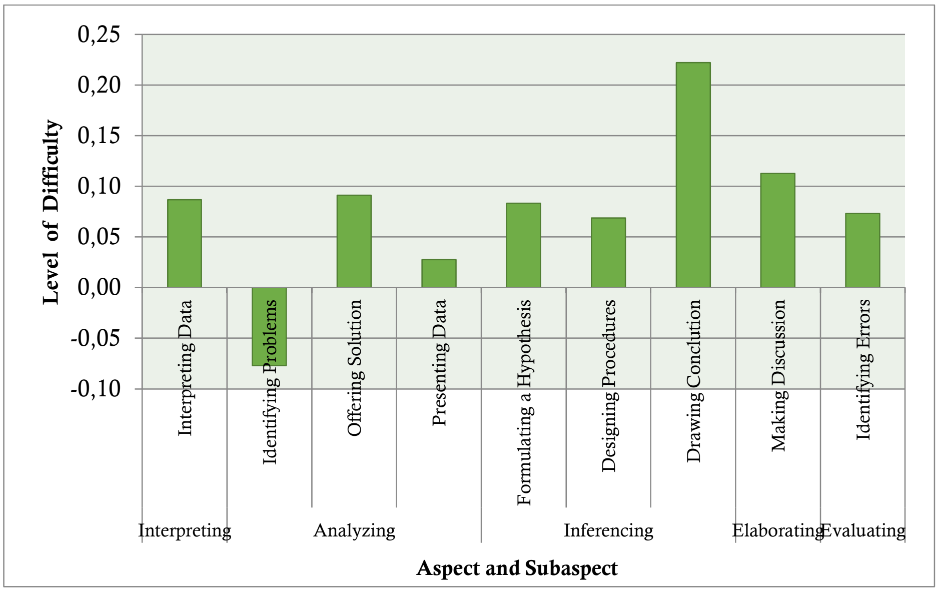

Figure 3. The Level of Difficulty of Each Item per Aspect and Sub-Aspect

In addition, based on the curve of item characteristics in Figure 4, it can be stated that the test instrument of critical thinking skills is suitable to measure the students' ability which ranged from -1.60 to 1.70 in the logit scale.

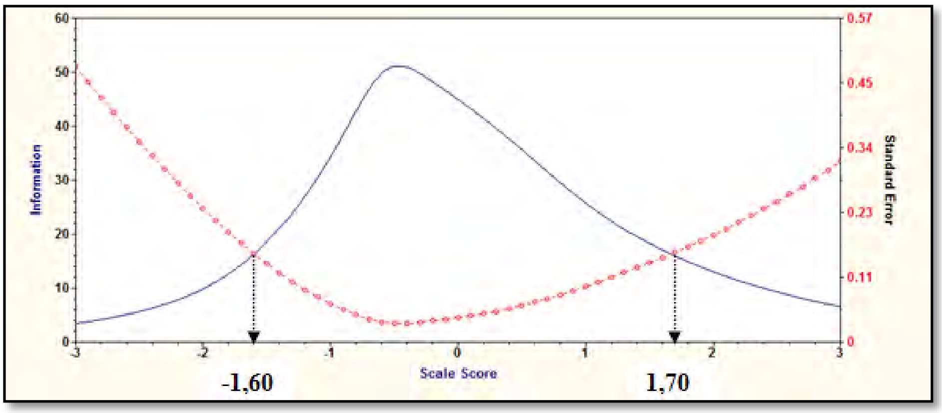

Figure 4. The Function of Information and SEM

Discussion and Conclusion

The evaluation conducted by the experts and practitioners on the developed performance assessment on the STEM-based critical thinking skill is declared suitable to be used. This is evident in the results of the Students' Worksheets, self-evaluation sheet, observation rubric, Students' Worksheet scoring rubric, and Report Scoring rubric which were categorized from "good" to "very good."

Somewhere along the line, all the items of the developed test instrument was classified as “valid” criteria based on the coefficient of the Aiken’s V formula.

The result of the preliminary field testing phase shows that there are some words in the performance assessment which need to be revised. Further, the main field testing phase also yielded the characteristics of the test instrument of critical thinking skills. The instrument was declared fit using PCM model based on INFIT MNSQ in which all the items were within the range from 0.77 to 1.30. The reliability coefficient was 0.81 and classified as “very high.” The level of difficulty of each item was between -0.60 and 1.24 with the average of 0.0 and the standard deviation of 0.33. The level of difficulty was between -2.00 and the instrument item was stated to have a good level of difficulty. The test instrument was suitable to measure students’ ability ranging between -1.60 and 1.70 in the logit scale.

In the main field testing phase, the data obtained from the reliability of the evaluation sheet was in the form of ICC and alpha. The alpha coefficient of the evaluation sheet is 0.929 and classified as “very high”; the alpha coefficient of the Students’ Worksheet scoring rubric is 0.982 and classified as “very high”; and the alpha coefficient of the report scoring rubric is 0.990 and classified as “very high.” In addition, the result of this phase is corroborated by the students’ and teachers’ positive responses to the implementation of the performance assessment.

The results of the research and discussion provide explanation of the developed product. Hence, it can be said that the performance assessment of the STEM-based critical thinking skill has fulfilled the feasibility characteristics implemented in the Senior High School physics subject.

Limitation and Suggestion

The developed performance assessment has some limitations such as inefficient use of papers. Thus, it is suggested that the further product development can be in the form of digital application to be installed in the students’ and teachers’ gadgets or computers. In addition, this product can be developed further in other physics materialsbesides the material on heat by implementing STEM and constructing the critical thinking skills.

Acknowledgment

This article was written thanks to the facility provided by the Graduate School Library, Yogyakarta State University, in the process of finding references. YSU has given support in the form of materials in the research and publication of this article. I am forever grateful for the assistance provided by the staff of Yogyakarta State University.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Fikroturrofiah Suwandi Putri earned her M.Pd.(Master of Education) in Physics Education from the Yogyakarta State University Graduate School, Yogyakarta, Indonesia.

Edi Istiyono is a Doctor in the field of Physics education testing and evaluation. Currently he works as a lecturer at the Faculty of Mathematics and Natural Sciences and the Yogyakarta State University Graduate School, Yogyakarta, Indonesia.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, (45), 131-142.

Arends, R.I. (2012). *Learning to teach* (9th ed.). New York: McGraw-Hill.

Azwar, S. (2016). *Penyusunan skala psikologi (edisi 2)*. Yogyakarta: Pustaka Belajar.

Becker, K. & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education*, 12 (10), 862-875. DOI**:** 10.12691/education-2-10-4.

Blattner, N.H. & Frazier, C.L. (2002). Developing a performance-based assessment of student's critical thinking skills.*Assessing Writing*, 8, 47-64. doi:10.1016/S1075-2935(02)00031-4

Borg, W.R., & Gall, M.D. (1983). *Educational research: An introduction* (4th ed.). New York: Longman.

Ennis, R.H. (1996). *Critical thinking*. New York: Pretince-Hall.

Gliem, J.A. & Gliem, R.A. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales*. Midwest Research to Practice Conference in Adult, Continuing, and Community Education.

Hambleton & Swaminathan (1991). *Fundamentals of item response theory*. Newbury Park, Calif: Sage Publications.

Kementerian Pendidikan dan Kebudayaan.(2013). *Peraturan Kementerian dan Kebudayaan No. 81A, Tahun 2013, tentang Implementasi Kurikulum*.

Kementerian Pendidikan dan Kebudayaan.(2013). *Peraturan Kementerian dan Kebudayaan No. 64, Tahun 2013, tentang Standar Isi*.

Ku, K.Y.L. (2009). Assessing student's critical thinking performance: urging for measurement using multi-response format. *Thinking Skills and Creativity*, *4,* 70-76. doi:10.1016/j.tsc.2009.02.001

Nitko, B. J. & Brookhart, S. M. (2011). *Educational assessment of student* (6th ed.). Boston: Pearson Education, Inc.

Oriondo, L.L. & Antonio, E.M.D. (1984). *Evaluating educational outcomes (test, measurement, and evaluation)*. Manila: Rex Book Store.

Partnership for 21st Century Skills. (2008). *21st Century Skills:How can you prepare students for the new Global Economy*. Paris: Cisco System,Inc. Diambil pada 25 Agustus 2015, dari http://www.oecd.org/site/educeri21st/40756908.pdf

Petrie, K., Akmal, T., & Lamb, R. (2014).Development of a cognition-priming model describing learning in STEM Classroom.*Journal of Research in Science Teaching*, *5,* 1-23. http://onlinelibrary.wiley.com/doi/10.1002/tea.21200/abstract

Qing, Z., Ni, S., & Hong, T. (2010).Developing critical thinking disposition by task-based learning in chemistry experiment teaching.*Social and Behavioral Science*, 2, 4561–4570.doi:10.1016/j.sbspro.2010.03.731

Retnawati, H. (2016). *Validitas, reliabilitas, &karakteristikbutir (Panduanuntukpeneliti, mahasiswa, danpsikometrian)*. Yogyakarta: Parama Publishing.

Sari, D.S. & Sugiyarto, K.H. (2015).Pengembangan Multimedia Berbasis Masalah Untuk Meningkatkan Motivasi Belajar Dan Kemampuan Berpikir Kritis Siswa. *Jurnal Inovasi Pendidikan IPA*. ISSN: 2477-4820 ,1 (2), 153-166. DOI: http://dx.doi.org/10.21831/jipi.v1i2.7501

Seon, Hi Sin. (2009). How to tread omitted respons in Rasch model based equating. Practical Assessment. *Research & Evaluation, 14(1),* 133-145. ISSN 1531-7714.

Sumintono, B. &Widhiarso. (2014). *Aplikasi model raschuntukpenelitianilmu-ilmusosial.*Cimahi: Trim Komunikata*Publishing House*.

Wagner, T. (2008). *The Global achievement gap*. New York: Basic Books.

Widarto, Pardjono, dan Widodo, N. (2012). Pengembangan model pembelajaran*soft skills* dan*hard skills*untuksiswa SMK. *CakrawalaPendidikan*, *3,* 409-423.

Appendix. Blue Print of the Performance Assessments of the STEM-based Critical Thinking Skills

No.	Aspect	SubAspect	Indicators
1.	Interpreting	Interpreting the data from the experiment results using technology and mathematics skills	1. Interpreting data in the form of tables. 2. Interpreting data in the form of graphs. 3. Interpreting data in the form of a proposition of negative-positive relationship between variables. 4. Interpreting data in the form of correlation coefficient or gradient.
2.	Analyzing	Identifying problems using the science skill.	1. Identifying problems related to the issue being presented. 2. Identifying problems related to the lesson materials. 3. Identifying problems and presenting them in a concise and clear affirmative proposition.
		Solving a problem as the basic skill in making experiments using science skills.	1. Offering solutions related to the identified problems. 2. Offering solutions which can be implemented in the experiment. 3. Offering solutions along with the negative consequences. 4. Offering solutions along with the positive consequences.
		Presenting the data of the experiment results using technology skills.	1. Making tables containing independent and dependent variables according to the experiment using the Excel program. 2. Making graphs with the independent variable in the *x* axis and dependent variable in the *y* axis according to the experiment using the Excel program.
3.	Inferencing	Formulating the experiment hypothesis using the science skill.	1. Formulating a hypothesis in the form of a logical proposition. 2. Formulating a hypothesis related to the experiment plan. 3. Formulating a hypothesis supported by a proposition from a relevant source. 4. Formulating a hypothesis containing the correlation between independent and dependent variables.
		Designing the experiment procedures using science and engineering skills.	1. Designing an experiment procedure equipped with factors supporting the experiments. 2. Designing an experiment procedure which can be used to test the hypothesis. 3. Designing an experiment procedure which can be used to control variables systematically. 4. Designing an experiment procedure equipped with a procedure of work safety.
		Drawing conclusion based on the experiment using science and mathematics skills.	1. Drawing conclusions related to discussion 2. Drawing conclusions based on the experiment objectives. 3. Drawing conclusions in the form of mathematical logic. 4. Drawing conclusions in the form of diagrams.

No.	Aspect	SubAspect	Indicators
4.	Elaborating	Discussion of the experiment results using science and technology skills.	1. Discussing the results supported by two relevant and competent references (book and Internet). 2. Elaborating the meaning of the experiment data interpretation in the discussion. 3. Elaborating reasons why a hypothesis is accepted or rejected in the discussion.
5.	Evaluating	Identifying errors in the experiment using science and technology skills.	1. Identifying errors based on the experiment facts. 2. Identifying errors based on the calculation resultsof the measurement uncertainty using the Excel program. 3. Identifying errors based on the theory of measurement uncertainty. 4. Identifying errors and offering suggestions.