

CAN STUDENT ASSESSMENT SHEETS REPLACE OBSERVATION SHEETS?

**Badrun Kartowagiran*, Eka Ary Wibawa, Fitri Alfarisa,
and Dian Normalitasari Purnama**

Universitas Negeri Yogyakarta

*e-mail: kartowagiran@uny.ac.id

Abstract: Observation is argued as the most suitable technique to assess the execution of authentic assessment. Unfortunately, it requires great amount of time and money. We need an alternative. Therefore, the purpose of this study was to develop an instrument in the form of a student assessment sheet on the implementation of authentic assessments in Mathematics subjects. This research is a development research that uses standard procedures for developing instruments. The result of the analysis with Aiken's formula showed that every item of the instrument was in a good category. The result of the analyses using Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Multitrait-Multimethod showed that the instrument had good construct validity. The result of reliability estimation using Cronbach Alpha (α) also showed that the instrument was in the reliable category. Thus, it can be concluded that the instrument in the form of student assessment sheets for assessing the implementation of authentic assessment in junior high school Mathematics learning is highly valid and reliable, which means that the developed instrument can replace the equivalent observation sheet.

Keywords: *assessment sheet development, authentic assessment, mathematics*

DAPATKAH LEMBAR PENILAIAN SISWA MENGGANTIKAN LEMBAR OBSERVASI?

Abstrak: Observasi dianggap sebagai teknik yang paling tepat untuk menilai implementasi asesmen autentik. Sayangnya, teknik ini memerlukan waktu dan biaya yang banyak, sehingga perlu dicarikan alternatifnya. Oleh karena itu, tujuan penelitian ini adalah mengembangkan instrumen yang berbentuk lembar penilaian siswa terhadap pelaksanaan asesmen autentik pada mata pelajaran Matematika. Penelitian ini merupakan penelitian pengembangan yang menggunakan prosedur baku pengembangan instrumen. Hasil analisis dengan formula Aiken menunjukkan bahwa semua butir yang ada pada instrumen termasuk kategori baik. Hasil uji analisis menggunakan *Exploratory Factor Analysis* (EFA), *Confirmatory Factor Analysis* (CFA), dan *Multitrait-Multimethod* (MTMM) menunjukkan bahwa validitas konstruk instrumen termasuk kategori baik. Hasil estimasi reliabilitas menggunakan Cronbach Alpha (α) juga menunjukkan bahwa instrumen tergolong reliabel. Penelitian ini menyimpulkan bahwa instrumen yang berbentuk lembar penilaian siswa terhadap pelaksanaan penilaian autentik di SMP dalam pembelajaran matematika memiliki validitas dan reliabilitas tinggi, yang berarti instrumen yang dikembangkan dapat menggantikan lembar observasi.

Kata Kunci: *pengembangan lembar penilaian, asesmen autentik, matematika*

INTRODUCTION

The quality of education defines the quality of a nation. Better education makes better nation. Today, many indicators show

that the level of education, particularly in Indonesia is still far from being ideal. The Government and every part of the community especially teachers have to put more efforts to improve education quality.

Teachers should stand on the frontline in the effort to make education better. They are essential factors in such an effort. Barber & Mourshed (2012) state that high-performing teachers and headmasters are the starting point of high-achieving students. Furthermore, Barber and Mourshed state that “*student placed with high performing teachers will progress three times as fast as those placed with low performing teachers*”.

Sallis (2002:150) writes that there are ten indicators that define schools' quality and the following is the value of each indicator: (1) access 5%, (2) available services for customers 5%, (3) leadership 15%, (4) physical environment and resources 5%, (5) teaching-learning process 20%, (6) students 15%, (7) staff 15%, (8) external connection 5%, (9) organization 5%, and standards 10%. High-performing human resource working with adequate resources and following correct processes gives a high-performing result. But high-performing human resources following incorrect process –even with abundant resources– will not be able to give optimum result (Massy, 1997: 249). This means that improving the learning process of a school is an essential part of the effort to improve schools' quality; better learning process means better school.

In the effort to improve the quality of the learning process, teachers have many options. One of them is to develop better learning quality assessment. This is just natural. Diranna, Osmundson, Topps, Barakos, Gearhart, Cerwin, ..., Strang (2008) states that instructional goals, teaching models and assessment techniques are linked to one another. For example, if producing graduates with strong characters is the instructional goal, the teaching process shall include trainings and activities that build students' characters and the assessment shall include assessment and description of students' characters.

In line with the above description, Reeves (2010) states that assessment processes and material mastery– which are

included in teaching strategies– are two substantial components in teaching processes. Furthermore Reeves (2010) states that in order to improve the quality of teaching through assessment, teachers must: (1) identify the essential components of the syllabus, (2) develop the performance assessment system (including formulating essays) with rubrics, (3) conduct examination with essay-based content, (4) evaluate the result of the examination using previously prepared rubrics, and (5) review the result of the examination upon evaluating them, including reviewing the competencies that have not been mastered by the students. In the next step, those competencies will serve as the basis for formulating the remedial program. In such a manner, the students have second opportunity to master those competencies.

There are two types of assessments: (1) assessment as a means of improving the capability of teachers in delivering lessons or assessment for learning (AfL) and (2) assessment as a means of improving the capability of students in receiving lessons or assessment as learning (AaL). Both types are the preparatory steps before conducting assessment on the result of the study or assessment of learning (AoL) (Arends & Kilcher, 2010). In principle, assessment must be able to drive teachers to deliver lessons better and also to encourage students to put more effort in their study.

Authentic assessment is the only assessment model that fulfills the above-mentioned principles. It uses the technique of triangulation and triangulation of the source of information and covers all phases of teaching (input, processing, and result). In line with the above argument, the Indonesian Regulation of the Minister of the Education and Culture Number 66 Year 2013 on Assessment Standards states that authentic assessment is a comprehensive assessment method that assesses every teaching phase: input, processing, and output. This method of assessment is

deemed to be comprehensive because it covers assessment on the area of knowledge, skills, and spiritual and social attitude. Frey & Schmitt (2007) argue that authentic assessment aims at measuring the capability of responding to given tasks or tests – which are formulated based on everyday real life problems. Gulikers, Bastiaens, & Kirschner (2004) add that authentic tasks incorporate knowledge, skills, and attitude aspects.

Still in connection with authentic assessment, Tombari & Borich (1999) state that authentic learning and authentic assessment are identification processes on individuals' knowledge, ideas, problem-solving capabilities, social skills and attitude in their daily interaction in their communities, work places and advanced courses. An authentic process assesses every material taught and practiced in the classroom and requires students to apply their skills, knowledge and ability to process things as they are practiced by adults in work place, presented in classroom activities and work book and required in real life. Moreover, Tombari & Borich (1999) mention some characteristics of authentic assessment as follows. (1) It assesses materials taught and practiced in the classroom. (2) It provides real-life-based task as a part of assessment process. (3) It is done continuously. (4) It has standards or criteria. (5) Its assessment condition is the same as that of real-world conditions. (6) It directly assesses students' performance when they are following training or in the process of solving problems.

Authentic teaching and authentic assessment are designed to produce better experience for students, so that they have better performance. Students perform better when they are offered with opportunities to demonstrate what they do and every time they get opportunity will be followed by specific performance improvement. Typical performance exists in the situation of assessment when students are provided with an opportunity to demonstrate the

result of their study with the assumption that they give their best.

Vu & Alba (2014) state that conventionally, assessment is considered authentic when the tasks are real-to-life or have real-life value. Wiggins (1998) states that in order to be authentic, assessment has to be realistic; it requires judgment and innovation, and “asks the student to “do” the subject, that is, to go through the procedures that are typical to the discipline under study”; is conducted in the context mirroring situations in which the skills are best performed; requires students to demonstrate various skills related to complex problems, including decision-making situation; and provides feedback, trainings and second opportunity to solve problem at hand. Some elements of authentic assessment aim not only at assessing competencies, but also at helping students prepare themselves to handle professional world in the future (Raymond, Homer, Smith, & Gray, 2012).

In line with the argument of most experts' in the field, this research defines authentic assessment as real, unpretentious assessment that continuously and sustainably assesses teaching input, process and output covering assessment on knowledge, skills, spiritual and social attitude. This means that full development of the students is only possible when teaching process includes assessment that is authentic. Furthermore, information concerning the authentic assessment implementation in learning processes can be collected through observation. Kartowagiran & Jaedun (2016) showed that observation regarding authentic assessment was replaceable with evaluation on authentic assessment implementation by students.

The challenge is the availability of an instrument in the form of student assessment sheets as a medium for students to assess the implementation of authentic assessment. That is exactly why this research is initiated. Therefore, the purpose of this study was to develop an instrument

in the form of a student assessment sheet on the implementation of authentic assessments on Mathematics subjects. This instrument is expected to replace an equivalent observation sheet.

METHOD

This research is a development research that uses standard procedures for developing instruments which were published by AERA (2014). The procedures are: (1) reviewing theories, (2) developing outline, (3) putting down instrument items, (4) conducting theoretical analysis on instrument items and revision, (5) testing the instrument content validity using expert judgment and then measuring content validity index (V) using Aiken's formula (Aiken, 1985), (6) conducting instrument readability test and revision, (7) conducting the first trial and then instrument construct validity evidentiary test using the Exploratory Factor Analysis (EFA) technique, (8) conducting second trial and then confirming instrument construct validity using the Confirmatory Factor Analysis (CFA) technique, (9) estimating reliability using Cronbach Alpha technique, and (10) determining the construct validity using *multitrait-multimethod* by correlating the data on the evaluation by students and the researcher observation data on the implementation of authentic assessment in the teaching of junior high school mathematics

The assessment grid and items were written by the first author, and analysed theoretically by three co-authors. The result of the theoretical analysis was used to revise the instrument. Furthermore, the validation of the content of the instrument was conducted by assessing the suitability of the instrument items with the indicators by five experts – three experts in educational research and evaluation and two experts in mathematics education. The items which were not very suitable with the indicators was given a score of one and the most suitable was given a score of five. The

data obtained from the five experts were analysed using the Aiken Formula to determine the V value. The items with the V value lower than the critical V value according to the Aiken Table had to be deleted.

The instrument which already had good content validity was tried out to five students who were going to apply the instrument, in order to get the information about its readability or in order to know the statements which could not be understood by the users. After the instrument was revised based on the result of the readability test, the first trial was conducted to 90 junior high school students. The data on the first trial were analysed using the EFA to show the construct validity of the instrument; the item with the loading factor of less than 0.3 had to be omitted (Hair, Ringle, Hult, & Sarstedt, 2014). Furthermore, the second trial was conducted to 150 junior high school students, and the data obtained were analysed using the CFA technique in order to confirm the construct validity of the instrument. The final stage was to measure the construct validity of the instrument using the MTM technique, by correlating the student evaluation data and researcher observation data on the implementation of authentic assessment in the teaching of junior high school mathematics. When the correlation coefficient was higher than 0.8, the student evaluation sheet could replace the observation sheet (Grewal, Cote, & Baumgartner, 2004).

FINDINGS AND DISCUSSION

Findings

The result of this research is an instrument in the form of student assessment sheets on the implementation of authentic assessment in mathematics subject. The instrument was developed based on four components: attitude, skill, knowledge, and teachers' discipline in implementing authentic assessment principles. The discipline in implementing

authentic assessment in this research consists of three principles: the assessment has to be realistic, it has to assess HOTS (Higher Order of Thinking Skills), and it has to be sustainable. The instrument grid can be seen in Table 1.

Table 1. Student Assessment Sheet on the implementation of Authentic Assessment

No.	Indicators	Item Number
1.	Attitude assessment	1,2,3
2.	Skill assessment	4,5,6
3.	Knowledge assessment	7, 8, 9, 10
4.	Discipline authentic assessment	11, 12, 13, 14, 15

Initially, the instrument in the form of student assessment sheets to evaluate the implementation of authentic assessment in mathematics teaching consisted of 15 items. The next was readability test and revision which then followed with review from the experts. The result of the review was then computed using the Aiken Formula. According to Aiken (1985), when the number of raters is five, the number of choices is also five, so that the minimum V

value is 0.80. The result of the analysis using the Aiken Formula showed that items 3, 7, and 13 were in the poor category with 0.73 content validity index (V) and the rest of the items were in the good category. The good items were items 2 and 5 with the V value of 0.80; items 8, 14, and 15 with the V value of 0.87; and items 1, 4, 6, 9, 10, 11, and 12 with the V value of 0.93. The distribution of the valid items resulted from the calculation using the Aiken Formula can be seen in Table 2.

Table 2 shows that after the omission of the items which were not valid, the instrument consisted of only 12 items. Later, the instrument was tried out at the first stage to 90 grade 11 students of 15 junior high schools in Yogyakarta Special Region, who took Mathematics. The data from the first trial were analysed using the EFA technique and the result showed that Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was at 0.743. Every item had anti-image coefficient is greater than 0.5, which means that it satisfied the requirement for the factor analysis.

Table 2. Calculation Result of Aiken V Index

Factor	Number of Item	Index of Aiken V	Information	Number of New Items
Discipline in Implementing Authentic Assessment	Item 1	0.93	Valid	1
	Item 2	0.80	Valid	2
	Item 3	0.73	Not Valid	-
	Item 4	0.93	Valid	3
Knowledge	Item 5	0.80	Valid	4
	Item 6	0.93	Valid	5
	Item 7	0.73	Not Valid	-
	Item 8	0.87	Valid	6
Attitude	Item 9	0.93	Valid	7
	Item 10	0.93	Valid	8
	Item 11	0.93	Valid	9
Skill	Item 12	0.93	Valid	10
	Item 13	0.73	Not Valid	-
	Item 14	0.87	Valid	11
	Item 15	0.87	Valid	12

The result of the first trial showed that 12 items had the loading factor greater than 0.7, which means that they were valid. Since the implementation of authentic assessment had four components, hence the

authentic assessment variants that can be explained using these four components were 65.845%. These components were: attitude, knowledge, skills and discipline in the implementation of authentic

assessment. This is in line with Frey, Schmitt, & Allen's research (2012) who describe context assessment as one of the dimensions of authentic assessment which consists of three aspects: realistic or context activity, performance-based task, and cognitively complex task.

In the second trial, the instrument was administered for 150 students and the analysis used the CFA technique. The goal of the second trial was to confirm the analysis result of the first trial. This is in line with Cramer (2003) who argues that EFA explores theories and CFA tests theories. The result of the confirmatory factor analysis with CFA is shown in Figure 1.

Figure 1 shows that the instrument construct of the implementation of authentic assessment is fit for its purpose. This means that the data supported the concept of student assessment sheets in the evaluation of the authentic assessment implementation in junior high schools; in

short, the instrument is valid. Moreover, the reliability of the assessment sheets was estimated using Cronbach Alpha and the result was at 0.810 which according to Feldt & Brennan (1989), the instrument can be categorized as reliable.

In addition to the factor analysis technique, the construct validity was also verified with *multitrait-multimethod* technique. Campbell & Fiske (1959) introduce this technique and claim that the technique aims at performing verification on the construct validity of an instrument that measures the same traits but measured with two or more different methods. The instruments with good construct validity show high degree of correlation among the measurement results of the same traits but with different methods (Azwar, 2013). In accordance with this, Mardapi (2017) states that when using *multi-trait-multi-method* validity to measure more than one trait, we need to apply more than one method.

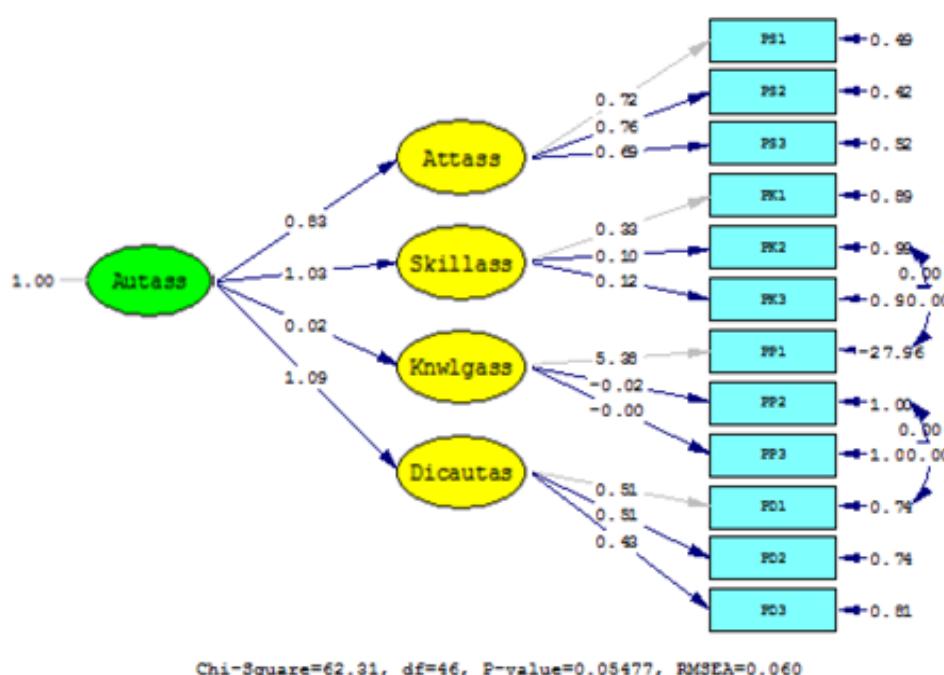


Figure 1. The Result of The Second Trial Analysis With CFA

In this research, to measure the construct validity with multi-trait-multi-method, we correlated the result of the assessment from the students with the result

of the observation on the implementation of authentic assessment. This step is important because both instruments measure the quality of the authentic assessment; the first

instrument was observation sheets used by the researchers and the second one was assessment sheets used by the students. In this research, the assessment sheets are deemed to have the construct validity when the correlation coefficient among assessment results scored at least 0.8 (Grewal, et al., 2004). In this research, the correlation between the result of the assessment and the result of the observation was at 0.965. This shows that there is a very strong correlation between the scores of students' assessment of the authentic assessment implementation and the results of the researchers' observations of authentic assessments. Thus the two methods proved empirically to measure the same trait, namely authentic assessment implementation. This also gives meaning that student assessment sheets can replace the observation sheet.

Figure 2 indicates that the score of each instrument points -both from students' assessment sheets and researchers' observation sheets in the evaluation of the implementation of authentic assessment- are consistent. Most of the items have almost the same scores, so the correlation between them is quite strong. Figure 2 shows that two items (Item 1 and Item 6) score poorly (below standard). Item 1 represents teachers' discipline in assessing students' attitude during the teaching and learning process. Basically, in varying degrees, all teachers have done this job, but most of them do not regularly record students' attitude in their journal or notebook. This is in line with the research by Kartowagiran & Jaedun (2016) which found that 47% of their sample teachers did attitude evaluation. The teachers' reasons for not doing attitude evaluation were: (1) that they could not make the instrument for measuring attitude competence, (2) that the class size was very big, and (3) that measuring attitude was the Counselling and Civic Education teachers' responsibility. Item 6 represents teachers' intensity in asking questions to students during classes.

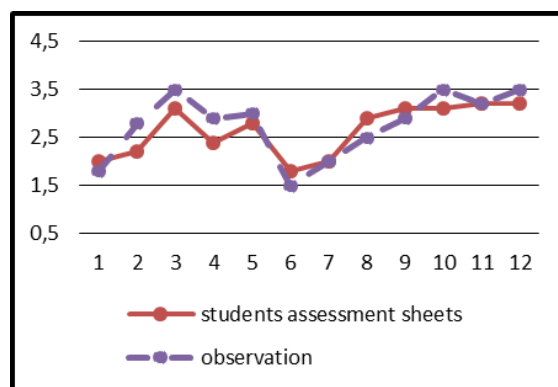


Figure 2. The Correlation between the Result of the Assessment from the Students and the Result of the Researchers' Observation

Discussions

The above-mentioned result shows that the instrument developed in this research has good content validity, construct validity, and reliability. Concerning the capability of the teachers in conducting the attitude assessment, we have to say that it does not look quite good. The low intensity of the assessment on students' attitude is the problem. This is due to the lack of teachers' understanding of how what's written in the Lesson Plan, taught and demonstrated by the teachers in classes affects students' attitude. This is in line with Kartowagiran & Maddini's research (2015) which reported that attitudinal competence developed in classes and demonstrated by the teachers had effects on students' attitude. Besides, in the assessment of the students, teachers have to pay more attention to the manner they communicate with students. Thus, they can improve it. This is also in line with the research of Retnawati, Kartowagiran, Arlinwibowo, & Sulistyaningsih (2017) that showed how the lack of teacher-student communication arose as one of the factors that held up students from achieving the best result in their study.

Item 6 represents teachers' intensity in asking questions to students during classes. Figure 2 shows that the teachers asked questions only once or twice in a

meeting. This means that the intensity was considered low and the teachers did not practice the ability to ask questions. In line with the result of the research by Ermasari, Subagia, & Sudria (2014) which found that there were four factors that hindered the teachers in asking question: the lack of understanding of the types of questions, the lack of planning in formulating and asking questions, the lack of training relevant to formulating and asking questions and the lack of awareness on the challenges the teachers had to deal with. The teachers need to improve their skills in asking questions and drive the intensity of the practice of asking questions. By such manners, the students have a chance to develop the ability to think critically.

Additionally, there were still unrealistic and/or irrelevant questions; the questions made sense mathematically, but not realistically. This type of items is not authentic items (content). In line with Frey, et al. (2012), authentic items (content) have to be composed of realistic and/or relevant questions. Let us return to Figure 2 for a moment. For Items 2, 3 and 4, there are wide gaps between the scores of the students' assessment sheets and those of the researchers' observation sheets. The scores of the researchers' observation sheets are significantly higher. This is reasonable since the observation was only conducted three times in one whole semester and the evaluation from the students was conducted in every class of the semester. Figure 3 below shows the students' assessment on the implementation of authentic assessment by mathematics teachers.

In Figure 3, based on the students' evaluation of the implementation of authentic assessment in the subject of Mathematics, there are two items (Items 1 and 6) that do not give optimum results. Item 1 indicates the low level of discipline in conducting assessment on students' attitude during the process of learning. In connection with the issue of discipline, the teachers confronted a number of challenges that hindered them from performing

authentic assessment optimally. One of them was that the authentic assessment technique required a great deal of time (c.f. Mintah: 2003). Furthermore, Mintah adds that the implementation of authentic assessment with high degree of discipline will deliver positive impact not only on students' development, but also on students' concept of self-development and motivation. Consequently, it is mandatory for teachers to improve the degree of discipline they put on in the assessment of students' attitude.

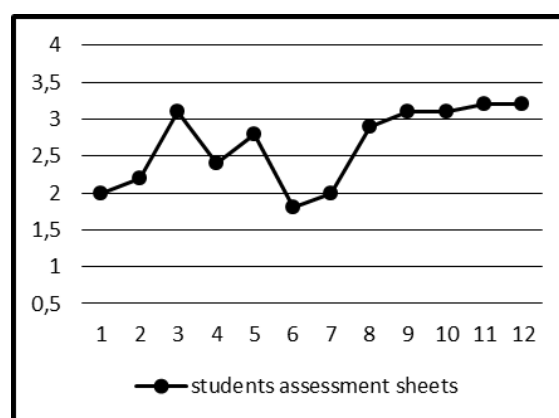


Figure 3. The Students' Evaluation on the Implementation of Authentic Assessment in the Subject of Mathematics

Furthermore, the teacher's questions are only at the second (understand) and third (apply) levels of the Bloom's Taxonomy, and they are not yet at the fourth level (analyse). The learning materials which are tested are not realistic; mathematically, the questions are correct, but they are not applied in the students' everyday life. For example, *Budi lifted a 50-kilogram ball and carried it running for 500 m*, and so on.

Authentic assessment is basically a complex concept. This makes attempting to apply an authentic assessment into practice might be an exhausting task for teachers. It is easy to fall into confusion in the discussion of this concept. It is clear that the concept of authenticity in the description of authentic assessment is significantly deeper than just mere realism

(or being realistic). Most of the publications that we have reviewed focus on class assessment. But some other experts, especially in their early publications, attempted to explore the characteristics of inauthenticity in most of large scale standard test.

Typically, only performance-based assessments or assessments with cognitively complex tasks –that do not put the value of the tasks outside classroom into consideration– are categorized as authentic assessment. It may also be defined that authenticity based on whether any students' arguments, students' team work, or students' involvement in defining scoring criteria are required.

On top of that, relevancy with real world tasks is also a commonly mentioned component of authenticity. Many real world tasks or works are cognitively complex, followed by clear and widely understood criteria of success. It is impossible to think of a real world task that is not performance-based. Obviously, it is improper to assume that the authenticity aspects –that are not focused on in the definitions from the publication–are not included in the real conceptualization of the experts.

Other concepts that potentially add to the teachers' confusion is the description from Frey, et al. (2012). Frey, et al. (2012) states that Oosterhof, Mertler and Popham argue that authentic assessment is a part of performance assessment. On the other hand, he also states that Kubiszyn and Borich, Taylor and Bobbit-Nolen and Airasian argue that performance assessment is a part of authentic assessment. This research stands with the concept expressed later: performance assessment is a part of authentic assessment. Performance assessment only focuses on specific competences, but authentic assessment focuses not only on a single competency, meaning to say it has a broader scope. In authentic assessment, teachers can use journals (teachers' notes), whereas in performance assessment teachers need not use journals. The instrument used for

evaluating performance is merely an observation sheet, and/or evaluation sheet, while that for doing authentic assessment is an observation sheet and/or evaluation sheet which must be accompanied with journal (teachers' notes on students' behavior).

The implementation of authentic assessment in mathematics teaching need to be done because there are many advantages of it. This is in line with Nitko & Brookhart, (2011) who writes that there are some advantages of authentic assessment. It possesses the ability to show students' development based on goals holistically and assessing skills to "do" in the area of knowledge and skills; it provides more meaningful assessment of students for students (Whitelock & Cross, 2012); it encourages students to improve their interest and skills (Svinicki, 2004; Gulikers, Kester, Kirschner, & Bastiaens, 2008); it improves students' confidence, knowledge and skills (Raymond et.al, 2012); it *enhances the integration of what students know and how they act with who they are becoming* (Vu & Alba, 2014). Moreover, authentic assessment also gives students chances to learn by doing and to support teachers in their effort to develop their teaching quality based on students' performance, resulting in more accurate assessment result (Linh, 2016). With these advantages in mind, the implementation of authentic assessment is beneficial to both students and teachers.

Meanwhile, Hargreaves, Earl & Schmidt (2002) state that authentic assessment encourages students to be more responsible for their study, produce assessment as an integral part of learning process, be more creative and implement – and not only memorize what they have learned. Furthermore, Hargreaves, et al. (2002) found that: (1) teachers were more comfortable with authentic assessment because they did not need to test examination content first; (2) authentic assessment was effective in building common collaborative understanding among teachers, students and parents

because authentic assessment assessed every students' activities and involved parents in many occasions; and (3) authentic assessment provided better feedbacks for teachers.

The advantages of applying authentic assessment in teaching are so many that it is logical that in the Curriculum 2013 used by Indonesian teachers, for example, who are obliged to apply authentic assessment. Nevertheless, it must be noted that there are still many teachers who cannot apply authentic assessment well. Such teachers need to be trained to improve their ability to apply authentic assessment. In order to make them serious in applying authentic assessment, evaluation needs to be done. The evaluation is done by the school principal who is helped by students using the developed assessment sheet.

CONCLUSION

The students' assessment sheets on the implementation of authentic assessment as an instrument was developed in the following procedures: (1) reviewing theories, (2) developing outline and writing down the points of the instrument, (3) analyzing the points of the instrument and conducting revision, (4) conducting trials and defining the characteristics of the instrument, (5) finalizing the instrument, (6) conducting instrument readability test and revision, (7) conducting the first trial and then instrument construct validity evidentiary test using the EFA technique, (8) conducting the second trial and then confirming instrument construct validity using the CFA technique, (9) estimating the instrument reliability by using Cronbach Alpha formula, and (10) confirming the construct validity using the multitrait-multimethod. It could be concluded that students' assessment sheets as an instrument in the assessment of the implementation of authentic assessment in junior high school Mathematics teaching has a high degree of validity and reliability, which means that the developed instrument can replace an equivalent observation sheet.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Research, Technology, and Higher Education of Indonesia for their financial support to this research.

REFERENCES

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement*, 45(1), 131-142.
doi:10.1177/0013164485451012.
- Arends, R. I., & Kilcher, A. (2010). *Teaching for student learning becoming an accomplished teacher*. New York, NY: Routledge.
- Azwar, S. (2013). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Barber, M. & Mourshed, M. (2012). *Profesional development international*. New York, NY: Pearson.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
doi:10.1037/h0046016.
- Cramer, D. (2003) *Advanced quantitative data analysis*. London: McGraw-Hill Education.
- Diranna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., ..., Strang, C. (2008). *Assessment-centered teaching (A reflective practice)*. London: Sage.
- Ermasari, G., Subagia, I. W., & Sudria, I. B. N. (2014). Kemampuan bertanya guru IPA dalam pengelolaan pembelajaran. *Jurnal Pendidikan dan Pembelajaran IPA Indonesia*, 4(1), 1-

12. Retrieved from http://oldpasca.undiksha.ac.id/e-journal/index.php/jurnal_ipa/article/view/1111.
- Feldt, L. S., & Brennan, R. (1989). Reliability. In R. L. Linn (Ed), *Educational measurement* (3rd ed.). New York, NY: Macmillan.
- Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, 18(3), 402-423. doi:10.4219/jaa-2007-495.
- Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical assessment, research & evaluation*, 17(2), 1-18. Retrieved from <https://pareonline.net/pdf/v17n2.pdf>.
- Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: implications for theory testing. *Marketing Science*, 23(4), 519-529. doi:10.1287/mksc.1040.0070
- Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational technology research and development*, 52(3), 67-86. Retrieved from <https://eric.ed.gov/?id=EJ732658>.
- Gulikers, J. T., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction*, 18(2), 172-186. doi:10.1016/j.learninstruc.2007.02.012.
- Hair, J. F., Ringle, C. M., Hult, T., & Sarstedt, M. (2014). *A primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks: Sage.
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95. doi:10.3102/00028312039001069.
- Kartowagiran, B., & Maddini, H. (2015). Evaluation model for islamic education learning in junior high school and its significance to students' behaviours. *American Journal of Educational Research*, 3(8), 990-995. doi:10.12691/education-3-8-7.
- Kartowagiran, B., & Jaedun, A. (2016). Model asesmen autentik untuk menilai hasil belajar siswa Sekolah Menengah Pertama (SMP): Implementasi asesmen autentik di SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(2), 131-141. doi:10.21831/pep.v20i2.10063.
- Linh, N. N. (2016, August). *Authentic assessment: A case study of its implementation in a lecturer's classes in Vietnam*. Paper presented at the International Conference on Education and Social Integration, Ho Chi Minh City, Vietnam.
- Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan, edisi kedua*. Yogyakarta: Parama Publishing.
- Massy, W. (1997). Teaching and learning quality-process review: The Hong Kong programme. *Quality in Higher Education*, 3(3), 249-262. doi:10.1080/1353832970030305.
- Mintah, J. K. (2003). Authentic assessment in physical education: Prevalence of

- use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in physical education and exercise science*, 7(3), 161-174. doi: 10.1207/S15327841MPPE0703_03.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of student*. Boston, MA: Pearson.
- Raymond, J. E., Homer, C. S. E., Smith, R., & Gray, J. E. (2012). Learning through authentic assessment: An evaluation of a new development in the undergraduate midwifery curriculum. *Nurse Education in Practice*, 13(5), 471-476. doi:10.1016/j.nepr.2012.10.006.
- Reeves, D.B. (2010). *Transforming profesional development into student result*. Alexandria: ASCD.
- Retnawati, H., Kartowagiran, B., Arlinwibowo, J., & Sulistyarningsih, E. (2017). Why are the mathematics national examination items difficult and what is teachers' strategy to overcome it? *International Journal of Instruction*, 10(3), 257-276. doi: 10.12973/iji.2017.10317a.
- RoI, Ministerial Regulation of Education and Culture 2013 No. 66, Standar Penilaian [Assessment Standard].
- Sallis, E. (2002). *Total quality management in education*. London: Routledge.
- Surya, A., & Aman, A. (2016). Developing formative authentic assessments based on learning trajectory for elementary school. *Research and Evaluation in Education*, 2(1), 13-24. doi: 10.21831/reid.v2i1.6540.
- Svinicki, M. D. (2004). Authentic assessment: Testing in reality. *New Directions for Teaching and Learning*, 100(Winter 2004), 23-29. doi: 10.1002/tl.167.
- Tombari, M.L & Borich, G.D. (1999). *Authentic Assessment in the classroom (application and practice)*. Upper Saddle River, NJ: Prentice Hall.
- Vu, T. T., & Alba, G. D. (2014). Authentic assessment for student learning: An ontological conceptualisation. *Educational Philosophy and Theory*, 46(7), 778-791. doi:10.1080/00131857.2013.795110.
- Whitelock, D., & Cross, S. (2012). Authentic assessment: What does it mean and how is it instantiated by a group of distance learning academics? *International Journal of e-Assessment*, 2(1), article 9. Retrieved from <http://journals.sfu.ca/ijea/index.php/journal/article/view/31>.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.