DEVELOPING FORMATIVE AUTHENTIC ASSESSMENT INSTRUMENTS BASED ON LEARNING TRAJECTORY FOR ELEMENTARY SCHOOL

[1]*Anesa Surya;* [2]*Aman*
[1]Sarjanawiyata Tamansiswa University; [2]Yogyakarta State University
[1]surya.anesa@yahoo.co.id; [2]aman@uny.ac.id

Abstract

This research aims to (1) develop research formative authentic assessment instruments based on learning trajectory that is eligible for the fourth grade students of elementary schools; and (2) develop formative authentic assessment instruments based on learning trajectory that is effective for the fourth grade students of elementary schools. This developmental study refers to the model developed by Borg & Gall. The developmental design was grouped into four development procedures, consisting of: (a) Exploration, (b) development of the draft/proto-type, (c) product testing and revisions, and (d) final validation. The experimental subjects are some pilot project elementary schools which used Curriculum 2013 in Ngawi Regency. The data were collected using interview guides, documentation, assessment sheets of the product instrument of authentic assessment, observation sheets of the students, achievement test, questionnaire responses of teachers and students. The data of the instrument reliability were analyzed in terms of item discrimination and item difficulty, and the agreement index was employed for the reliability of the instrument. The research finding reveals that the research instrument in terms of attitude assessment, assessment of knowledge and skills according to subject-matter experts and experts in evaluation is categorized as 'very good'. The results of the item discrimination analysis show that nonnegative and item difficulties range from easy to hard. The formative authentic assessment instruments are categorized as 'reliable' by the agreement index of ≥ 0.75.

Keywords: *authentic assessment instrument, formative assessment, thematic-integrative, learning trajectory*

Introduction

The change of time and development of technology demand positive renewals in education system. Positive renewals happen due to the changing of Curriculum 2006 into Curriculum 2013. The curriculum change was carried out to adjust to the social situation and condition of the developing society. According to the Team of Curriculum of Professional Basic Subjects or *Mata Kuliah Dasar Profesi* (MKDP) and Learning Development (2011, pp. 17-42), 'a curriculum is developed based upon philosophy, psychology, sociology, and the development of science and technology.' The change is expected to answer the challenge of education needs based on the change of time and the development of technology, and also to support students to dynamically survive being with people from other countries based on the goal of national education.

The use of Curriculum 2013 in elementary school is basically due to the implementation of thematic-integrative teaching. Competencies formulated based on the subjects are changed into competencies that develop subjects according to certain themes. The system of assessments in Curriculum 2013 will be applied to four aspects, namely: Spiritual, social, knowledge, and skill.

Fogarty (2009, p.92) explains that 'the integrated curricular model represents a cross-disciplinary approach similar to the shared model, the integrated model blends the four major disciplines by setting curricular priorities in each and finding the overlapping skills, concepts and attitude in all four.' Therefore, thematic-integrative teaching is conducted in the following steps. First, the teacher should select concepts, skills and attitudes that will be given from different subjects, and then, the teacher chooses some concepts, skills and attitudes that are closely related to some subjects with the themes.

The theme selection in the integration of teaching materials needs to consider the aspects of consequences and attractiveness to students. Glatthorn and Jailall (2009, p. 103) state that 'in theme focused, developers begin by identifying major themes that would be of interest and consequence to students.' Besides, the themes have to be appropriate for students' real lives. With this regard, Meldrum and Peters (Webb & Pearson, 2012, p.17) claim 'a thematic model as one that adopts themes that are suggestive of a range of teaching ideas and often integrate several topics.' Suggestive themes are those which have to be appropriate for students' real lives and the learning materials that can develop students' critical thinking.

The teaching approach used in Curriculum 2013 changes as well. The approach used is a scientific approach. It focuses on observing, questioning, formulating, trying, and communicating (Ministry of Education and Culture, 2013, p.21). It is an approach which is able to provide the whole learning opportunities for students in supporting the thematic-integrative teaching.

Curriculum 2013 demands the implementtation of an authentic assessment in assessing students' competencies as well. An authentic assessment gives more complete data about students' abilities and it is based on the teaching and learning process, valuing products, and valuing the process (O'Neil, 1992, pp.14-19). It is an alternative assessment to avoid assessments which emphasize only on comprehension tests. An authentic assessment requires students to actively get involved in actual activities such as writing, doing work and projects, and creating products. Newman, King, and Charmichael (2007, p.3) state that 'characteristic authentic intellectual works as construction of knowledge, through the use of disciplined inquiry to produce discourse, product or performance that have value beyond school.'

Based on the Regulation of the Minister of Education and Culture No. 54 Year 2013 about qualification of graduates' skills which includes attitudes, knowledge and skills, an authentic assessment in Curriculum 2013 emphasizes attitude, knowledge, and skill assessments. This regulation is in line with the statement proposed by Guliker, **Bastiaens, and Kirschner** (2004, p.69), who say that 'authentic assesment requires students to use the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in pro-

fessional life.' Assessments of the three aspects are done in every actual activity conducted by students during the teaching and learning process.

Attitude is the tendency to like or dislike an object (Zakaria, 2010, p.1). Hence, attitude can determine someone's behavior towards an object. Johnson and Johnson (2002, p.168) explain that 'attitude is an important determinant of behavior.' Attitude assessments in the Curriculum 2013 will be applied to two aspects, namely: Spiritual and social. One of the techniques that can be used to assess attitude is self-assessment. Stiggins (1994, p.94) explains that 'self-assessment would properly be considered as questionnaire.' Self-assessment has many benefits, one of which is to train students to be honest (Kunandar, 2013, p.130).

Knowledge is a cognitive aspect. The assessment of knowledge in Curriculum 2013 covers factual, conceptual, procedural, and metacognitive knowledge. One of the techniques that can be used to assess knowledge is a written technique in the form of essays. Essays emphasize students' communicating skills in organizing their knowledge (Kubiszyn & Borich, 2010, p.163). Therefore, essays are authentic tests to assess students' skills compared to multiple choice tests.

Skills are the application of students' knowledge in a real life situation. The techniques used in assessing skills are work, project, product techniques based on the basic competencies and indicators. Wright (2008, p.246) explains, 'performance assessment measures both the skill and knowledge acquired by students and also assesses the application of judgment and insight on the students.' A project assessment is an assessment toward an assignment that has to be finished in certain period of time (Ministry of Education and Culture, 2013, p.94). Meanwhile, a product assessment is an assessment toward a process of making a quality product (Hosnan, 2014, p.406).

The implementation of the techniques explained earlier needs to be done to reflect students' skills. The preliminary step to do an authentic assessment is to create authentic assessment instruments. Instruments can be

defined as a measuring tool to collect data (Riduwan, 2013, p.1).

The process of creating authentic assessment instruments needs an instrument which can assess students' competencies and the instruments can be used in any circumstances. The process of creating indicators in an authentic assessment includes: (1) Identifying instructional goal, (2) pre-assessing the learners, (3) providing relevant instruction, and (4) assessing the intended learning outcomes (Gronlund, Linn & Miller, 2009, pp.32-34).

One of the key activities in the process of creating indicators of authentic assessment instruments is to identify students' learning needs (Gronlund, Linn, & Miller, 2009, p.34). Students' learning needs can be defined as students' plot diagnosis in understanding learning materials. One of the learning concepts which discusses students' learning plot components is called learning trajectory.

Learning trajectory is gained based on a hypothetical learning plot. A hypothetical learning plot needs to be tried to students to gain the real learning plots (Van den Akker, 2006, p.9). It consists of three components: The objective of teaching, the objective of learning, and students' development in teaching and learning processes (Clements & Sarama, 2009, p.3). Creating the development of students needs a theory to create the diagnosis.

Thematic-integrative teaching brings constructivism as the philosophy basis of teaching (Rusman, 2011, pp.332-339). The teaching theories that belong to constructivism are cognitive development and sociocultural theories (Schunk, 2012, pp.332-339). A sociocultural theory is a teaching theory which is more appropriate for the thematic-integrative teaching. According to Midoro (Liu & Wang, 2010, p.26), 'As far as thematic teaching is concerned, peer and student-teacher cooperation are consistent with Vygotsky's concept of scaffolding, which aims to increase Vygotsky's Zone of Proximal Development — the difference between what a learner can do independently and what the same learner can do when he/she is tutored, can improve learning results.' In conclusion, the thematic-

integrative teaching which employs a socio-cultural theory can improve students' learning results.

Further, Vygotsky's sociocultural theory emphasizes how children socially interact with adults in their environments and how they organize their learning experiences into some ways to make them gain cognitive skills. Thus, a sociocultural theory is the right theory to make a diagnosis on the students' learning development. To make a diagnosis of students' learning needs, students' grades have to be considered, since every grade has different learning plots as Wittek (2013, p.75) explains, 'Students' trajectories of learning bring together third parties in unique ways.'

The learning trajectory-based authentic assessment instruments can be developed in the forms of formative and summative assessments. Formative assessments aim to supervise students' learning development during the teaching and learning process, to give feedbacks to complete learning programs, and to know weaknesses which need to be fixed. Meanwhile, summative assessments are conducted when a learning experience or the whole learning materials have been done. These assessments aim to determine students' scores based on the level of learning results (Arifin, 2011, pp.35-36).

Learning trajectory-based authentic assessment instruments both in formative and summative assessments have to have standard criteria of instruments which can measure students' skills. The standard criteria are validity and reliability as Riduwan (2013, p.1) asserts, 'Valid and reliable instruments are needed to do an assessment.'

Valid instruments are the instruments which can be used to measure what will be measured (Arikunto, 2010, p.219). Instrument validity can be classified into four, namely: Validity based on the content of the test, response processes, internal structures, and the relation with other variables (Mardapi, 2008, p.8). Reliability is the constancy of the results of measurement (Sukmadinata, 2013, p.229). Instrument reliability can be classified into the following classifications: Internal consistency, stability, and between-assessor reliability (Mardapi, 2008, p.36).

Validity and reliability are measured by using classical and modern theories. Classical theories empirically emphasize the validity of items with the fulfillment of discrimination index and the level of difficulties (Subali, 2012, p.114). Discrimination is the ability of items to distinguish students who have already mastered competencies from those who have not. Miller (2008, p.132) suggests that 'Item discrimination provides an index of how an item discriminates between students who scored high and those who scored low in a test.'

The level of difficulties is an item category, easy or uneasy items to students. The level of difficulties can be understood by calculating the number of students who answer correctly. Miller (2008, p.130) explains that 'Item difficulty indicates the percentage of students who responded correctly to a test item.' Meanwhile, instrument reliability can be calculated using one of the types of reliability mentioned earlier.

The main problem which is currently arising about authentic assessments in Curriculum 2013 is the teachers' less understanding of authentic assessments. In addition, learning instruments which is proposed by the government are not practical, so the teachers find it difficult to use them to assess students one by one in all of the competencies, especially in thematic-integrative teaching in elementary school.

Realities found in the field show that teaching and learning processes and assessment instruments do not really consider students' learning plots. Teachers assess and arrange teaching materials on the basis of assumptions and predictions, even teachers sometimes make question items taken from existing question items in books (Mulyana, 2012, p.12).

The government publishes teachers' handbooks and students' textbooks in order to support them in understanding authentic assessment instruments. Assessment drafts are written on the teachers' handbooks, but the drafts have not included all aspects yet, especially spiritual aspect. The indicators used in assessing social, cognitive, and skill aspects have not been provided yet. The scoring

guidelines do not give enough clear explanation for teachers. It causes the teachers to less understand authentic assessment instruments and how to develop them.

In addition, character education is an important component that should be considered to grow moral values according to the goal of national education. It is learning toward strengthening and developing the entire children's behavior which is based on certain values that the schools refer to (Kusuma, Triatna & Permana, 2011, p.5). One of the character values that can be embedded to students since they were in Elementary School is to care about environments. The subtheme which is discussed to form environmental education is the subtheme of the diversity of flora and fauna. It teaches students to preserve the environment especially flora and fauna. Besides, social aspect will also teach students to cooperatively preserve the environment.

This research employed the subtheme of the diversity of flora and fauna. Therefore, the assessment instruments developed are formative assessment instruments. The targets are Grade IV students. Grade IV is the beginning of a high level of class. Grade IV students' learning plots are higher than grade III students'. Grade IV students develop their attitudes and skills. It is shown by the development of their spiritual attitudes, while their independent attitudes are shown that they do not depend on their teachers and friends. It is supported by the concept of sociocultural development theories.

The objective of this research is to develop learning trajectory-based formative authentic assessment instruments which are eligible and effective for Grade IV students. Eligible instruments are based on expert judgment, and effective instruments are based on classical item validity and empirical reliability instruments. The theoretical significance of this research is as scientific contribution for the next research. For teachers, this research is as a reference to develop learning the trajectory-based authentic assessment instruments for grade IV students. For students, this research is important to give feedbacks toward their learning results (students know their strengths and weaknesses), and for schools, it is as a reference to develop assessment instruments for other subthemes.

Method

Development Models and Research Procedures

The development model used in this research was a design proposed by Borg and Gall (1983, pp.775-778). In addition, the development procedures used the procedures also proposed by Borg and Gall (1983) which were modified into four development procedures, namely: (1) Exploration; (2) prototype development; (3) field testing; and (4) final validation. More clearly, the procedures can be seen in Figure 1.
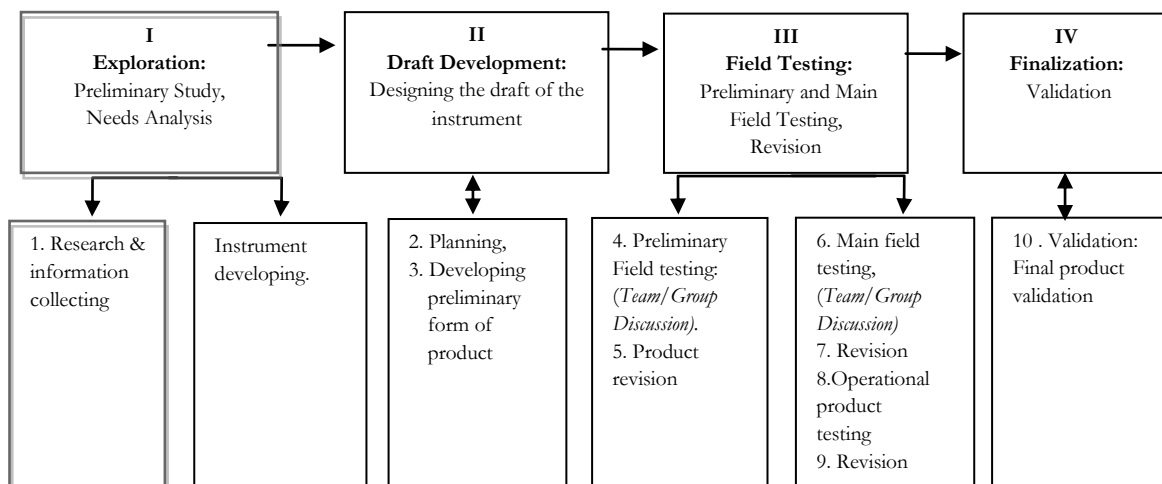

Figure 1. Development procedures

Testing Design

The testing design in this research includes the following steps: (1) Small group testing; (2) analysis of small group testing based on suggestions and feedbacks from one practitioner and 17 students; (3) revision; (4) large group testing; (5) analysis of results of large group based on questionnaires of three practitioners and 118 students and also the measurement of discrimination index, the level of difficulties and reliability by determining coefficient of agreement index; (5) revision; and (6) final product. The subjects of the testing were all elementary schools in Ngawi Regency which have applied Curriculum 2013. The subject of small group testing was SDN Jogorogo 1 (Jogorogo 1 Public Elementary School) with 17 students, while the subject of large group testing was SD Margomulyo 1 (Margomulyo 1 Elementary School) with 118 students. The schools were selected based on the characteristics of the elementary school, geographical location, and the number of students.

Data Collection Techniques and Instruments

The data collection techniques which were used were interviews, documentation, questionnaires, observation, and also tests. Interviews were conducted in the exploration stage in order to collect information from the teachers of pilot project elementary school which has applied the Curriculum 2013. Meanwhile, documentation was carried out in the exploration stage as well by analyzing documents that had been made by the teacher. Questionnaires were given to the evaluation and materials experts, teachers, and students. The questionnaires for the evaluation and materials experts were given before they were tested in the field. Teachers' and students' questionnaires were given when the small and large group testing was done. Observation was conducted in small and large group testing. Tests were administered during the teaching and learning process in large group testing. Meanwhile, the developed research instruments were interview guides, documentation sheets, questionnaires, observation sheets, and a test.

Data Analysis Techniques

The data analysis techniques in this research were divided into four. They were preliminary study, product development process, product eligibility, and product effectiveness techniques. The data of the preliminary study were analyzed descriptively. The data of product development process were analyzed qualitatively. The data of product eligibility were analyzed qualitatively, and then conversed into a scale of five to know the quality of the product. The data of product effectiveness was analyzed by calculating the classical item validity by calculating the discrimination index and the level of difficulties, while reliability was analyzed by calculating agreement index which was seen from z score and coefficients of Cronbach Alpha/ KR 21. The data of observation were processed descriptively and the data of teachers' and students' questionnaires were analyzed by calculating the mean, and then conversed into a scale of five to know the quality of the instruments.

Findings and Discussion

Results of Preliminary Study

The results of the preliminary study were obtained from the needs analysis and literature review. The results of needs analysis were obtained from the interviews, documentation, and observation. These results clearly show that teachers need learning trajectory-based formative authentic assessment instruments for elementary school. In addition, the results of the literature review discuss the following points: Authentic assessment instruments, formative assessments, learning trajectory, and thematic-integrative learning.

Results of Product Development Process

The initial products in this development research are assessment instruments of attitude, knowledge, and also skills. Attitude assessment is a self-assessment. Knowledge assessment is essays, and skill assessment is an assessment using working, project and product techniques.

Results of Product Eligibility

Expert Validation

The results of product eligibility were obtained from experts. The criteria of attitude and skill assessments are 'Good', with a range of 40.80 < X ≤ 50.40. For cognitive assessments, the criteria are in the range of 37.39 < X ≤ 46.19. Generally, the assessment instruments are categorized in 'Good'. The results of the expert judgment of the assessment instruments can be seen in Table 1.

Table 1. The results of experts' judgment

Learning	Attitude	Knowledge	Skill
1	55.50	43.50	55.00
2	54.00	44.00	53.50
3	48.83	46.25	50.17
4	50.00	45.00	53.00
5	53.75	45.00	54.25
6	52.25	45.00	51.25

Based on Table 1, it can be concluded that the assessment instruments of learning attitudes 1, 2, 5 and 6 gain scores more than 50.40, which is in 'Very Good' category, while for learning 3 and 4, the gain score is less than 50.40, which is in 'Good' category. The assessment of learning cognitive 1 to 6 shows the gain score of more than 50.40, which is in 'Very Good' category, and that of learning cognitive 3 shows the gain score of less than 50.40, which is in 'Good' category. Meanwhile, suggestions and feedbacks were given to aspects of materials for assessments of attitude, language for assessments of knowledge and constructive aspects for assessments of skills.

Results of Small Group Testing

The results of small group testing were obtained from classroom observation and teachers' and students' questionnaires. The results of students' questionnaires show that: (1) Students do not understand how to choose words and make complex sentences, (2) students understand decimal addition and subtraction symbols, (3) students need the teacher's guidance when carrying out projects, (4) students are unable to dig complex information from various sources. The results

of classroom observation are considered to revise the instrument products.

The criterion of the results of teachers' questionnaires is 'Good', with the range of 40.80 < X ≤ 50.40. The mean of the results of teachers' questionnaires for attitude, knowledge and skill assessments in small group testing is 'Good'. The results of the teachers' responses in small group testing can be seen in Table 2.

Table 2. The results of teachers' questionnaires in small group testing

Learning	Attitude	Knowledge	Skill
1	52.00	52.00	55.50
2	56.00	52.00	54.00
3	53.67	54.67	55.20
4	53.00	52.50	53.00
5	55.50	54.00	56.00
6	55.00	55.50	54.00

Based on Table 2, it can be concluded that assessment instruments of attitude, knowledge, and skill gain more than 50.40 with 'Very Good' category, while suggestions and feedbacks given include linguistic aspects which need to be simplified to make students familiar with it, and constructive aspects for skill assessments by replacing the column of total score with the column of score mean to ease teachers in assessing students' skills.

Criterion of the results of students' questionnaires is 'Good' with score range of 61-80. Therefore, the assessment instruments have to gain minimal score of 61 to be categorized as 'Good'. Students gave responses for attitude and knowledge assessments. Questionnaires for skill assessments were not given to students because they were related to the teachers. The results of students' questionnaires for small group testing are shown in Table 3.

Table 3. The results of students' questionnaires in small group testing

Learning	Attitude	Knowledge
1	79.04	88.97
2	90.44	69.41
3	82.35	89.71
4	84.40	87.82
5	87.39	88.62
6	84.87	87.06

Based on Table 3, it can be concluded that assessment instruments of skills gain scores more than 80.00 with 'Very Good' category, while knowledge assessments for learning 1, 2, 4, 5, and 6 gain scores of more than 80.00 with 'Very Good' category, and for learning 2, it gains score of 69.41 with 'Good' category.

Results of Large Group Testing

The results of large group testing were gained from classroom observation, teachers' and students' questionnaires. The results of classroom observation in large group testing show that students' teaching and learning activities are not different from those in small group testing. Therefore, assessment instruments developed in large group testing are not continued to be revised.

The criterion of the results of teachers' questionnaires in large group testing is the same as that in small group testing which is 'Good' with the range of $40.80 < X \leq 50.40$. Instruments need to have scores more than 40.80 to belong to 'Good' category. The mean of the results of teachers' questionnaires for assessment instruments of skill, knowledge and attitude in small group testing is categorized as 'Very Good'. For clearer explanation, the results of teachers' questionnaires are shown in Table 4.

Table 4. The results of teachers' questionnaires in large group testing

Learning	Attitude	Knowledge	Skill
1	55.50	55.50	55.50
2	56.00	53.00	52.00
3	54.67	55.00	56.60
4	56.50	56.50	54.00
5	55.50	55.50	55.50
6	55.50	58.00	57.00

Based on Table 4, it can be understood that the assessment instruments of attitude, knowledge and skill in large group testing gain score more than 50.40 with 'Very Good' category. The suggestions and feedbacks which were given by the teachers are about constructive aspects to add column of predicate in skill assessments. The column of predicate is put next to the column of score mean in order to ease teachers in determining

feedbacks during the teaching and learning process after knowing the competencies that their students had achieved.

The criterion of the results of students' questionnaires in a large group testing is the same as that in small group testing which is categorized as 'Good' with the score range of 61-80. The minimum total score for the attitude assessments that has to be owned is 61 with 'Good' category. Further, the results of students' questionnaires in large group testing are shown in Table 5. Based on Table 5, it can be concluded that the assessment instruments of attitude and knowledge gain scores about more than 80.00 which belong to 'Very Good' category.

Table 5. The results of students' questionnaires in large group testing

Learning	Attitude	Knowledge
1	92.29	94.22
2	95.75	83.89
3	94.56	91.88
4	92.28	88.13
5	91.41	88.32
6	90.07	90.33

Beside the results of observation and also teachers' and students' questionnaires in large group testing, a data analysis of students' learning tests was carried out to empirically determine the validity and the reliability. This research used a theory of classical tests to measure the validity and reliability of the instruments. The validity of classical items can be seen empirically from the fulfillment of item requirements by measuring them. This research used criterion-referenced test. Therefore, to fulfill the validity empirically, this research used item discrimination with a condition that it was not negative and the level of item difficulties from easy to difficult was appropriate with criterion-referenced test tests.

Item discrimination is the ability of test items to distinguish students who already have and have not achieved competencies yet. Furthermore, the results of the measurement of item discrimination mean in every teaching and learning process can be seen in Table 6.

Table 6. Item discrimination mean of learning trajectory-based formative authentic assessment instruments

Learning	Attitude	Knowledge	Skill
1	0.61	0.53	0.39
2	0.48	0.45	0.31
3	0.50	0.54	0.38
4	0.54	0.41	0.36
5	0.46	0.49	0.31
6	0.39	0.54	0.29

Based on Table 6, it can be concluded that the assessment instruments of skill, knowledge, and attitude gain the mean of co-efficient of discrimination more than 0, so the coefficient of discrimination is not negative. It means that it fulfills the requirements for test with criterion-referenced test.

The level of difficulties is an item category, which is easy or uneasy to do by students. It can be understood by calculating the number of students who answer correctly divided with the number of students who join the test. The results of the mean of level of difficulties in every teaching and learning process are presented in Table 7.

Table 7. Mean of level of item difficulties of learning trajectory-based formative authentic assessment instruments

Learning	Attitude	Knowledge	Skill
1	0.59	0.71	0.78
2	0.66	0.59	0.59
3	0.71	0.61	0.76
4	0.73	0.77	0.74
5	0.76	0.53	0.76
6	0.78	0.60	0.76

The criterion of level of difficulties is appropriate with the aim of measurement of criterion reference, which is, easy to difficult. To determine the categories of level of difficulties is to see the coefficient of level of difficulties with the following categories: (1) Level of difficulties < 0.3 means 'Difficult'; (2) $0.3 <$ level of difficulties ≤ 0.7 means 'Fair'; and (3) level of difficulties > 0.7 means 'Easy' (Team of Educational Research Center, 2010, p.36). Based on Table 7, it can be said that the mean of level of difficulties of authentic assessment instruments gains various scores with easy to difficult categories.

The afore-mentioned results of data analysis for classical item validity fulfill the criteria of discrimination and level of dif-ficulties based on criterion-referenced tests. Therefore, it can be said that classical item validity is fulfilled. In addition to item validity, the instrument reliability for the criterion-referenced tests have to be fulfilled as well. Reliability for criterion-referenced tests has to be stated with Kappa index or agreement index (Subali, 2012, p.119). Agreement index was obtained from z score (standard score) and coefficient from Cronbach Alpha or KR 20/ 21 based on data scale which was used in the developed range of instruments

In this research, attitude assessments employ scale 1 and 0, while knowledge and skill assessments use scale were 1-4. There-fore, the agreement index of attitude assess-ments was gained based on z score and coefficient KR 21. Meanwhile, knowledge and skill assessments of agreement index were gained from z score and coefficient Cronbach Alpha. The results of agreement index for attitude, knowledge, and skill assessments are shown in Table 8.

Table 8. Instrument reliability of learning trajectory-based formative authentic assessments

Learning	Attitude	Knowledge	Skill
1	0.79	0.86	0.85
2	0.76	0.80	0.82
3	0.81	0.83	0.83
4	0.76	0.87	0.89
5	0.76	0.79	0.87
6	0.81	0.84	0.87

The criterion of reliable instruments has the agreement index of at least 0.75 (Frisbie, 2005, p.26). Based on Table 8, it can be concluded that the instruments are 'reliable'. Therefore, the requirements of classical item validity and instrument reliability are fulfilled.

Product Revision

Product revision was done in three stages: Revision based on experts' suggestions and feedbacks, suggestions and feedbacks in small and large group testing. The draft re-vision to create main instruments was based on suggestions and feedbacks covering ma-

terial, construction and linguistic aspects. The revision of material aspects was about attitude indicators that needed to be suited with the subtheme of 'The diversity of flora and fauna'. The revised indicator of spiritual aspects became 'I am grateful for the diversity of flora and fauna'. Then, the revised indicator of social aspects became 'I cooperatively take care of animals, water plants and clean up the classroom and school with my friends'.

The construction aspects were revised by adding answer keys for skill assessments in order to ease teachers to assess students' skills. The format aspects were revised by changing font and font size into 'Times New Roman' 12. The linguistic aspects revised the test items in knowledge assessments which the students were not familiar with. The revised test items were those in learning 2 point 2 and 3, learning 3 concluding the content of the text point 2, learning 4 concluding the content of the text point 2, and learning 5 interviewing point 1 and 2.

Learning 2 point 2 was revised became 'Explain our rights toward the beauty of flora'. Point 3 was revised into 'Explain the examples of our actions to keep and preserve flora and fauna'. Learning 3 point 2 for concluding the content of the text was revised into 'Explain our rights to keep and preserve pine trees'. Point 4 was revised into 'Explain the beauty of Cendrawasih'. Learning 5 interviewing point 2 was revised into 'Explain our rights toward the environments around us'. Point 3 'Explain our obligation toward the environments around us' was revised into 'Explain examples of our actions to keep and preserve environments'.

Product revision for small group testing included construction and linguistic aspects. The construction aspects in instruments needed some revision: (1) The total score in skill aspects was revised into score mean to ease teachers in assessing, (2) fulfillment guidelines of attitude assessments were revised into 'I will give "check" (√) in column: Yes, if I did it and Never, if I did not do it', (3) assessment rubrics of coloring and drawing skills were detailed in three stages: Preparation, implementtation, and final stage assessments, (4) assessment rubrics of project skills were suit-

ed with students' learning plots for preparation aspect by attaching students' guidance.

Linguistic aspects in attitude assessments needed some revisions, such as 'I don't kill animals'. The statement was not understandable for students, so it was revised into 'I kill animals'. The later statement was 'I don't kill plants' which became 'I kill plants'.

Knowledge assessments needed some revisions mainly in learning 1 concluding the content of the texts. The previous question 'What is the main idea (*gagasan utama*) of the text above?' was revised into 'What is the main idea (*ide pokok*) of the text above?' (fit with students' learning plots).

Product revision in group testing did not get a lot of revisions, since the measurement of classical item validity and reliability had fulfilled the requirements. The column of 'Predicate' was added in skill assessment as a revision in large group testing of construction aspects. This 'Predicate' column was next to the column of score mean to know the results of students' learning suited with the determined indicators.

Final Product Review

Final product review covers some advantages: (1) Oriented; (2) comprehensive; (3) authentic; and (4) practical. Being oriented is that a final instrument refers to the effort to achieve the goal of Curriculum 2013. The goal of Curriculum 2013 is to improve balance of soft skill and hard skill (attitude, knowledge and skills). The final instrument measures the competencies of students' attitude, knowledge and skills. Being comprehensive is that a final instrument is comprehensible to assess students' competencies in all aspects, such as attitude, knowledge, and skills based on learning indicators and students' learning plots, so the assessment will show students' competencies based on their characteristics. Being practical is that a final instrument is practical, because it eases teachers to assess students based on the literature review and students' competencies.

Limitation of the Research

This research has two limitations. First, the development of learning trajectory-based

formative authentic assessment instruments in subtheme of 'The Diversity of Flora and Fauna' does not certainly result in valid, reliable and practical instruments and based on students' learning plots in the next subtheme, since students' learning plots improve and their competencies are different.

Second, learning trajectory in the development of assessment instruments is limited in Grade IV in elementary schools in Ngawi Regency. Grade IV students' learning plots are not always suitable for grade IV students in other regions, since students have different learning plot characteristics. Third, some pieces of students' questionnaires are not simple enough, so they are not comprehensible to students.

Conclusion and Suggestions

Conclusion

Based on the results of the research and the earlier-mentioned discussion, it is concluded that learning trajectory-based formative authentic assessment instruments which are contemplated from the assessment instruments of attitude, knowledge and skill aspects are eligible with 'Very Good' category, based on the experts of evaluation and materials. Besides, learning trajectory-based formative authentic assessment instruments which are contemplated from the assessment instruments of attitude, knowledge, and skill aspects are effective with the measurement of empirical item validity, have non-negative discrimination, and also level of difficulties with easy to difficult categories. Meanwhile, the results of agreement index of instrument reliability is ≥ 0.75, so it can be said that it is reliable.

Suggestions

Some suggestions are proposed as follows. Teachers are hoped to use learning trajectory-based formative authentic assessment instruments to assess students' competencies and as a reference to make similar assessment instruments with different subthemes. For school, it is hoped that these assessment instruments can be used as a reference to develop assessment instruments for other subjects. For national education, it is hoped that these assessment instruments can be used as a reference of training and education program to develop teachers' competencies in understanding authentic assessment instruments based on Curriculum 2013.

References

Arifin, Z. (2011). *Evaluasi pembelajaran* [Learning evaluation]. Bandung: Rosdakarya.

Arikunto, S. (2010). *Prosedur pengembangan* [Developmental Procedure]. Jakarta: Rineka Cipta.

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

Clements, D.H. & Sarama, J. (2009). *Learning and teaching early math: The Learning Trajectories approach*. New York, NY: Routledge.

Fogarty, R. (2009). *How to integrate the curricula*. Palatine, IL: Skylight.

Frisbie, D.A. (2005). Measurement 101: Some fundamentals revisited educational measurement: Issues and practice. *Fall*, 24(38), 21-28.

Glatthorn, A., & Jailall, M. (2009). *The principal as curriculum leader shaping what is taught and tested* (3rd ed.). Thousand Oaks, CA: Corwin Press.

Gronlund, N.E., Linn, R.L., & Miller, M.D. (2009). *Measurement and teaching* (10th ed.). Upper Sadle River, NJ: Merril.

Guliker, J.T.M., Bastiaens, T.J., & Kirschner, P.A. (2004). A five dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86.

Hosnan, M. (2014). *Pendekatan saintifik dan kontekstual dalam pembelajaran abad 21* [Scientific and contextual approaches in the 21st century learning]. Bogor: Ghalia Indonesia.

Johnson, D. & Johnson R. (2002). *Meaningful assessment: A management and cooperative process*. Boston, MA: Allyn and Bacon.

Kubiszyn, T. & Borich, G. (2010). *Educational and measurement: Classroom application and practice* (9rd ed.). Hoboken, NJ: John Willey & Sons.

Kunandar. (2013). *Penilaian autentik (Penilaian hasil belajar peserta didik berdasarkan Kurikulum 2013)* [Authentic assessment (Assessment of learning participants' learning results based on Curriculum 2013)]. Jakarta: Rajawali Press.

Kusuma, D., Triatna, C., & Permana, J. (2011). *Pendidikan Karakter* [Character education]. Yogyakarta: Diandra.

Liu, M.C. & Wang, J.Y. (2010). Investigating knowledge integration in web-based thematic learning using. *Journal of International Forum of Educational Technology & Society (IFETS)*, 13, 25-39.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes* [Test and non-test instrument arrangement techniques]. Yogyakarta: Mitra Cendekia.

Miller, P.W. (2008). *Measurement and teaching.* Munster, IN: Patrick W. Miller and Associates.

Ministry of Education and Culture. (2013). *Kurikulum 2013 untuk sekolah dasar (SD)/ madrasah ibtidaiyah (MI)* [Curriculum 2013 for elementary schools/ madrasah ibtidaiyah]. Jakarta: Kementerian Pendidikan dan Kebudayaan.

Mulyana, T. (2012). Pengembangan bahan ajar melalui desain penelitian [The development of lesson material through research design]. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, 2, 21-28.

Newman, F.M., King, M.B., & Charmichael, D.L. (2007). *Authentic instructional and assessment.* Des Moines, IA: Grimes State of Office Building.

O'Neil, J. (1992). Outting performance assessment to the test. *Educational Leadership*, 49, 14-19.

Riduwan. (2013). *Skala-skala pengukuran variabel* [Variable measuring scales]. Bandung: Alphabeta.

Rusman. (2011). *Model-model pembelajaran* [Learning models]. Jakarta: Rajawali Press.

Schunk, H.D. (2012). *Learning theories: An educational perspective.* Boston, MA: Pearson.

Stiggins, R. J. (1994). *Student classroom assessment.* New York, NY: Macmillan.

Subali, B. (2012). *Prinsip assessmen dan evaluasi pembelajaran* [Principle of assessment and learning evaluation]. Yogyakarta: UNY Press.

Sukmadinata, N.S. (2013). *Metode penelitian pendidikan* [Educational research method]. Bandung: Remaja Rosdakarya.

Team of Curriculum of Professional Basic Subjects and Learning Development. (2011). *Kurikulum dan pembelajaran* [Curriculum and learning]. Jakarta: Rajawali Press.

Team of Educational Research Center. (2010). *Analisis soal secara manual* [Problem analysis manually]. Jakarta: Pusat Penilaian Pendidikan Balitbang Kemendiknas.

Van den Akker, J. (2006). *Educational design research.* Macmilan: Routledge.

Webb, P. & Pearson, P. (2012). Create unit and lesson planning through a thematic/integrated to teaching games for understanding (TGFU). *The Journal Physical Education*, 45, 17-22.

Wittek, L. (2013). The activity of writting for learning in a nursing program: Trajectories of meaning making. *Sociology Journal*, 14, 73-94.

Wright, R.J. (2008). *Educational assessment test measurement in the age of accountability*. Los Angeles, CA: Sage.

Zakaria, T.R. (2010). *Pedoman penilaian sikap* [Attitude assessment guidance]. Jakarta: Pusat Penilaian Pendidikan Balitbang Kemendiknas.