



Validity of Evaluation Instrument on the Implementation of Performance Assessment to Measure Science Process Skills

Marina Ramadani¹*, S. Supahar¹, Dadan Rosana¹

¹ Department of Science Education, Graduate School of Universitas Negeri Yogyakarta. Jalan Colombo No. 1, Karangmalang, Yogyakarta, 55281, Indonesia

* Corresponding Author. Email: marinaramadani91@gmail.com

Received: 30 August 2017; Revised: 17 October 2017; Accepted: 23 October 2017

Abstract

This study aims to develop an evaluation instrument CIPP (context, input, process, product) model that suitable at implementation of performance assessment instrument to measure students' science process skills of junior high school. Methods in this study are research and development (R & D) which is adopted by model Borg & Gall model. This evaluation instrument used one of evaluation CIPP model with context, input, process, and product components. The evaluation instrument is used to determine the implementation of performance assessment instrument to measure junior high school students' science process skills in excretion system topic. This development study involves some experts to give validation of the development of feasibility products. Experts involved, they are two lecturers of evaluation expert, two junior high school science teachers, and four peers. The validity of evaluation instruments was analyzed using V'aiken formula. The results of this study are evaluation instrument of CIPP model of the implementation of performance assessment to measure junior high school students' science process skills of grade VIII in excretion system topic. The results of the validity of this product indicates that this evaluation instruments are eligible in using with V'eiken coefficient of 0.86. It indicates that the evaluation instruments are valid in substance, construction, and language aspects.

Keywords: evaluation instrument, performance assessment, science process skills

How to Cite: Ramadani, M., Supahar, S., & Rosana, D. (2017). Validity of evaluation instrument on the implementation of performance assessment to measure science process skills. *Jurnal Inovasi Pendidikan IPA*, 3(2), 180-188. doi:<http://dx.doi.org/10.21831/jipi.v3i2.15534>

Permalink/DOI: <http://dx.doi.org/10.21831/jipi.v3i2.15534>

INTRODUCTION

Society in the 21st century is increasingly aware of the importance of preparing the young generation who are flexible, creative, and proactive. At this time, it is necessary to build young generation who are skillful to solve problems, wise to make a decision, creative thinking, active in discussing, can communicate ideas effectively, and they are able to work efficiently both individuals and groups. To build that generation, only mastery of the concept is not enough. Through learning science, students need to be equipped with skills in facing all the symptoms and science phenomena based on the basis of scientific methods, so that students are able to apply their knowledge in daily life. Based on competency standards of science graduated in this learning activity, students are required to have observation skills with the

appropriate equipment, conducting science experiments based on the procedures that have applied, noting the results of observations and measurements, interpreting the result of observations to tables and graphs, and concluding and communicating the results of observations towards peers in orally or writing. To determine student achievement in fulfilling all of these skills, the teachers can use one of assessment instruments refer to measure the science process skills, one of them is performance assessment instrument. Assessment is a conventional activity, which is practiced in schools on a day-to-day basis. It is also a process, which helps in developing student's learning. It provides the teacher with an opportunity to review their own teaching to enhance students' learning (Azim & Khan, 2012, p. 11). Assessment for learning includes all those activities undertaken by

teachers, and by the students in assessing themselves. Which provides information to be used as feedback to modify the teaching and learning activities in which they are engaged (Earl & Giles, 2011, p. 12).

Chabanlengula explains that in learning science, an assessment is used to examine and describe students' achievement and process in one or more domains of science such as the cognitive, affective, psychomotor, applications, creative, and the nature of science. In fact, domain that is still rare in assessing is psychomotor domain. Domain psychomotor often designed as a performance or practice skills, one of them in learning science is science process skills including observation, measurement, experimentation, classification, communication, inference, and prediction aspects. In general, the domain of psychomotor skills can be realized and demonstrated by students through hands-on activities in laboratory work. The performance assessment can provide the basis for teachers to evaluate the effectiveness of the process through data collection and product through the report and task that teachers can used in learning activities (Chabalengula, Mumba, Hunter, & Wilson, 2009, pp. 168–169). The performance assessment is useful to record the ability in applying knowledge to the students' authentic situation. In assessing the performance assessment, teachers must have good teaching skills in order to achieve the learning objectives that are multi-domain (McLellan, 2008, p. 2).

Furthermore, Oloruntegbe (2010, pp. 13–14) explains that the main purpose of the assessment is to provide quantitative data of students' performance especially in science learning. This is related to systematic of authentic assessment that in this assessment gained the students to implement their knowledge and skills at the same time and can be used in real life outside school activities. One of authentic assessments is performance assessment. This assessment is not only requires students to memorize the basic concept but also invites students to demonstrate their knowledge and understanding through product, performance or exhibition. There are a variety of tasks that are applied in performance assessment including trains students' responses, submit assignments and portfolios, teachers' observation and categorize questions as a composite of responses; product; and performance, skills training, essay writing, oral presentation, solve the problem, and etc. The experts say that performance

assessment is possibly be a meaningful indicator of what students know and what they can do. It is to promote active learning activities. It is also a change in teaching practices to support effective science learning (Oloruntegbe, 2010, p. 13).

In this era, assessment instrument has already developed to measure it on science process skills of excretion system topic or junior high school students of grade VIII. By developing performance assessment instrument, it is a bit much to answer the demands in the implementation of curriculum 2013 especially on a scoring system. According to Xiufeng Liu, performance assessments are appropriate to assess students' understanding related to learning activities that this performance assessment is a hands-on test that required students to show their performance by applying scientific procedure, conducting a scientific investigation, or producing useful products. In addition, the implementation of assessment is not only used at the end of the lesson, but also should be used during the learning time. By combining the performance assessment into learning activities continuously, it can raised learning goals and assessment. Both of them are definitely needed in learning (Ghazali, 2016, p. 156). The performance assessment is very useful when learning targets want to see students' performance from the learning process or form learning outcomes. Teachers are possibly to observe the process, assess the results, or even both of them. To assess students' skills using performance assessment, the teachers should develop performance task first. Performance tasks that are well designed will provide the opportunity for students to apply their learning in a new situation that more real. With this performance task, it will help students to connect their skills specifically. In this case, science process skills with their knowledge in applying the learning materials at schools with activities in their daily life. Besides, in performance assessment must also include an assessment rubric. With the scoring rubric that students can know, they will tend to improve their performance and they will use their skills actually and knowledge during the learning process (Brookhart & Nitko, 2008, p. 168). Performance-based assessment describes one or more approaches for measuring student progress, skills, and achievement. Authentic assessment are considered tasks of real world application of content knowledge. Performance assessment encompasses authentic

performance assessment to convey alternative forms of assessment. Teacher include quality assessment in their instruction activities to provide consistent guidance for planing and instruction. Student can participate from the very beginning of instruction by demonstrating their strengths through authentic performance assessment. Through the use of pre-assessing with alternative assessment measures, students can demonstrate their abilities, strenght, knowledge, likes, and desires that can guide classroom instruction (Obergh, 2011, p. 2). Performance assessment can be concluded that performance assessment is an assessment which requires students to do a task that will be assessed later by the teacher using rubric (Pratama & Rosana, 2016, p. 103).

One is main points of science is look at science as a process. In learning science, the scientific approach can be applied to explore students' science process skills. Science process skills are difened as the adaptation of the skills used by scientists for composing knowledge, thinking of problems and making conclusions. Science process skills also defined as facilitating basic activities in regards to learning science, gaining research method and techniques, helping students to be active and to make learning permanent. Science process skills are classified as basic (observation, testing, classification, relating number with space, and recording data), casual (prediction, determination of variables, and drawing a conclusion) and experimental (making hypothesis, modeling, doing the experiment, changing and testing variables, and making a decision) (Karsli & Şahin, 2009, p. 2). In teaching science concepts, teachers should train and teach students to use science process skills that appropriate to the science material that is taught (Goldston & Downey, 2013, p. 130). The science process is the number of skills to examine a natural phenomenon with certain ways to acquire science knowledge. Thus, when students understand science concepts using their science process skills, so students can learn science based on what the scientists did (Bundu, 2006, p. 12). In obtaining the science process skills, it is considered as "learning how to learn" because students learn how to learn to think critically and creatively using information (Rauf, Rasul, Mansor, Othman, & Lyndon, 2013, pp. 47–48). Basic science process skills of observing, comparing and classifying, concluding, predicting, defining the operational, measuring, and creating charts and graphs

(Ergül et al., 2011, p. 53). Science process skills are used in real life as well as in science. Student are required to explain how real life event occur. Science process skills involve creativity and critical thinking along with scientific thinking. It is known that those who can think creatively and critically are an important factor in the development of a country (Karsli & Şahin, 2009, p. 3).

The development of performance assessment instrument is a learning program that the conducting of the programs need to be evaluated of its implementation. Evaluation of a program is a method to determine the performance of a program by comparing with pre-determined criteria or objectives to be achieved with the results that would be achieved. The results achieved in the form of information used as a consideration for making a decision and decide the policy (Retnawati, 2013). In order to the evaluation of this program runs systematically and knows actually of a success program, it is necessary to develop an evaluation instrument aimed to evaluate the implementation of performance assessment instruments that already exist. There are several models of evaluation that can be used to evaluate a program, one of them is CIPP evaluation model that is developed by Stufflebeam. This evaluation model based on context, input, process, and product. Each aspect of the fourth aspect of this CIPP evaluation model has different criteria (Stufflebeam, 2001, pp. 280–282).

Specifically, the context evaluation component of the Context, Input, Process, and Product evaluation model can help identify service providers' learning needs and the community's need. The input evaluation component can help perscribe a responsive project that can best address the identified needs. Next, the process evaluation component monitors the project process and potential procedural barriers, and identifies needs for project adjusment. Finally the product component measures interprets, and judges project outcomes and interprets their merit, worth, significance, and probity (Zhang et al., 2011, pp. 57–59). The CIPP model is the most widely known and applied model that is used by evaluator. This model was developed by Stufflebeam. Stufflebeam stated that the core concept of the CIPP model denoted by the CIPP acronym, which stands for the evaluation context, input, pprocess, and product. The CIPP evaluation model belongs in the improvement/accountability category, and is one of the most widely applied evaluation models. One of

the strenght of CIPP model is, especially, that is a useful and simple tool for helping evaluators produce questions of vital importance to be asked in an evaluation process (Divayana, Sanjaya, Marhaeni, & Sudirtha, 2017, p. 2).

The CIPP evaluation model is useful framework for analyzing the interrelationship between the four evaluation dimensions. These four dimensions of evaluations also serve planning, structuring, implementing, and recycling decisions respectively. As such context evaluation involves confirming the present objectives, to modify the existing objectives or develop new ones. Input assess the strategies, personnel, resources or procedures in achieving the program's objectives. Process evaluation is looking at everything related to the implementation of already selected design, strategies or action plan. Product evaluation determines and examines the spesific outcomes of the program (Abdullah, Wahab, Noh, Abdullah, & Ahmad, 2016, pp. 5–6). It is consistent with the definition of operational assessment in this study the evaluation process is a draw, collect, and provide useful information to choose alternative decisions for improvement. The level of context, input, process, product can provide important strategies or action plans for the structuring of a more robust implementation of the performance assessment to measure science process skills.

Aspects of evaluation context help in analyzing the needs to be achieved by the program that will be implemented so it can help in planning decisions. Input evaluation aspects help to build the readiness of human resources and the surrounding conditions. In this case, the input aspects of the three objectives are the evaluation of students' characteristics, teachers' competencies, and the availability of schools' facilities and infrastructures. Process aspects help in implementing a program. Thus, in this aspect, the evaluators observe the conducting of a program. In this case, it is the enforceability of the implementation of assessment instruments. Product aspects evaluate the result of applying a program (Stufflebeam, 2001, pp. 280–282). Therefore, the aim of this study is to develop an evaluation instrument with CIPP model on the implementation of performance assessment instrument to measure the science process skills students of grade VIII on the topic of the excretion system. In each study, the instrument is something that has a very important position because the instrument will determine the quality of the data collection. The higher the

quality of the instrument, the higher also of the evaluation results (Arikunto & Jabar, 2009, pp. 92–94).

An instrument is valid when it is measuring what is supposed to measure. In the other words, when an instrument accurately measures any prescribed variable it is considered a valid instrument for that particular variable. There are four types of validity, face validity, criterion validity, content validity, or construct validity. Face validity is looking at the concept of whether the test looks valid or not on its surface. Criterion validity is a concept which will be demonstrated in the actual study as to establish it needs a good knowledge of theory relating to the concept and a measure of the relationship between our measure and those factors, whereas content validity is looking at the content of items. Finally is the construct validity, which measures the extent to which an instrument accurately measures a theoretical construct that it is designed to measure. There is a relationship between validity and reliability. Any instrument can be reliable but not valid however, it cannot be valid if it is not reliable. In other words, if an instrument is valid, it must be reliable. And in general, checking for validity of an instrument is more difficult than checking for reliability because validity is measuring data related to knowledge whereas reliability only concerns with the consistency of scores (Ghazali, 2016, p. 80).

METHOD

This research is a development research that is developing an evaluation instrument on the implementation of performance assessment instruments to measure science process skills topic in the excretion system. The development model used in this research is to integrate 4D model proposed by Thiagarajan with non-test instrument of development model. 4D development model has steps such as define, design, develop, and disseminate, while the development model of non-test instrument has 7 steps: (1) disseminate the specifications of the instrument, (2) write the instrument, (3) determine the scale of the instrument, (4) determine the scoring system, (5) comprehend the instruments, (6) validate the instrument, and (7) analyze the instrument.

Development product in this study is evaluation instrument CIPP model that consists of context, input, process, and product. From each aspect of the CIPP model evaluate different

components. In context components, evaluators analyze the needs. In this case, the needs analysis is intended to build up information about the importance of the implementation of performance assessment instrument related to the demands of curriculum 2013. In input component techniques, the evaluation focused on teachers' competencies, students' characteristics, and the availability of schools' facilities and infrastructures. In evaluation process components, it is forced at the conducting in implementing of performance assessment instrument in science learning. In product component, evaluation is focused on the results of the implementation of performance assessment instrument and students' learning outcomes related to the science process skills. These are grilles of evaluation instrument CIPP model on the implementation of performance assessment instrument to measure the science process skills.

Table 1. Grilles of evaluation instrument CIPP model on the implementation of performance assessment instrument to measure the science process skills

Component	Aspect	Indicators
<i>Context</i>	needs analysis	Teachers' needs
	Teachers' competency	Pedagogic, professional, characteristic, and social
<i>Input</i>	Students' characteristic	Students' interest
	Facilities and infrastructures	Laboratory condition and the availability of equipment and material
<i>Process</i>	Planning	Planning of the use of performance assessment instrument
	Conducting	The implementation of performance assessment instrument
<i>Product</i>	Result of performance assessment instrument implementation	The accuracy of using, skills improvement, and the effectiveness of instrument implementation

This development study used several data collection techniques such as used observation and questionnaires so the data collection instru-

ments using observation sheets and questionnaire sheets. Observation sheets used to evaluate the program of input, process, and product component like observation items. It was used questionnaire sheets to evaluate the program in context component. Besides, questionnaire sheets were also used of expert lecturers, science teachers, and peers to see the validity of the instrument. To determine the validity of this evaluation instrument CIPP model, data analysis techniques were used is Aiken. Content validity was gained from scores given by every aspect of the assessment. The assessment scores obtained from seven validators. Then, content validity coefficient was calculated using Aiken formula. Each questionnaire validation item, it was converted into qualitative data with V index ranges between 0 and 1. Score calculation results of V compared with the table category of V Aiken. According to Aiken for numbers of assessment categories are 4 to 7 rater, limited V number called good if ≥ 0.86 (Aiken, 1985, pp. 131-134).

RESULTS AND DISCUSSION

This development research aims to produce an evaluation instrument on the implementation of performance assessment instrument to measure students' science process skills that fulfill valid criteria. The product of this research is evaluation instrument CIPP model that aims to evaluate the implementation of performance assessment to measure science process skills. This evaluation instrument was developed based on 4D model and development model of non-test instrument..

Define Stage

Define stage is a stage of preliminary study that aimed at preparing the product to be developed to collect information and analyze the factors that caused problems so it required product development to solve those problems. Step of analysis needs start from conducted interviews with several junior high school science teachers. These analysis needs were conducted to determine the needs for an evaluation of the conducting on the implementation of performance assessment instrument to measure the science process skills. The results of the interview was obtained an information that need to be evaluated toward the product that has developed. However, evaluation program activities will be more structured and obtain maximum results, it needs using evaluation

guidelines including evaluation instruments that could be the basis of the evaluator in observing and assessing the enforceability of the implementation of the performance instruments.

Based on the results of observation of the application of the system of teacher assessment conducted during the learning process on December 7, 2016 in SMP 3 Tempel (Junior high school) obtained the fact that in science learning aspect of the achievement of the results of the study are rarely measured are aspects of skill. In addition, the current performance assessment instruments developed to measure skills in the process of science students on the material system excretion compiled by Deby Kurnia Dewi with IPR (Intellectual Property Right) number C22201604796 in 2016. Based on the results of the research of Deby Kurnia Dewi, performance assessment instrument science process skills have been applied in SMPN 15 Yogyakarta (Junior high school) but the results, the effectiveness of the application of performance assessment instrument is not yet known through the evaluation. In an evaluation of the activities, thing that is very important to the evaluation instrument is prepared. The instrument is the thing that has a very important position, because by using the instruments will determine the quality of the data collected. The information gathered in the evaluation activities are very useful in decision-making and follow-up of the program policies, because the evaluation of the results of the input evaluator can determine the follow-up of the programme has been implemented, in this case is the application of performance assessment instruments. There are four possibilities that policy can be made based on the results of the implementation of a program, namely (1) terminate the programe, (2) revise the program, (3) continue the program, or (4) disseminating the program.

Besides analyzing the needs, further analysis evaluation model that will be used in the evaluation of the program. There are several evaluation model developed by several evaluation experts. In this study, the researchers chose to use an evaluation CIPP model that is developed by Stufflebeam. CIPP Model has four components which each evaluation of each component has a different criteria. In context components, main point of evaluation is on analysis of the needs of teachers. In input component, the main point of evaluation is on teachers' competency, students' characteristic,

and the availability of facility and infrastructure. In process component, the main point of evaluation is in planning and implementing the use of evaluation instruments in the learning activities by teachers. In product component, the main point is on the results of the evaluation of the implementation of performance assessment instrument itself.

Design Stage

The design stage is a stage of drafting or it can be called as early design product that will be developed. The first step is to choose the form and format of an evaluation instrument. Determination of shape and this format should be adapted to the evaluation model selected. Appropriate forms of evaluation instrument are observation and questionnaires. A questionnaire sheet is used to evaluate the components of context because the data was obtained based on answers of science teacher. An observation sheet is used to observe the implementation of the program especially on input, process, and product components. The development of evaluation instruments should also specify the format that will be developed. The first format specified is determining the title of the instrument. It is "Evaluation instrument CIPP Model on The Implementation of Performance Assessment Instrument to Measure Students' Science Process Skills". Next step is preparation of the instructions for use of performance assessment that aims to make the observer to understand how to use the evaluation instrument.

The next stage is determining the design of evaluation indicators of program implementation. The arrangement of this indicator is based on the analysis of each criterion of CIPP evaluation model components. After determining the indicator, then the next step is to make design of evaluation instrument that are interpreted into statement items that suitable with the indicator. The next step is writing evaluation instrument. This instrument is written based on design by considering the substance, construction, and language that make thi instrument easier to use directly.

The next design stage is determining the scale of evaluation instruments. The scale used in this evaluation instrument development is rating scale with a scale of 1 to 5. Thus, the scoring system used in this study refers to a scale of 1 to 4 based on the observation result of appearance options that provide of each items. To make observer easier in scoring, the

researchers arrange the rubric as a basis in scoring. Rubric contains five possible conditions that will arise. Then, it was compared with the ideal conditions that has determined by the researcher.

Develop Stage

Product development is done to produce an evaluation sheet that is ready to be experimented. The early step in an evaluation instrument development stage of CIPP model are arrange early draft. Then, it validated by expert lecturers, science teachers, and peers. The preparation of these instruments is based on grilles that obtained from the synthesis of literature review. Validation of evaluation instrument CIPP model is viewed from several aspects including substance, construction, and language aspects. Then, the experts give a decision of each item. Data results from validators was analyzed descriptively by calculating the content validity coefficient using V Aiken analysis later. Based on V Aiken formula, it can be determined the feasibility of the product of evaluation instrument CIPP model. Rate of validation product of evaluation instrument CIPP model using a scale of 4 with the 7 rater. Th summary of V Aiken value can be seen of the Table 2.

Table 2. Results of V Aiken Analysis

Evaluation Component	Item Number	V value	Description
Context	1,2,3,4,5,6	0,90	Valid
Input	1 - 17	0,95	Valid
Process	1,2,3,4	0,90	Valid
Product	1,2,3,4	0,88	Valid

Based on Table 2, it can be seen that the results of the validation of 7 V Aiken rater shows the value of each evaluation component ranges in the range of 0.86 to 1.00. Therefore, the calculation of Aiken of each item ≥ 0.86 so the whole instrument items are declared valid. After finding the V value, then it analyzed the decision of the experts whether evaluation instrument CIPP model is appropriate to use without revision, appropriate to use with according to suggestion, or inappropriate to use. The result of the experts of evaluation instrument CIPP model completely is appropriate to use with revision based on the suggestions.

CIPP model evaluation instrument has been developed by Ghazali (2016). The CIPP evaluation instruments used to evaluate School-Based Assessment System. Based on the results

of the research of Hasnida model CIPP evaluation suitable to evaluate the assessment instruments because the CIPP model is based on the management-oriented evaluation approaches that will help decision makers to plan, implement, and evaluate programs. Evaluation of model CIPP involve decision makers, so that all components of the CIPP model presents a staple in the decision in the form of planning, structuring, implementation, and repeated in a row processing where information obtained from each stage of the previous evaluation components [14]. A good evaluation instruments are instruments that meet certain conditions, can provide accurate data, and only measure specific behaviors samples. One of the conditions of good evaluation instruments are valid. Instruments that are developed must be able to measure what is measured.

Assessment is a vital component in education. The interaction between assessment, curriculum and instruction is very important if we were to improve the teaching and learning process in school. Performance assessment is one of the main elements that contribute to this. Currently, the instrument to evaluate the implementation of performance assessment to measure science process skills. According to (Ghazali, 2016, p. 156) presenting the value of validity of an instrument is important so that other researchers are confident with the quality of the data.

CONCLUSION

The results of this study are evaluation instrument of CIPP model of the implementation of performance assessment to measure junior high school students' science process skills of grade VIII in excretion system lessons. The results of the validity of this product indicates that this evaluation instruments are eligible in using with V'eiken coefficient of 0.86. It indicates that the evaluation instruments are valid in substance, construction, and language aspects.

REFERENCES

- Abdullah, N., Wahab, N. A., Noh, N. M., Abdullah, E. M., & Ahmad, A. (2016). The evaluation and effectiveness of school based assessment among science teachers in Malaysia using CIPP Model. *International Journal of Advanced and Applied Sciences*, 3(11), 1–7. <https://doi.org/https://doi.org/10.21833/ija>

- as.2016.11.001
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Arikunto, S., & Jabar, C. S. A. (2009). *Evaluasi program pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Azim, S., & Khan, M. (2012). Authentic assessment: An instructional tool to enhance students learning. *Academic Research International*, 2(3), 314–320. Retrieved from http://ecommons.aku.edu/pakistan_ied_pdf
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Prentice Hall.
- Bundu, P. (2006). *Penilaian keterampilan proses dan sikap ilmiah dalam pembelajaran sains SD*. Jakarta: Depdiknas RI.
- Chabalengula, V., Mumba, F., Hunter, W., & Wilson, E. (2009). *A model for assessing students' science process skills during science lab work. Problems of Education in the 21st Century* (Vol. 11). Retrieved from <http://www.scientiasocialis.lt/pec/node/206>
- Divayana, D. G. H., Sanjaya, D. B., Marhaeni, A. A. I. N., & Sudirtha, I. G. (2017). CIPP evaluation model based on mobile phone in evaluating the use of blended learning platforms at vocational schools in Bali. *Journal of Theoretical and Applied Information Technology*, 95(9), 1983–1995. Retrieved from <http://www.jatit.org/volumes/Vol95No9/10Vol95No9.pdf>
- Earl, K., & Giles, D. (2011). An-other look at assessment: Assessment in learning. *New Zealand Journal of Teachers' Work*, 8(1), 11–20.
- Ergül, R., Simsekli, Y., Çalis, S., Özdilek, Z., Göçmençebebi, S., & Sanli, M. (2011). The effects of inquiry-based science teaching on elementary school students' science process skills and science attitudes. *Bulgarian Journal of Science and Education Policy*, 5(1). Retrieved from <http://see-articles.ceon.rs/data/pdf/1313-1958/2011/1313-19581101048E.pdf>
- Ghazali, N. H. M. (2016). A reliability and validity of an instrument to evaluate the school-based assessment system: A pilot study. *International Journal of Evaluation and Research in Education (IJERE)*, 5(2), 148–157. <https://doi.org/http://dx.doi.org/10.11591/ijere.v5i2.4533>
- Goldston, M. J., & Downey, L. (2013). *Your science classroom*. Los Angeles, CA: Sage.
- Karsli, F., & Şahin, Ç. (2009). Developing worksheet based on science process skills: Factors affecting solubility. *Asia-Pacific Forum on Science Learning and Teaching*, 10(1). Retrieved from https://www.eduhk.hk/apfslt/download/v10_issue1_files/sahin.pdf
- McLellan, S. (2008). When students teach: Performance based assessment. *Transformative Dialogues: Teaching & Learning Journal*, 2(2), 1–12. Retrieved from https://www.kpu.ca/sites/default/files/Teaching_and_Learning/TD.2.2_McLellan_When_Students_Teach.pdf
- Oberg, C. (2011). Guiding classroom instruction through performance assessment. *Journal of Case Studies in Accreditation and Assessment*, 1–11. Retrieved from <http://www.aabri.com/manuscripts/09257.pdf>
- Oloruntegbe, K. O. (2010). Approaches to the assessment of science process skills: A reconceptualist view and option. *Journal of College Teaching and Learning*, 7(6), 11.
- Pratama, S., & Rosana, D. (2016). Pengembangan performance assessment untuk mengukur dan memetakan practical skills IPA siswa pada guided-PjBL di SMP. *Jurnal Inovasi Pendidikan IPA*, 2(1), 100. <https://doi.org/10.21831/jipi.v2i1.8372>
- Rauf, R. A. A., Rasul, M. S., Mansor, A. N., Othman, Z., & Lyndon, N. (2013). Inculcation of science process skills in a science classroom. *Asian Social Science*, 9(8), 47.

<https://doi.org/10.5539/ass.v9n8p47>

- Retnawati, H. (2013). *Evaluasi program pendidikan*. Jakarta: Universitas Terbuka.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation, 2001*(89), 7–98.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). *Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs*. *Journal of Higher Education Outreach and Engagement* (Vol. 15). Institute of Higher Education and the Office of the Vice President for Public Service and Outreach, University of Georgia. Retrieved from <http://openjournals.libs.uga.edu/index.php/jheoe/article/view/628>