

Authentic assessment: Evaluation and its application in science learning

Maria Hanifah¹*, Alfred Irambona²

¹ Universitas Negeri Yogyakarta. Jalan Colombo No 1, Karangmalang, Yogyakarta, 55281, Indonesia

² Burundi University. Boulevard Mwezi Gisabo, Bujumbura, Burundi

* Corresponding Author. E-mail: maria.hanifah05@gmail.com

Received: 6th September 2018; Revised: 10th September 2019; Accepted: 2nd October 2019

Abstract

The research aims to reveal: (1) the suitability of the implementation of science authentic assessment in Kulonprogo Regency, and (2) the forms of science authentic assessment applied. The study was used the CIPP Stufflebeam evaluation model. The quantitative data were analyzed with T-score, while the qualitative data were analyzed by using Miles & Huberman. The results of this research are as follows. (1) The implementation of science authentic assessment is wholly fairly effective be seen from T-score analysis 52.44 from the score extent 20-80. (2) The forms of science authentic assessment are practice, portfolio, teacher transcript journal, and daily test, whereas project activity, self assessment, and peer assessment are not done for the heat matter and its movement with T-score 45.14 wholly fairly effective from the score extent 20-80.

Keywords: implementation, natural science authentic assessment, 2013 Curriculum

How to Cite: Hanifah, M., & Irambona, A. (2019). Authentic assessment: Evaluation and its application in science learning. *Psychology, Evaluation, and Technology in Educational Research*, 1(2), 81-94. doi:<http://dx.doi.org/10.33292/petier.v1i2.4>



INTRODUCTION

Curriculum is basically demanded to change in accordance to the turn of the century. In relation to the demand, the new curriculum deals with challenges to manifest its implementation by targeting three key elements of success. The three key elements of success are students, teachers and materials (Sparapani, Callejo Perez, Gould, Hillman, & Clark, 2014). Thus, these key elements of success become the subject of evaluation. Therefore, prior to the implementation of the new curriculum, the old or the previous curriculum should be evaluated first in order to gain better description for the policy of the new curriculum implementation. The reason is that the implementation of a curriculum should be in balance with the capacities and the readiness of the curriculum implementers namely the teachers, the students and the principals since these people hold the immediate function of the curriculum implementation. A good example over such situation is provided by the implementation of the 2013 Curriculum. The implementation of the new curriculum has brought about several impacts to the assessment of the subjects, including Natural Science for Junior High School Degree. One of the impacts might be traced in the Regency of Kulon Progo, which covers curriculum planning, implementation and evaluation.

The 2013 Curriculum emphasizes that the teachers should implement the authentic assessment within the learning process. The authentic assessment in this 2013 Curriculum refers to the Minister of Education and Culture Regulation Number 104 of 2014 on the Standards of Educational Assessment. According to (Kunandar, 2013, pp. 35–36), authentic assessment refers to the activities of gathering the data that provide descriptions on the development of the students, both in terms of process and of results, by means of assessment instrument.

The Natural Science Authentic Assessment in the Junior High School Degree throughout the Regency of Kulon Progo might be actually traced by conducting an evaluation research. The pilot schools that have implemented the 2013 Curriculum for three semesters consecutively is obliged to

keep implementing the 2013 Curriculum. The results of the 2013 Curriculum Implementation later will be evaluated by using the Natural Science Authentic Assessment with the assessment standards for the material "Heat and Its Transfer." In this evaluation, the CIPP Model will be adopted.

The authentic assessment should be conducted because it offers wide usefulness. The characters of the authentic assessment are namely: (1) having relevance with the real world; (2) encouraging the students to develop relevant relationship with the real-world tasks; (3) provide complex tasks continuously; (4) observing different perspective tasks; (5) displaying self-reflection; (6) displaying teamwork; (7) pursuing meaningful artwork; and (8) generating competitive problem-solving paradigm (Basuki & Hariyanto, 2014, p. 170). The authentic assessment will create active, innovative, socialist and scientific students.

Authentic assessment should be conducted comprehensively in order to assess the input, the process and the output of the learning process. An assessment thus becomes a fundamental development when both the teachers and the students use information from the assessment as the teaching information and the renewal of the learning process. Then, assessment in general might describe the test in the end of a chapter, in the rapport of the semester, in the daily measurement of the skills and in the composition of final laboratory rapport (Atkin, Black, & Coffey, 2001; National Research Council, 2001). The assessment therefore will cover the overall process that the students experience and will be ended by the communication on the results of the assessment that has been performed.

According to Kunandar (2013, p. 42), the component that should be given attention in conducting the authentic assessment consists of three aspects. The first aspect is authentic in terms of instrument, meaning that the authentic assessment instrument should be varied. The varied instruments might be written test, oral test, project test and performance test. Then, the second aspect is authentic in terms of measurement, meaning that the authentic assessment instrument should measure the learning results based on the competencies in the attitude, knowledge and skills. Last but not the least, the third or the last aspect is authentic in terms of condition, meaning that the authentic assessment instrument should oversee the input, the process and the output.

Furthermore, the authentic assessment should be continuous or include the overall implementation of the assessment activities. The authentic assessment is designed in order to accompany the standardized test assessment; thereby, the authentic assessment by the teacher is performed in order to complete the existing standardized test (Pantiwati, 2013). As a result, the authentic assessment might cover the standardized test, the assignment and the project during the learning process. The authentic assessment ensures that all students are given the opportunity to show what they are capable of while teachers have the necessary information for establishing the balanced and fair assessment to each student. In other words, the authentic assessment is able to facilitate the overall needs of the students in their development and might be summarized in the portfolio assessment. Portfolio assessment delivers a response that the students is able to understand the learning objectives and expand their learning horizon (Merilia, Fajaruddin, & Arbain, 2019; Nezakatgoo, 2011).

As having been implied, the authentic assessment includes three aspects namely attitude, knowledge and skills. According to the Ministry of Education and Culture (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2014), the aspects of attitude assessment might consist of observation, self-assessment, peer assessment and journal whereas the aspects of knowledge assessment might consists of written test, oral test and assignment; similarly, the aspects of skills assessment might consist of practice, project and portfolio. In relation to the so-called authentic assessment, (Lukum, 2015, p. 64) defines that Natural Science contains empiric, systematic and logical scientific process altogether with the attitude of being curious, appreciating evidence, being patient, being critical, being persistent, being creative and being innovative. Natural Science is a process that should be learned under scientific manner by means of numerous methods and models in the form of facts, concepts, principles, laws and theories.

According to Arikunto and Jabar (2009), evaluation refers to the activities of identifying the achievement of a program. Evaluation provides information for the policy-making activities under two manners. First, evaluation provides information for the policy makers about the educational situations or the learning achievement of a certain group of people. Second, evaluation provides information that might be benefitted as an administrative tool for implementing certain policies. Departing from the two manners of evaluation, the evaluation in Natural Science aims at describing the learning programs of Natural Science for the Junior High School Degree in the Regency of Kulon Progo. Thus, it might

be inferred that assessment program evaluation refers to the activities of evaluating the assessment process within the Natural Science learning process in accordance to the assessment standards, which consists of assessment planning, assessment implementation and assessment process altogether with their impact toward the students.

In specific definition, evaluation is an important process that covers all authentic assessments. In the case of Natural Science, the evaluation research in this domain does not entail any specific measurement for determining the final decision as the judgment. The results of such evaluation might be adopted as a matter of final recommendation (Levin-Rozalis, 2000, p. 24). Consequently, evaluation demands intelligence supported by logical steps in delivering the final decisions. Then, in the context of the present study, the evaluation research is conducted in order to solve the problems that have been related to: (1) the fitness between the Natural Science authentic assessment implementation in the Regency of Kulon Progo and the 2013 Curriculum assessment standards; and (2) the Natural Science authentic assessment forms that have been implemented in the 2013 Curriculum throughout the Regency of Kulon Progo.

The results of the evaluation study are expected to be benefitted by the teachers as a matter of consideration for improving the quality of the Natural Science learning process in responding the policy of curriculum change. For the schools, it is expected that the results of the evaluation study might serve as a matter of consideration for the learning process evaluation and might also provide several inputs for the 2013 Curriculum implementation in the future. In addition, the results of the study are also expected to be benefitted by the policymakers as a matter of consideration for tackling the polemic behind the curriculum change that has been spread within the society so that there will be better preparation within the implementation of the 2013 Curriculum. Last but not the least, for the researchers it is expected that the results of the evaluation study might serve as a conceptual review in answering the curriculum-related problems behind the Natural Science learning process within the 2013 Curriculum so that the results of the evaluation might be benefitted as the reference for the conduct of the research and development study in the future.

Natural Science should essentially have the dimensions for opening the students' paradigm toward investigation and knowledge expansion so that there will be interaction between the social aspects and the technological aspects within the students (Chiappetta & Koballa, 2010, p. 105). In this case, Natural Science develops the scientific process in order to shape the students' scientific thinking and therefore there should be special assessment technique in measuring the essence of Natural Science. Furthermore, the assessment that has been taking place should be an activity integrated into the teaching and learning process. The teacher should assess every single process that the students have experienced (Chiappetta & Koballa, 2010, p. 6). Therefore, the assessment of the Natural Science does not take place at the end of the learning process but also during the learning process. In the context of 2013 Curriculum, the assessment technique that should be adopted is the authentic assessment. Authentic assessment applies the concepts and the theories that the students have with the skills and the capacities that might be observed immediately (Andayani & Mardapi, 2012, pp. 166–167). This kind of Natural Science learning assessment emphasizes on the skills of the students who have been active in each learning process.

Furthermore, the essence of Natural Science is to develop the competencies in the scientific skills and nurturing the scientific products and attitudes (developing the concepts that have been related to the daily experiences). According to (Ali, Suastra, & Sudiarmika, 2013), Natural Science is related to the systematic discovery in the nature; consequently, Natural Science does not only deal with the mastery of a set of knowledge in the facts, concepts, principles, laws and theories but also deals with the process of discovery. In the same time, the materials in Natural Science provide a set of integrations from several domains namely Physics, Chemistry, Biology, Geography and Astronomy (Hewitt, Lyons, Suchocki & Yeh, 2007, p.xvi). Therefore, it might be concluded that Natural Science emphasizes the learning process that should be performed through a sequence of scientific-processing activities namely observations or experiments based on the Natural Science learning materials.

The main objective of Natural Science learning process is to involve the paradigm of all students and to deliver final definition to the students so that the students might expand their knowledge based on the materials that have been delivered. In relation to the statement, a Natural Science teacher should understand the lesson plans that will be implemented what the students should learn (Chiappetta & Koballa, 2010, p. 5). A good Natural Science learning process shapes knowledge

outcome, reasoning outcome, skill outcome, product outcome and affective outcome (Stiggins, 1994, p. 3). Basically, the learning process will shape the knowledge and provide the materials for the problem-solving activities, the inventing activities and the creative activities; in the same time, the learning process will also invite the students to apply their skills in mastering the learning materials. Therefore, the lesson plans should be equipped with the on-going assessment.

The on-going assessment, or the lifelong assessment, is part of the authentic assessment. Thus, the authentic assessment deals with some kind of evolution in relation to activeness of the students throughout the learning period, meaning that the assessment is conducted once throughout the course of certain period. Therefore, the authentic assessment should be multiphase, ongoing, interactive and collaborative. Furthermore, the implementation of the authentic assessment will deliver several insights for both the teachers and the students (Murphy, 1996, pp. 23–24). The authentic assessment should be ongoing so that the activeness of the students might be shaped well and the development of the students might be contained well by the teachers.

As having been asserted, the authentic assessment should be ongoing. The implication of the statement is that the authentic assessment should cover the implementation of all assessment aspects. The reason is that the authentic assessment has been designed in order to complete the standardized assessment; in other words, the authentic assessment by the teachers is benefitted to complete the existing standardized test assessment (Pantiwati, 2013, p. 5). Then, the authentic assessment might cover the standardized test, the assignment and the project during the learning process. The reason is that the authentic assessment strives to ensure all students have the equal opportunity to show what they are capable of and the all teachers have the necessary information for establishing fair and balance for all students. In addition, the authentic assessment might facilitate the overall needs of the students in their development and might also be summarized in the form of portfolio assessment.

The progress on the implementation of the Natural Science authentic assessment might be identified through the conduct of the evaluation study since evaluation will involve the use of measurement activities for making decisions with regards to the valuable objects (Van Blerkom, 2009). This object might be a program. Specific to the context of the present study, the object in the evaluation study will be the Natural Science assessment programs. Then, the evaluation model that will be adopted in the study is the CIPP Model. According to (Stufflebeam & Shinkfield, 1985), CIPP Model insists that evaluation is the process of delineating, obtaining and providing descriptive and judgment information in order to guide decision-making serve needs for accountability and promote understanding of the involved phenomena. In other words, Stufflebeam & Shinkfeld propose that CIPP Model is able to deliver useful description, results and information as a matter of consideration for making decisions responsibly. As having been implied, evaluation specifically delivers an important process that covers all aspects of authentic assessment. However, the evaluation in the Natural Science does not entail any specific measurement unit in delivering the final decision as part of the judgment. As a result, the results of this evaluation might only be adopted as the matter of final recommendation (Levin-Rozalis, 2000). Most importantly, the conduct of an evaluation should be supported by the logical procedures within the delivery of the final decision.

METHOD

The study was a program evaluation research. The program that had been evaluated in the study was the Natural Science assessment program within the 2013 Curriculum. Then, the technique that had been adopted in the conduct of the study was the descriptive qualitative technique and the descriptive quantitative technique. The evaluation itself was conducted toward the junior high schools that had joined the pilot project of 2013 Curriculum implementation throughout the Regency of Kulon Progo. In these junior high schools, the object that had been studied was the Natural Science learning process of 2013 Curriculum for Grade VII. This evaluation took place for approximately three months at the end of the even semester.

The population within the conduct of the study was both the teachers and the students from the junior high school degree that had implemented the 2013 Curriculum. With regards to the teachers, the teachers who had been evaluated were the teachers of Natural Science who had been teaching in Grade VII. Therefore, automatically the students who had been provided with the evaluation were the Grade VII students. The samples from the Grade VII students were divided into two groups by using the

purposive sampling technique. The classroom that had been taught by the selected teachers was selected as the classroom which learning process should be observed. The selected teachers, then, should select the two classrooms as the samples for the conduct of the study based on their consideration. In selecting the samples for the study, the formula that had been adopted was the Isaac and Michael Formula. The formula might be elaborated as follows.

$$S = \frac{\lambda^2 NPQ}{d^2 (N-1) \lambda^2 NPQ} \tag{1}$$

Note:

S = Number of sample

N = Number of population = 682

λ^2 = Chi-Square, which depends on the degree of freedom and rate of errors, $df = 1$, rate 5% and thus the Chi-Square is 3.841.

P = Q = 0.5

d = Differences between the sample mean score and the population mean score namely 0.05

The population that had been involved within the conduct of the study consisted of 682 Junior High School students with the following composition: (1) 55 students from Grade VII of Pangudi Luhur 1 Junior High School Kalibawang; (2) 192 students from Grade VII of State 1 Junior High School Wates; (3) 150 students from Grade VII of State 1 Junior High School Galur; (4) 97 students from Grade VII of State 1 Junior High School Samigaluh; and (5) 188 students from Grade VII of State 2 Junior High School Lendah. From the total 682 students, 246 students were selected as the sample for the study by using the S-sample calculation (Sugiyono, 2010). Then, this figure was equivalently distributed to the five schools of the 2013 Curriculum Pilot Project. Consequently, 49 students were selected from each junior high school as the sample with an exception that 50 students were selected from the State 1 Junior High School Wates since this junior high school had the highest figure of the students. The detail thus might be viewed in Table 1.

Table 1. Details of Research Sample

No.	Name of School	Subjects	Number of Respondents
1.	SMP Negeri 1 Wates	Natural Science Teachers	1
		Students	50
2.	SMP Negeri 1 Galur	Natural Science Teachers	1
		Students	49
3.	SMP Negeri 1 Samigaluh	Natural Science Teachers	1
		Students	49
4.	SMP Negeri 2 Lendah	Natural Science Teachers	1
		Students	49
5.	SMP Panggudi Luhur 1 Kalibawang	Natural Science Teachers	1
		Students	49
Total			246

Procedures

The data that had been gathered in the study were the oral information in the form of facts, written notification and numbers that supported the proposition of the problems behind the implementation of the Natural Science authentic assessment within the 2013 Curriculum throughout the Regency of Kulon Progo. Therefore, the data gathering technique that should be adopted was documentation study, questionnaire, observation and interview. In addition, data triangulation should also be adopted in order to ensure that there had not been any case of data overlap so that the accuracy of the data analysis might be maintained by using the three data gathering techniques that had been implemented. Then, the examples of the instrument guideline for the teacher questionnaire might be consulted in Table 2.

In addition, the instrument guidelines for the student questionnaire had also been designed. The student questionnaire was distributed in order to capture the data on the process of authentic assessment implementation. The student questionnaire itself consisted of 23 items and might be consulted in Table 3.

Table 2. Example of Instrument Guideline for Teacher Questionnaire

Aspects	Sub-Aspects	Indicators	Items		Total	
			Positive	Negative		
Planning	Teacher Capacity Analysis	1. Assessment Standards Understanding	1,2	-	2	
		2. Authentic assessment preparation training	3,4,5	-	3	
	Facility Analysis	3. Procurement of papers for assessment	6	-	1	
		4. Procurement of practical tools and equipment	7	-	1	
	Instrument Planning	5. Design of authentic assessment in the lesson plan	5. Design of authentic assessment in the lesson plan	8	-	1
			6. Design of authentic assessment in learning media	9	-	1
		7. Design of authentic assessment in student worksheet	7. Design of authentic assessment in student worksheet	10	-	1
			8. Design of assessment rubric	11	-	1
		9. Design of assessment framework	12	-	1	
		10. Clear instrument formulation	-	13	1	
		11. Design of self-assessment indicators that might facilitate the students' understanding in mapping their capacities	14	-	1	
		12. Design of authentic assessment plan in laboratory practice	15	-	1	
		13. Design of authentic assessment in project activities	16	-	1	
		14. Design of authentic assessment plan by means of portofolio submission	-	17	1	
Implementation	Preparation of assessment implementation	15. Delivery of aspects that will be assessed	18	-	1	
		16. Delivery of assessment period	19	-	1	
		17. Delivery of assessment forms that will be implemented	20	-	1	
	Attitude Assessment	18. Explanation on the directions of self-assessment instrument.	18. Explanation on the directions of self-assessment instrument.	21	-	1
			19. Peer-assessment to three classmates or vice versa	22	-	1
			20. Note-taking toward all findings during the learning process by the teachers	23	-	1
			21. Presence of daily test or assignment at the end of each chapter	24	-	1
	Knowledge Assessment Skills Assessment	22. Performance of experimental activities	22. Performance of experimental activities	25	-	1
			23. Note-taking on the findings by the students	26	-	1
		24. Conclusion drawing on the findings by the students	24. Conclusion drawing on the findings by the students	27	-	1
			25. Presentation on the experimental results by the students in the classroom	-	28	1
		26. Explanation on the experimental activities by the teachers	29	-	1	
		27. Explanation on the dangerous parts of the experiments by the teachers (safety-lab)	30	-	1	
		28. Conduct of the role as facilitator and provision of guidance to the students	31	-	1	
		29. Design of independent project activities	32	-	1	
30. Submission and storage of the students' work results in one folder and reminder for the students to mark on the date of the submission and storage		33	-	1		
31. Provision of revision period for the works that have not met the standards		-	34	1		
32. Check and return by the teachers for the students' work with educative feedback and comment		35	-	1		
Assessment Principles		33. Transparency of assessment results	33. Transparency of assessment results	37	36	2
	34. Use of assessment instrument		38	-	1	
	35. Implementation of enrichment program		39, 40	-	1	
	36. Implementation of remedial program		41	-	1	
	37. Implementation of revision program		42	-	1	
Report	Report Standards	38. Analysis on the results of the report	43,44	-	2	
		39. Composition of the report for the results of the students' learning process	45,46	-	2	
Total			41	5	46	

Table 3. Instrument Guidelines of Student Questionnaire

Aspect	Sub-Aspect	Indicators	Items		Total
			Positive	Negative	
Implementation	Assessment Preparation	1. Delivery of assessment aspect	1	-	1
		2. Delivery of assessment period	2	-	1
	Attitude Assessment	3. Delivery of assessment instrument form	3	-	1
		4. Explanation on the self-assessment instrument procedure	-	4	1
	Knowledge Assessment Skills Assessment	5. Implementation of peer-assessment to three classmates or vice versa	5	-	1
		6. Presence of daily test or assignment at the end of each chapter	6	-	1
		7. Performance of experimental activities and note-taking on the findings	-	7	1
		8. Conclusion-drawing on the findings by the students	8	-	1
		9. Presentation of the experimental results by the students in the classroom	-	9	1
		10. Explanation on experimental activities by the teachers	10	-	1
		11. Explanation on the dangerous parts of the experiment by the teachers (safety lab)	11	-	1
		12. Conduct of the role as facilitator and provision of guidance to the students	12	-	1
		13. Design of independent project activities	13,14	-	2
		14. Submission and storage of the students' work results in one folder complete with the submission date	15,16	-	2
	Assessment Principles	15. Provision of revision period for the works that have not met the standards	-	17	1
		16. Check and return for the students' works complete with the educative feedback and comments	18	-	1
		17. Transparency of assessment results	-	19	1
		18. Implementation of remedial and enrichment program	20	-	1
		19. Administration of daily test for the students who attend the remedial learning process	21	-	1
		Reporting Standards	20. Submission of the students' learning results	22,23	-
Total			18	5	23

Data, Data Gathering Instrument and Data Gathering Technique

The data gathering instrument was designed based on the indicators. The indicators that had been defined were in accordance to the needs of the study for each research subject. Within the study, the scale that had been adopted was the 4-point Likert Scale in the form of checklist (Widoyoko, 2012, p. 105). This scale was adopted for both the teacher questionnaire and the student questionnaire. Then, the use of the data gathering instrument might be classified into four parts. The first part was related to the teacher questionnaire guidelines and the teacher interview guidelines. Next, the second part was related to the input data or the planning data that had been attained from the "Input" stage. This part included the teacher questionnaire guidelines, the teacher documentation guidelines and the teacher interview guidelines. The teacher questionnaire should be completed first and this process proceeded to the interview and the documentation study with the teachers. In addition, the instrument analysis was also performed as complimentary activities within the documentation study. Furthermore, the third part was related to the process data or the data that had been attained by means of observation sheet and questionnaire. The process evaluation in this regard was performed by observing and documenting the assessment activities in order to enrich the relevant information. In this part, both the teachers and the students were observed during the conduct of the learning process. In order to check the accuracy of the observation results, both the teacher questionnaire and the student questionnaire were distributed. Last but not the least, the fourth part was related to the product data or the data that had been attained from the overall conclusion toward the conduct of the Context evaluation, the Input evaluation and the Process evaluation activities by means of teacher questionnaire, student

questionnaire, student interview guideline and documentation of final learning results. The flowchart for the conduct of the study might be consulted to Figure 1.

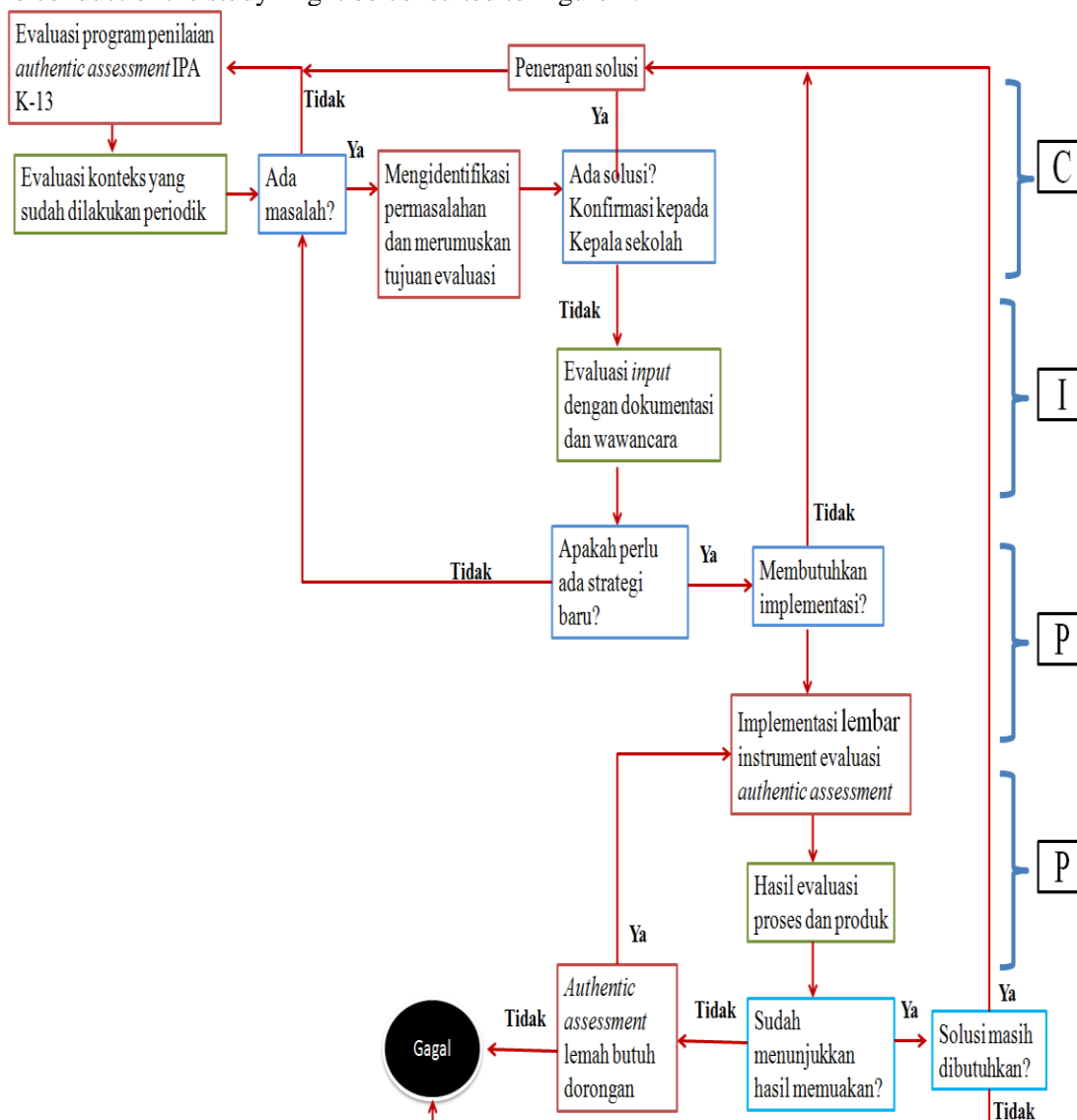


Figure 1. Flowchart of CIPP Evaluation Model

All of the instruments that had been administered should go through the validity test and the reliability test. The results of the validity test confirmed whether the instrument that had been designed were valid or not. On the other hand, the results of the reliability test confirmed whether the instrument that had been designed were reliable or not. The reliability of the research instrument, specifically the observation sheet and the questionnaire, was confirmed by conducting the Cronbach's Alpha estimates value calculation, while the reliability of the student questionnaire was analyzed by using the estimates reliability.

Furthermore, the content validity of the research might be confirmed by using the rational analysis. The conduct of the rational analysis described the items within the instrument indicators from the variables that had been measured within the study. The instrument items were validated based on the indicators by means of expert judgment in order to gather the necessary revisions from the experts. The data that had been attained from the expert judgment were in the form of 1 – 5 score. The validity of these data were confirmed by calculating the coefficient of content validity using the Aiken's V Formula. These data later were categorized based on the criteria of success. The formula proposed by Aiken (Azwar, 2015) might be elaborated as follows.

$$V = \frac{s}{\sum [n(c-1)]} \tag{2}$$

$$s = r - l_0 \tag{3}$$

Note:

- l_0 = The lowest score of validity test
- c = The highest score of validity test
- r = The score assigned by the rater
- n = the number of rater

In order to attain the high content validity, the research instrument should be validated by seven raters ($n = 7$). The validation by the seven raters adopted the number rating of categories 5. Therefore, the allowed (valid) range for the results of Aiken analysis was 0.82 (Aiken, 1985, p. 137). The results of the Aiken analysis might be consulted in the following sections.

Data Analysis Technique

The data analysis technique that had been adopted within the conduct of the study was the descriptive quantitative technique and the descriptive qualitative technique. Each technique might be elaborated in the following sub-sections.

Descriptive Quantitative Analysis

The quantitative data were attained from the questionnaire, the observation and the documentation. Then, the level on the implementation of the authentic assessment was defined by categorizing the data into five scale. The classification might be consulted in Table 4. This category of assessment was used for calculating the data of the achievement on the implementation of the Natural Science authentic assessment that had been conducted. Next, the raw data was processed into the standardized data by means of Z-score and was altered by means of T-score test so that the data did not yield any negative value.

As having been explained, the evaluation model that had been adopted was the CIPP Model. The CIPP Model was implemented in order to measure the success of the evaluation process. The criteria of success that had served as the reference in the study was the 5-scale criteria. According to (Rosana, 2014), the criteria of assessment for the success of a program might be consulted as follows in Table 4.

Table 4. Score Interval Formula and Assessment Criteria

No.	Score Interval	Assessment Criteria
1.	$X > \bar{X} + 1.5 SDx$	Highly Effective
2.	$\bar{X} + 0.5 SDx < X \leq \bar{X} + 1.5 SDx$	Effective
3.	$\bar{X} - 0.5 SDx < X \leq \bar{X} + 0.5 SDx$	Moderately Effective
4.	$\bar{X} - 1.5 SDx < X \leq \bar{X} - 0.5 SDx$	Ineffective
5.	$X \leq \bar{X} - 1.5 SDx$	Highly Ineffective

Note:

- \bar{X} = Ideal mean score
- SDx = Ideal standard deviation score
- X = Respondents' score

The ideal score standard deviation (SDx) was attained from $1/6$ times the subtraction of the ideal score by the lowest ideal score. Then, \bar{X} was attained from $1/2$ times the addition of the highest ideal score by the lowest ideal score. The maximum ideal score in each instrument was attained when the respondents selected the highest scale on all items, whereas the minimum ideal score in each instrument was attained when the respondents selected the lowest scale on all items.

The evaluation by means of CIPP Model yielded the quantitative data that had been analysed by using the z-score formula (Arikunto, 2016, p. 303). Z-score refers to the number of comparisons that discern the individual score from the mean score and the standard deviation score. The Z-score formula was implemented in the study in order to identify the equation of several instruments that had adopted different scales. In implementing the Z-score formula, the mean score of each class (the mean score of each data instrument) and the standard deviation of the class (the instrument) should be identified.

Table 5. Score Interval Formula and Assessment Criteria

No.	Score Interval	Assessment Criteria
1.	$X > 65$	Highly Effective
2.	$55 < X \leq 65$	Effective
3.	$45 < X \leq 55$	Moderately Effective
4.	$35 < X \leq 45$	Ineffective
5.	$X \leq 35$	Highly Ineffective

The calculation of the T-score and the Z-score

$$Z = \frac{x - \bar{X}}{SD} \tag{4}$$

Note

Z = Standard score

\bar{X} = Mean score of instrument data

X = Instrument indicator data

With the Z-score results that ranged between -3 to 3, these results should be rounded into the interval of 50 by using the T-score formula. As a result, the score from each indicator should be calculated by using the T-score formula. The T-score formula itself refers to the number that implements $M = 50$ and $SD = 10$. The scale of the T-score is calculated by timing the Z-score to 10 (Arikunto, 2016, p. 306). The T-score formula might be elaborated as follows:

$$T = 10Z + 50$$

Descriptive Qualitative Analysis

The data that had been attained from the interview and the interpretation of the T-score calculation should be elaborated further. The results of the data interpretation might be explained qualitative by using the theory analysis proposed by (Miles, Huberman, & Saldaña, 2014, p. 33).

RESULTS AND DISCUSSIONS

In overall, the implementation of the Natural Science authentic assessment within the Regency of Kulon Progo has been moderately effective. The T-score results that have been attained in the distribution of CIPP Model is 52.44 within the range 20.00 – 80.00. The authentic assessment has been conducted comprehensively in order to assess the learning input, the learning process and the learning output.

Based on the results of the teacher questionnaire, the interview and the documentation study with regards to the authentic assessment plan, it is found that the Natural Science teachers throughout the Regency of Kulon Progo have implemented the lesson plan well. The data from the interview with 5 Natural Science teachers show that the Natural Science teachers have designed the assessment instrument as having been contained in the lesson plan that has been prepared. The assessment plan has not been prepared independently but, instead the assessment plan has been prepared collaboratively in either the Subject Teacher Workshop (*MGMP: Musyawarah Guru Mata Pelajaran*) or in other activities. The results of the analysis toward the overall instrument show that the T-score of the teachers is 50.32 and belongs to the “Moderately Effective” category within the range 20.00 – 80.00.

Within the implementation of the Natural Science authentic assessment, the teachers only conduct several skills assessment and knowledge assessment, in addition, the teachers only conduct the attitude assessment once in the whole semester or even in the whole academic year. Consequently, the data from the attitude assessment has not been apparent within the observation sheet. However, the results of the data analysis show that the T-score of the Natural Science authentic assessment implementation is 49.45 and belongs to the “Moderately Effective” category within the range 20.00 – 80.00.

The data analysis of the results toward the assessment plan has been supported by the interview data with regards to the authentic assessment toward 5 teachers. The results of the interview with the 5 teachers indicate that the teachers have designed the assessment plan as having been described in the lesson plan that has been prepared. The assessment plan has not been independently designed; instead,

the assessment plan has been collaboratively designed through the Workshop of Subject Teacher (*MGMP, Musyawarah Guru Mata Pelajaran*) or in the other activities.

Table 6. T-Score Data from the Implementation of Natural Science Authentic Assessment

Model	<i>T-score</i> within range (20.00 – 80.00)				Σ	Conclusions
	Teacher Questionnaire	Student Questionnaire	Observation	Documentation		
Load	4	5	5	4		
C	48.51	-	-	57.77	53.14	Moderately Effective
I	52.94	-	-	48.33	50.64	Moderately Effective
P	48.49	49.67	50.00	-	49.45	Moderately Effective
P	55.62	56.22	-	57.78	56.51	Effective
		Result			52.44	Moderately Effective

Furthermore, the results of the interview with the five teachers show that the teachers have been committed to implement the authentic assessment. However, not all of the teachers who have been interviewed always conduct the daily test for each chapter. Then, based on the results of the observation, it is found that the other forms of authentic assessment that has been implemented are portfolio, practice and journal for each chapter. Therefore, it might be concluded that the teachers have been moderately effective in implementing the Natural Science authentic assessment.

The data on the implementing of the Natural Science authentic assessment within the 2013 Curriculum throughout the Regency of Kulon Progo have been attained by the followins instruments: (1) teacher questionnaire; (2) student questionnaire; and (3) observation. In overall, the authentic assessment that has been implemented is moderately effective with the T-score 45.14. With regards to this finding, it is found that from three sub-aspects within the skills authentic assessment the T-score indicates the effectiveness of one sub-aspect only while the T-score for the remaining two sub-aspects indicates lower level of effectiveness. The reason is that there has not been any teacher who implements the self-assessment and the peer-assessment during the learning process of the chapter “Heat and Its Transfer.” Despite this finding, the knowledge assessment implementation has been moderately effective with the T-score 54.79. In the same time, the implementation of the portfolio and the performance sheet have been moderately effective with the T-score ≥ 50.00 .

The forms of the Natural Science authentic assessment that have been implemented within the Regency of Kulon Progo for the chapter “Heat and Its Transfer” are, namely, performance assessment, portfolio assessment, teacher journal and daily test. In overall, the Natural Science authentic assessment has been moderately effective with the T-score 45.14 in the range 20.00 – 80.00. In addition, the Natural Science authentic assessment has been in accordance to the assessment standards. For measuring the aspects of knowledge, the forms of Natural Science authentic assessment that have been adopted are written test and oral test. Then, for measuring the aspects of skills, the forms of Natural Science authentic assessment that have been adopted are observation, performance assessment, assignment, portfolio and journal. Last but not the least, for measuring the aspects of attitude, the forms of Natural Science authentic assessment that have been adopted are observation, peer-assessment and self-assessment.

The other forms of authentic assessment such as project, self-assessment and peer-assessment should have been implemented but these forms have not been implemented for the chapter “Heat and Its Transfer.” Departing from the results of the interview with the five teachers, the project assessment has not been implemented because the conduct of the project for the chapter “Heat and Its Transfer,” namely designing a solar stove, has been difficulty for the students of Grade VII. The reason is that the glass as the base materials for designing the bowl within the solar stove has been enormously expensive and demanding special expertise. As a result, the T-score for the project assessment has been 32.39 with the category “Highly Ineffective” within the range 20.00 – 80.00. This T-score has been the lowest of all scores. Therefore, it might be implied that the project assessment has been the most highly ineffective form of authentic assessment for the chapter “Heat and Its Transfer.” In overall, it is also found that the project assessment has not been implemented in the five schools that have been sampled.

Table 7. The T-Score Data for the Aspects of Natural Science Authentic Assessment Implementation

Type of Assessment	Sub-Assessment	Teacher Questionnaire	Student Questionnaire	Observation	∑	Status
Attitude Assessment	Self-Assessment	50.60	56.89	35.53	47.46	Moderately Effective
	Peer-Assessment	33.88	36.77	35.53	35.50	Ineffective
	Teacher Journal	37.22	-	46.62	44.50	Ineffective
Knowledge Assessment	Daily Test	47.26	57.89	57.72	54.79	Moderately Effective
Skills Assessment	Performance Assessment	47.26	54.57	58.64	53.93	Moderately Effective
	Project Assessment	35.55	26.71	35.53	32.39	Highly Ineffective
	Portfolio Assessment	49.49	4.69	50.78	50.02	Moderately Effective
Remedial and Enrichment	Enrichment	45.58	55.55	41.08	47.53	Moderately Effective
	Remedial with Different Test Items	37.22	-	41.08	39.36	Ineffective
	Remedial with Same Test Items	40.57	49.61	52.17	47.94	Moderately Effective
Total					4514	Moderately Effective

In addition, the self-assessment and the peer-assessment have been moderately effective and ineffective respectively. The reason is that the Natural Science teachers have only implemented these forms of authentic assessment once throughout the semester or even throughout the academic year. Consequently, the T-score that has been attained for self-assessment has been 47.46 with the category “Moderately Effective” while the T-score that has been attained for peer-assessment has been 35.50 with the category “Highly Ineffective” within the range 20.00 – 80.00. Departing from these findings, it might be concluded that the peer-assessment has been more effective in comparison to the self-assessment.

In accordance to the assessment standards, teachers should guarantee that there shall not be any fraud during the conduct of the daily test or examination since the daily test or examination should be held at the end of every chapter. The reason is that the authentic assessment should be conducted in accordance to the principles of assessment that have been referred. According to the standards of educational assessment, the principles of good assessment cover the following characteristics: valid, objective, fair, integrated, open, ongoing, systematic, accountable, and educative. These principles should always be standards for the conduct of the authentic assessment.

CONCLUSIONS

In overall, the implementation of the Natural Science authentic assessment within the 2013 Curriculum throughout the Regency of Kulon Progo has been moderately effective. This conclusion has been confirmed by the T-score namely 52.44 within the range 20.00 – 80.00. Then, the results of the study with regards to the implementation of the Natural Science authentic assessment have provided the information about the forms of the authentic assessment that have been adopted. The forms of the Natural Science authentic assessment within the 2013 Curriculum that have been implemented throughout the Regency of Kulon Progo consist of project assessment, self-assessment, teacher journal and daily test. Despite this finding, certain forms such as project assessment, self-assessment and peer-assessment have not been implemented for the chapter “Heat and Its Transfer.” These forms have only been implemented once throughout the course of the semester or even the academic year.

Departing from the conclusions of the study, it is apparent that most of the teachers have not mastered the assessment instrument. Therefore, it is suggested that the government should hold training programs for designing the instrument of the Natural Science authentic assessment so that the teachers will be able to understand the essence within the implementation of the Natural Science authentic assessment. In the same time, the government should motivate the teachers with regards to

the conduct of the project assessment, especially in the conduct of the solar stove project, so that the project assessment might also be implemented in the future.

Furthermore, the creativeness of the Natural Science teachers should be improved both in terms of learning activities and in terms of assignments so that the authentic assessment will be effective activities. Since there are too many aspects (attitude, skills and knowledge) that should be observed within the authentic assessment, there should be practical and conductible by one teacher. This practicality and conductivity might be based on the experience of one teacher who has assessed the laboratory practice of the students in Grade VII.

REFERENCES

- 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>
- Ali, L. U., Suastra, I. W., & Sudiarmika, A. (2013). Pengelolaan pembelajaran IPA ditinjau dari hakikat sains pada SMP di Kabupaten Lombok Timur. *Jurnal Pendidikan Dan Pembelajaran IPA Indonesia*, 3(1). Retrieved from http://119.252.161.254/e-journal/index.php/jurnal_ipa/article/view/750
- Andayani, S., & Mardapi, D. (2012). Performance assessment dalam perspektif multiple criteria decision making. *Prosiding Seminar Nasional Penelitian, Pendidikan Dan Penerapan MIPA, Fakultas MIPA, Universitas Negeri Yogyakarta*. Yogyakarta: Universitas Negeri Yogyakarta.
- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (5th ed.). Jakarta: Bumi Aksara.
- Arikunto, S., & Jabar, C. S. A. (2009). *Evaluasi program pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Atkin, J. M., Black, P., & Coffey, J. (2001). *Committee on classroom assessment and the national science education standards (2001) Classroom assessment and the national science education standards*. Washington, DC.: Center for Education National Research Council National
- Azwar, S. (2015). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Basuki, I., & Hariyanto, M. S. (2014). *Asesmen pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Chiappetta, E. L., & Koballa, T. R. (2010). *Science instruction in the middle and secondary schools: Developing fundamental knowledge and skills*. New York: Allyn & Bacon.
- Kunandar, K. (2013). *Penilaian autentik (Penilaian hasil belajar peserta didik berdasarkan Kurikulum 2013)*. Jakarta: Rajawali Pers.
- Levin-Rozalis, M. (2000). Abduction: A logical criterion for programme and project evaluation. *Evaluation*, 6(4), 415–432. <https://doi.org/10.1177/13563890022209406>
- Lukum, A. (2015). Evaluasi program pembelajaran IPA SMP menggunakan model countenance stake. *Jurnal Penelitian Dan Evaluasi Pendidikan/evaluasi Pendidikan*, 19(1), 25–37. <https://doi.org/10.21831/pep.v19i1.4552>
- Menteri Pendidikan dan Kebudayaan Republik Indonesia. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia nomor 104, tahun 2014, tentang penilaian hasil belajar oleh pendidik pada pendidikan dasar dan pendidikan menengah. , Pub. L. No. 104, Peraturan menteri pendidikan dan kebudayaan Republik Indonesia (2014).
- Merilia, S., Fajaruddin, S., & Arbain, A. (2019). An assessment of an English textbook of vocational school. *Arisen (Assessment and Research on Education)*, 1(2). <https://doi.org/10.26486/ARISEN.V1I2.46>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Los Angeles: Sage.
- Murphy, A. A. (1996). *A model for authentic assessment utilizing portfolios*.
- National Research Council. (2001). *Classroom assessment and the national science education standards*. National Academies Press.
- Nezkatgoo, B. (2011). The effects of portfolio assessment on writing of EFL students. *English Language Teaching*, 4(2), 231–241. <https://doi.org/10.5539/elt.v4n2p231>

- Pantiwati, Y. (2013). Authentic assessment for improving cognitive skill, critical-creative thinking and meta-cognitive awareness. *Journal of Education and Practice*, 4(14), 1–9. Retrieved from <https://iiste.org/Journals/index.php/JEP/article/view/6790>
- Rosana, D. (2014). Pendekatan saintifik dalam pembelajaran IPA secara terpadu. *Seminar Nasional Unnes*, 4–21. Semarang: Universitas negeri se.
- Sarapani, E. F., Callejo Perez, D., Gould, J., Hillman, S., & Clark, L. (2014). A global curriculum? Understanding teaching and learning in the United States, Taiwan, India, and Mexico. *SAGE Open*, 4(2), 215824401453640. <https://doi.org/10.1177/2158244014536406>
- Stiggins, R. J. (1994). *Student-centered classroom assessment*. Merrill New York.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Kluwer-Nijhoff.
- Sugiyono. (2010). *Metode penelitian kuantitatif kualitatif dan R & D*. Bandung: Alfabeta.
- Van Blerkom, M. L. (2009). *Measurement and statistics for teachers*. New York, NY: Routledge.
- Widoyoko, E. P. (2012). *Teknik penyusunan instrumen penelitian*. Yogyakarta: Pustaka Pelajar.