JERE 7 (2) (2018) 154 – 162

Journal of Educational Research and Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Development of Authentic Assessment Instrument For Performance in Learning Mathematics in Linear Program

Zainul Arifin✉, **Kartono, Supriyadi**

Universitas Negeri Semarang, Indonesia

Article Info

Article History:
Accepted 17 May 2018
Approved 13 August 2018
Published 24 August 2018

Keyword:
Authentic assessment, Performance, Linear program

Abstract

Problems often faced by teachers of mathematics, namely the validity and reliability of measuring instruments assessment of performance, preparation of test performance is still very limited knowledge and understanding of teachers, the assessment results are often influenced by the objectivity of the teacher as a rater because in doing the assessment is done without the involvement of other teachers as collaborator. This study aims to examine the level of validity, reliability, and the practicality of authentic assessment instruments for performance in linear mathematics learning programs. The instrument is useful for teachers to innovate in the world of education and for students as study material to determine the level of understanding of students' abilities. This research method uses research and development techniques. The sample was used for a study of 34 students for small-scale tests and 44 students for large-scale tests. The results showed that the instrument was declared to be valid in the contents of 0.83. ICC reliability estimation between assessors 0.730. The scale test shows the value of KMO 0.548 and Barteletts sig test. 0,000. But there is one item that has an MSA value <0.5. Large scale test shows the value of KMO 0,532 and Bartlett's Test sig. 0,000. All items obtained MSA value> 0.5. The instrument forms 2 components. The value of Eigenvalues component 1 is 2, 007 and component 2 is 1.183. The estimated reliability of the instrument on a small scale test was 0.674 and a large scale of 0.661. Practical test gets a score of 155. The conclusion of this study is that the instrument developed has proven to be valid, reliable and practical. Very practical proven instruments based on teacher ratings. The suggestion of this research is for teachers to use this instrument in learning linear mathematics programs to produce easy and precise measurements.

✉Correspondence Address:
Kampus Pascasarjana UnnesJalan Kelud Utara III Semarang 50237, Indonesia
E-mail: azainul630@gmil.com

P-ISSN 2252-6544
E-ISSN 2502-356X

Journal of Educational Research and Evaluation 7 (2) (2018) 154 -162

INTRODUCTION

A session is carried out as an effort to measure the level of achievement of indicators of learning and gather information on students' learning progress in various aspects (Astuti, Prasetyo, & Rahayu, 2012). Mahendra (2016) said the assessment has a very important role and benefits teachers and students.

A session is a series of activities to obtain, analyze and interpret data about students' learning processes and results that are carried out systematically and continuously, so that it becomes meaningful information in decision making (Widowati, Aminah, & Cari, 2016).

Various efforts have been taken by the government to improve the quality of education, one of which is the 2013 curriculum policy aimed at preparing Indonesian people to have the ability to live as individuals and citizens who are faithful, productive, creative, innovative, affective and able to contribute to the life of the nation, state, and perarkan world (Kustitik & Hadi, 2016).

The use of authentic assessments is now a necessity, given the pillars of education echoed by UNESCO not only learning to know but also for the skills to use what is learned (learning to do); achieve self-actualization in the real world (learning to be) and be able to become part of a harmonious society (learning to live tougether) , the four pillars of education then become a reference for curriculum development in formal education institutions or schools throughout the world (Marhaeni & Artini, 2015) .

A performance segment is needed to measure other aspects beyond cognitive, namely seven basic abilities are (1) visual-spatial , (2) bodilykinesthetic , (3) musical-thythmical , (4) interpersonal , (5) Intrapersonal , (6) logicamthematical , (7) verbal linguistic, only the last two abilities (logicamthematical and verbal linguistics) are measured and assessed by many people, while the other five abilities have not been widely revealed. From this statement it is clear that the assessment process, especially performance assessment is the main focus of assessment (Majid, 2014, pp. 58-59).

Sumardi (2017) said that almost all mathematics teachers at junior high school had not conducted authentic assessments due to several obstacles. The teacher does not carry out authentic assessments because: (1) there is no time if each meeting is held an assessment, (2) the number of students is too much, both in each class (class size) and the number of students each teacher must teach, other obstacles (3) most teachers do not understand how to make instruments and how to conduct assessments (Kartowarigan & Jaedun, 2016). A authentic session cannot be carried out as a whole, this is reinforced by only 6% of lecturers who conduct authentic assessments, and 63% of lecturers do not understand authentic assessment well (Afrida, 2016).

According to Elliott, 1995 there are two main concepts that describe performance assessment: "1) Performance: Students active generation of responses that can be observed either directly or indirectly through permanent products, 2) Authentic: The nature of the task and the context in which the assessment occur relevant and represent "the real world or problem" (Dikli, 2003), if the performance assessment is carried out on a number of students not designed first and carried out carelessly the results cannot be justified because they are inconsistent, thus the teacher may be unfair to a number of students in assessing their performance.

Wiggins (2005: 2-3) mentions that designing and carrying out performance assessments is very efficient, because it is smooth or consistent (read reliably) (Majid, 2014: 59), but there are also weaknesses in performance assessments, namely: (1) performance assessments spend a lot of time compiling performance tasks in the field the same, (2) requires a relatively expensive fee, (3) performance assessment also requires a long time in assessing and giving a score. To overcome the shared weaknesses that must be

Journal of Educational Research and Evaluation 7 (2) (2018) 154 -162

done are: (1) performance assessment must be supported by a clear rubric, which must be understood by teachers and students; (2) the performance criteria must be used appropriately and consistently by the teacher and students, and (3) provide feedback to all participants assessed. (Utsman, 2013)

The rubric can help the supervisor to determine the level of achievement of the expected performance, by communicating the rubric to the respondent or even by composing a rubric together between the supervisor and respondent, the rubric is also expected to be a motivator or motivator for respondents in improving their performance, because by using assessment Learning outcomes based on the will of the determined rubric will obtain information on student learning outcomes that are accurate, fair and objective (Sutama et al, 2015)

Based on the results of observations and initial interviews in the field (especially towards mathematics learning at Bhakti Nusantara Mranggen Vocational School) there are several obstacles faced by teachers in assessing student performance, namely: first the guidelines for scoring in unclear instruments so difficult to use, the components assessed difficult to observe, so it tends to be ignored; secondly, the assessor (rater) is generally only one person, namely the teacher in the field of study, while the components assessed and the number of students who are assessed are quite large, so that it is difficult to get a comparison to be taken into consideration; third, there may be a tendency to give a high score or vice versa, this is caused by the instrument used does not meet the requirements of validity, reliability and practicality.

Based on the background above, it is necessary to develop an authentic assessment instrument for participants in mathematics learning. This study aims to examine the level of validity, reliability, and the level of practicality of authentic assessment instruments for performance in learning mathematics in linear programs.

METHODS

This study uses research and development techniques to produce products (Sukmadinata, 2007: 169-170). The research procedure based on Borg & Gall's ten steps of research and development was adapted into three phases of activities namely, (1) introduction, (2) product development, and (3) presentation.

The source of the research data came from Vocational Mathematics Teachers, Grade XI Vocational Students, and experts. The sample in this study was taken randomly from the 11th grade students of Accounting Study Program taken by 34 students for small scale tests and 44 students for large scale tests.

The technique used to collect data is interview techniques using interview guidelines, observation techniques using observation sheets, questionnaires, and tests. Data validity testing techniques are using source triangulation techniques. Validity test consists of testing the validity of the content by 3 experts with the criteria declared valid if the validity coefficient ≥ 0.30 means that the item can be said to be valid (Azwar, 2014: 143).

Aiken's V

$$\mathbf{V} = \sum \mathrm{s/n[(c-1)]}$$

V = Rater agreement index regarding item validity
s = r - lo
lo = The lowest validity rating (in this case = 1)
c = The highest validity rate (in this case = 5)
r = Figures given by an appraiser

The construct validity is carried out by exploratory factor test to determine the correlation between variables; eigenvalue is used to calculate the percentage of variance that is explained, as well as drawing the screeplot (Retnawati, 2016: 23). Reliability testing is divided into two, namely reliability testing of experts using Inter Class Correlation and empirical reliability testing using Cronbach

Journal of Educational Research and Evaluation 7 (2) (2018) 154 -162

Alpha formula on small scale tests and large scale tests. The results of data analysis were analyzed qualitatively and quantitatively using IBM SPSS version 24 software.

RESULTS AND DISCUSSION

The researcher found that the assessment conducted by the teacher on the mathematics subject of linear program material was still general in nature that was only assessing the final results without assessing in detail each student's performance. The instruments used in assessing learning do not yet have clear scoring guidelines and are tested for validity, reliability and practicality.

Development of assessment instruments authentic show work on learning linear program mathematics has special specifications. The instrument consists of Grids, Taks, Rubrics, and assessment sheets. The grid is adjusted to the Basic Competencies of Mathematics class XI, namely 4.2 Resolving contextual problems related to linear programs of two variables. Grid consists of on one the core competence of the KI 4 m engolah, reasoning, and menyaji in the realm of the abstract realm kongkerit and related to the development of learned in school independently, to act effectively and creatively as well as being able to use the method according to the rules of science.

The assessment uses a rating scale of 1 to 4, with the highest scale of 4 which means that students show the maximum expected ability. Researchers develop an instrument consisting of 4 items that will be accumulated to produce the final value, namely the formula **N =js/16 x 100** JS JS is the number of scores obtained and 16 is the highest score of all items. The four questions in question are as follows:

1. Draw a graph of the set of solutions for the linear inequality of the following two variables! $2X + 3Y \leq 6$ for X and Y $\in$ R
2. Determine the set region of completion of the linear inequality of the following two variables! $4X - 3Y < 12$, for X and Y $\in$ R
3. Draw a graph of the settlement set of the linear inequality system with the following two variables. $X \geq 0$, $Y \geq 0$, $2X + Y \leq 6$, $X + 2Y \leq 9$, for X and Y $\in$ R
4. A small industry produces two types of goods (item A and item B) using two machines (machines M_1 and machine M_2). One unit of item A is made by operating the machine M_1 for 2 minutes and the machine M_2 for 4 minutes, while one unit of goods B is made by operating the machine M_1 for 8 minutes and the machine M_2 for 4 minutes. In one day machine M_1 and machine M_2 operate for more than 8 hours. Net profit obtained from one unit of item A is IDR 250.00 and one unit of item B is IDR 500.00. Make a mathematical model of the linear program problem above if the net profit is expected to reach the maximum. Then draw a graph of the mathematical model.

Content Validity

The experts too give questionnaire containing conclusion expert judgment to instrument assessment authentic show work on learning Linear program mathematics. Aiken index V's instrument in whole worth 0.83, so the instrument is whole stated valid fill in by experts shown on Table 1.

Table 1. Coefficient Expert Agreement

No. Item	Aiken's V index	Conclusion	No. Item	Aiken's V index	Conclusion
1	0.83	Valid	7	1.00	Valid
2	1.00	Valid	8	0.67	Valid
3	0.75	Valid	9	1.00	Valid
4	0.83	Valid	10	0.83	Valid
5	0.83	Valid	11	0.67	Valid
6	0.67	Valid			

Validity Construct

The results of draft analysis instrument 2 on a small scale test obtained KMO value > 0.5 that is equal to 0.548, then the requirements of sample adequacy are fulfilled to be further analyzed. The Barteletts Test test shows sig. < 0.5 so that there is a correlation between variables so that they can be further analyzed. The results of the analysis continued with

Journal of Educational Research and Evaluation 7 (2) (2018) 154 -162

looking at the anti image matrices table found 1 item which shows a value of < 0.5, which does not meet the criteria and cannot be analyzed further. The item that has a correlation value of <0.5 is B1 of 0.401.

The test results for instrument 2 show that there is 1 item that is measured invalid so that the researcher makes a revision or improvement then produces draft instrument 3 which will be carried out a large scale test. The results of the analysis can be seen in Table 2.

Table 2. Results Test Appropriateness Instrument on Test Scale Big

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.532
Bartlett's Test of Sphericity	Approx. Square	Chi-65,531
	Df	6
	Sig.	.000

When the analysis of instrument 3 in the large scale test is obtained the value of KMO > 0.5 is equal to 0.532, then the requirements of sample adequacy are fulfilled to be further analyzed. The Barteletts Test test shows sig. < 0.5 so that there is a correlation between variables so that they can be further analyzed. The results of the analysis continued by looking at the anti image matrices table found all valid items. The results of the analysis can be seen in Table 3.

Table 3. Results of Anti Image Correlation

No.	Item	Anti Image Correlation
1	Determine settlement something linear linear inequality variable	0.523
2	Determine settlement something linear linear inequality variable	0.682
3	Complete linear program problem two variable	0.519
4	Complete linear program problem two variable	0.520

Step next that is look many possible factors formed on factor analysis with amount a sample of 44 participants student on test scale big. After do exploratory factor analysis with help with the IBM SPSS version 24 program. Next is results The Total Variance table is explored in Table 4.

Table 4. Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%
1	2007	50,173	50,173	2007	50,173	50,173
2	1,183	23,564	79,736	1,183	23,564	79,736
3	.687	17,185	96,921			
4	.123	3,079	100,000			

Two components that are formed and could represent the number of indexors. 4 items in analysis evidently have eigenvalues> 1 means that 4 items item it can grouped to 2 factors. Factor 1 has value of 2,007 and able to explain variance amounting to 50,173%. Factor 2 has value of 1.183 and able to explain variance amounting to 23,563. For determine how many components / factors used to get it explain total diversity then seen from big eigenvalues value, the component that eigenvalues> 1 is components used. in whole Results of Total Variance could described in Figure 4.1.

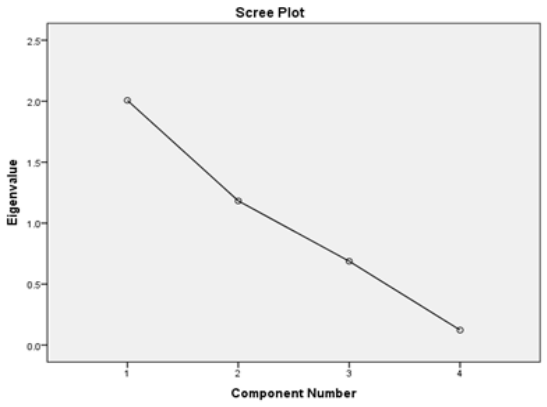

Figure 1. Scree Test Plot Scale Big

Based on the Scree Plot above seen that there are 2 points above eigenvalue is 1 and the other points are below value 1. This is describing

that there are 2 components that have the eigenvalue above value 1. Next do determination each item will enter into which factor from second existing factors. Grouping item and the size loading factor from one factor seen from the value of the loading factor that has value > 0.3. Grouping item into factors can do with look Rotated Component Matrix Table. Following is the results of Rotated Component Matrix on Table 5 .

Table 5. Rotated Component Matrix

Item	Component 1	2
Determine settlement something linear linear inequality variable	-.069	.843
Determine settlement something linear linear inequality variable	.231	.749
Complete linear program problem two variable	.961	.077
Complete linear program problem two variable	.960	.087

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. [a]
a. Rotation converged in 3 iterations.

The rubric from the rotation factor seen that grouping the indicator index into factors and the amount of loading that is obtained seen on table above. Seen that determination of input input indicator to factors certain follow on big correlation between variables with factor that is to the correlation big. With thereby then factors that are formed with a item presented on Table 6.

Table 6 . Grouping Item in Something Factor

Item No.	Item Instrument	Factor formed	Value correlation	Factor name
3	Complete linear program problem two variables.	1	.961	Linear program problems are two variables
4	Complete linear program problem two variables.		.960	
1	Determine settlement something linear linear inequality variable.	2	.843	Completion of a non-equal two variables
2	Determine settlement something linear linear inequality variable.		.749	

The grouping of 4 items becomes 2 factors that are formed and naming each factor. The components formed in factor 1 are named linear program variables two variables, the instrument items consist of items 3 and 4. The components formed in factor 2 are named the settlement of a non-equal two variables, the instrument consists of items 1 and 2.

Reliability

Test reliability do on test reliability assessment expert, test scale small and test scale big . E stimulation reliability assessment expert using an ICC of 0.730 shows experts corresponding in assess assessment instruments assessment authentic show work on learning linear mathematics program. Estimation reliability test scale small amounting to 0.674 more big from 0.5 so could interpreted that a reliable instrument. Results test scale big show coefficient reliability test scale big amounting to 0.661 more big from 0.5, so could interpreted that a reliable instrument seen on Table 7 .

Journal of Educational Research and Evaluation 7 (2) (2018) 154 -162

Table 7 . Data Analysis Reliability Test Small Scale

Test	Cronbach's Alpha	Intraclass Correlation Coefficient	N of items
ICC Expert	0.890	0.730	4
Small Scale	0.674		4
Large Scale	0.661		4

Practicality

The test of the practicality of the design of assessment instruments authentic as follows: a) assessor 1 gets a score of 52 in the category "Very Practical"; b) assessor 2 gets a score of 52 in the category "Very Practical"; c) assessor 3 gets a score of 51 in the category "Very Practical". The mean score of the three assessors was obtained at 51.6, so that in general the authentic assessment instrument for performance in learning mathematics was linear in the category "Very Practical".

Development of authentic assessment instruments performance on learning mathematics linear programs carried out several tests to produce good instruments. The instrument is carried out by validity analysis based on expert opinion. This is consistent with Azwar's statement that content validity is the validity estimated by testing the feasibility or relevance of the test content through rational analysis by a competent panel or through expert judgment. Test the content validity to see the suitability of the contents of the instrument carried out by 3 experts / validators.

The results of expert opinion were analyzed using Aiken's formula V. authentic assessment instruments performance on linear mathematics learning program consisting of 4 items obtained value> 0.3 that is 0.83 representing all items, so that it can be concluded that authentic assessment instruments Mathematics learning is a linear program relevant and can be tested on tests in the field. This is in accordance with the criteria stated by Azwar (2014: 34) that the validity coefficients greater than or equal to 0.3 can be said to be adequate or valid. However, if the validity coefficient is smaller than o, then the item is declared inadequate or invalid.

The results of the content validity test were continued by estimating the reliability of the expert agreement using the Inter Class Correlation (ICC) test. The estimated reliability of the content based on the expert test obtained an ICC coefficient of 0.730. ICC coefficient value is greater than 0.7 so that it can be concluded that the three experts are consistent in assessing instruments and categories of High category. This is in line with Stainer and Norman (2000) that the measuring instrument has adequate stability if the inter-gauge ICC> 0.50 is of high stability if the ICC is between the gauges ≥ 0.80.

Authentic assessment instruments for performance on linear mathematics learning programs have proven to be valid. Instrument appraisers have done a good assessment as evidenced by the consistency of the three-year assessment. Based on the content validity and reliability through expert agreement it can be stated that the instrument developed has fulfilled the requirements of validity and reliability so that the instrument can be continued to be carried out for small and large scale tests.

Authentic assessment instruments for performance in linear mathematics learning programs followed by small scale tests. Small-scale tests were carried out involving 34 samples, which obtained sample adequacy requirements by looking at KMO values of 0.548 and Bartelett's test obtained by sig. smaller than 0.05 so that data can be continued for further analysis. The results of the instrument test on the table anti image correlation found one item that obtained an MSA value below 0.5 yng that did not meet the requirements so it could not be further analyzed because the correlation value <0.5. Items that have a value below 0.5 are item B1 which has an MSA value of 0.401. So the analysis cannot continue. Researchers looked at the results of instrument reliability on small scale tests. Tests are carried out based on the results of small usage tests by obtaining an estimation of instrument reliability of 0.674. The condition is

said to be reliable is more than 0.5, so the researchers make improvements to the instrument to get the MSA value for point 1 greater than 0.5 and get a higher reliability value to produce reliable instruments.

The researchers went on to try out the revised instrument for a wider test on large-scale tests. Large-scale tests involved 44 respondents. The large-scale test results were explored by exploratory factors and obtained a KMO value of 0.532 which met the sample adequacy requirements of more than 0.5 and Barttelet's test showed sig. smaller than 0.05 so the analysis can be continued by looking at the value of the MSA item. The MSA value for four items is entirely more than 0.5 so it can be continued to see the Total variance explained which shows there are two components that have eigenvalues more than 1. H this is in accordance with Azwar (2014 : 143) that if the KMO value is ≥ 0 , 5, Anti image Correlation ≥ 0.5, Eigenvalues ≥ 1 and Loading factor ≥ 0.3 then factor analysis can be continued.

Based on the plot screen, it can be seen that there are 2 points in a bag of value 1 and the other points are below the value of 1. This shows that the instrument forms 2 components which have an eigenvalue above the value of 1. The names of the factors formed including Solving linear program problems and a two-variable linear inequality system.

The instrument performance assessment on linear mathematics learning program on a large scale test followed by testing reliability. The estimation of the reliability of the instrument on a large scale test can show consistency in measuring instruments. The instrument test results were tested using Alpha Cronbach relability test obtained reliability coefficient of 0.661. This is in line with what Naga stated that the reliability coefficient of 0.5 and above is sufficient to be accepted as good reliability (Khumaedi, 2012: 13).

Authentic assessment instruments performance on linear mathematics learning program seen practicality based on practicality test. Practical test involves 3 practitioners / teachers who directly know the implementation of the instrument test. This is in line with research conducted by Setiawan (2017: 8) practicality tests carried out with two techniques, namely field trials and assessments from teachers.

Value practicality very tin g gi indicates that worthy instrument used to measure basic competencies resolve the contextual problems associated with two variable linear programs. The instrument is easy to use, easy to understand, easy to get value results and easy to document the measurement results. Authentic assessment instruments performance on learning Mathematics linear programs developed are stated to be very practical and feasible in terms of users, namely teachers to use to measure Basic Competencies Resolving contextual problems related to linear programs of two variables.

CONCLUSSION

Based on the results of the research and discussion that has been carried out, it can be concluded as follows. Valid instrument fill in based on expert judgment. Instrument declared valid construct and form two fa c tors. The instrument proven consistent in assessment based on test reliability. Instrument could with easy used by the teacher inside learning because to be on category very practical.

SUGGESTION

Authentic assessment instruments for performance in linear mathematics learning programs that have been developed based on development steps by having valid and reliable criteria. The instrument can be used by the researcher then the performance is developed again. The instrument has been developed in can assist teachers in assessing the performance of the linear program that can be used as a reference in the manufacture of instrument performance in learning Mathematics.

REFERENCES

Afrida, R. A. (2016). Pengembangan Model Penilaian Otentik untuk Mengukur Capaian Pembelajaran Mahasiswa. *Jurnal Biologi Dan Pembelajaran Biologi*, *1*(2), 137–147.

Astuti, W. P., Prasetyo, A. P. B., & Rahayu, E. S. (2012). Pengembangan Instrumen Asesmen Autentik Berbasis Literasi Saint pada Materi Sistem Ekskresi. *Lembaga Ilmu Kependidikan*, *41*(1), 39–43.

Azwar, S. *2014. Reliabilitas dan Validitas (Edisi IV).* Yogyakarta: Pustaka Belajar Dikli, S. 2003. Assesment at a Distance: Traditional vs Alternative Assesment. *The Turkish Online Journal of Educational Technology*-TOJET, 2(3): 13-19

Kartowarigan, B., & Jaedun, A. (2016). Model Asesmen Autentik untuk Menilai Hasil Belajar Siswa SMP:Implementasi Asesmen Autentik di SMP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *20*(2), 131–141.

Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan

Kustitik, & Hadi. (2016). Pengembangan Perangkat Penilaian Autentik Mata Pelajaran Prakarya dan Kewirausahaan di SMK. *Jurnal Pendidikan Vokasi*, *6*(2), 184–197.

Mahendra, E. I. (2016). Contextual Learning Approach and Performance Assesment in Mathematics Learning. *International Research Journal of Management, IT and Social Science*, *3*(3), 11–26.

Majid, A. (2014). *Penilaian Autentik Proses dan Hasil Belajar*. Bandung: Remaja Rosdakarya.

Marhaeni, A. A. I. ., & Artini, L. P. (2015). Asesmen Autentik dan Pendidikan Bermakna: Implementasi Kurikulum 2013. *Jurnal Pendidikan Indonesia*, *4*(1), 499–511.

Retnawati, H. 2015. "Hambatan Guru Matematika Sekolah Menengah Pertama dalam Menerapkan Kurikulum baru"*, Cakrawala Pendidikan,* 34 (3): 390-403

Setiawan, H. (2017). Pengembangan instrumen asesmen autentik kompetensi pada ranah keterampilan untuk pembelajaran tematik di sekolah dasar. *Jurnal Pendidikan*, *2*(7), 874–882.

Sukmadinata, N. S. 2007. *Metode Penelitian Pendidikan*. Bandung: Remaja Rosdakarya

Sumardi. (2017). Penilaian Autentik dalam Pembelajaran Matematika Berbasis Lesson Study di SMP Muhammadiyah Kabupaten Sukoharjo Jawa Tengah. In *The 5Th Urecol Proceeding*.

Sutama, Narimo, S. & Samino. 2015. "Management of Curriculum 2013 Mathemathic Learning Evaluation In Junior High School", *International Journal of Education,* 7(3): 164-174

Utsman. 2014. "Penilaian otentik berbasis kurikulum 2013": *Seminar Nasional Evaluasi Pendidikan tahun 2014.*

Widowati, T., Aminah, S. N., & Cari. (2016). Pengembangan Instrumen Penilaian Otentik Berbasis Scientific Literacy pada Pembelajaran Fisika di SMA sebagai Implementasi Kurikulum 2013. *Jurnal Inkuiri*, *5*(2), 8–19.