The development of an authentic assessment instrument to measure students skills in learning biology in grade 8 junior high school

P D Yuniawan
Program Pendidikan Sains, Pascasarjana, Universitas Negeri Surabaya, Jl. Ketintang No. 30, Ketintang, Gayungan, Surabaya, 60231, Indonesia
puguhdwi@gmail.com

Abstract. Authentic assessment in the 2013 Curriculum is echoed for all lessons. The 2013 Curriculum emphasizes that student was expected to have knowledge competence, skill competence, and attitude competence. This research was conducted to develop authentic assessment instruments in the implementation of the final assessment of the school, especially on the skills aspect so as to obtain valid, practical, and effective assessment instruments. This assessment instrument research uses a 4-D development model. The objective of the research is the skill aspect assessment instrument that was tested on 34 graders 8 Junior High School in the first semester with the experimental “One Shot Case Study” model. Data analysis using descriptive qualitative. The result of data analysis showed: validation of skill assessment instruments was a very valid (3.56), readability of assessment instruments in categories was interesting and easy to understand, implementation of the assessment instruments entered in a very good category with a very high percentage (97.92%), reliability showed a reliable result (r>0.70). It can be concluded that the development of authentic assessment instruments to measure students skills in learning biology grade 8 junior high schools was valid, practical, and effective.

1. Introduction
The curriculum changes will have an impact on the learning process and the evaluation process in the school. The 2013 Curriculum was a competency-based curriculum that provides students with attitude, knowledge and skills competencies so that assessment is expected during the learning process and at the end of the lesson. Implementation of the 2013 Curriculum in schools was still an obstacle for some teachers. The low knowledge of madrasah teachers on assessment standards, the lack of training on assessment of learning and madrasah teachers does not yet know how to report the results of the 2013 Curriculum [1]. In the monitoring and evaluation of the implementation of the 2013 Curriculum by the Directorate of Junior High School in 2014 conducted in 29 provinces and involving 76 districts showed teachers of junior high school that understand and use teacher manual and student book of 33% -45%, lessons plan are 50% -59%, carry out learning with scientific approach of 42% -66%, and develop and carry out an authentic assessment of 25% -37%. The data shows that overall teachers competent to implement the 2013 Curriculum were still below 60%. The lowest competence was developing instruments and carrying out authentic assessments, thus requiring a lot of attention and support [2]. From interviews with teachers of Junior High School 6 Bontang, they were still difficulties in preparing authentic assessment instruments. To obtain authentic assessments that were valid and reliable need to interview and see the assessment instruments used by teachers [3].
The approach to the assessment of the learning process recommended in the 2013 Curriculum was an authentic assessment approach that assesses students' readiness, processes, and overall learning outcomes. The authentic assessment required students to use the same competencies, or combinations of knowledge competencies, skills competencies, and attitudes competencies, that they need to apply in real life context [4]. The authentic assessment was the activity of assessing learners that emphasize what should be assessed, both process and outcome with various assessment instruments tailored to competing demands that exist in the Competency Standards or Core Competencies and Basic Competencies [5]. The authentic assessment was a form of assessment in which students were assigned real-world tasks that demonstrate meaningful application of essential knowledge and skills [6]. Authentic judgments exist that define judgments that require students to use the same competencies or combine the competencies of the knowledge, skills, and attitudes they need to apply in real-life situations [7].

Assessment of skills was an assessment aimed at assessing the ability of students in applying knowledge in performing certain tasks in various contexts in accordance with indicators of achievement of competence. Skill assessment includes the skills of trying, processing, tasting, and reasoning. Skills assessment in the concrete realm involves the activities of using, parse, stringing, modifying, and making. While in the abstract realm, these skills include writing, reading, counting, drawing and composting [8]. In science learning, students need to master process skills up to integrated process skills so as to improve the process skills of students, teachers in learning using process skills approach. The process skills approach is a process of learning that is structured in such a way that students can find facts, build concepts and theories with their own intellectual and scientific skills [9]. There were various skills assessment techniques, including practice assessment, product assessment, project appraisal, and portfolio assessment.

With the low monitoring result of the implementation of Curriculum 2013 which shows teachers of junior high school subject who have the ability to understand and use teacher manual and student book that is 33% -45%, carry out authentic assessment spanning 25% -37%, and considering the importance of skill competence for students then this study develops an authentic assessment instrument guidance on the skills aspect so that a valid, practical, and effective judgment instrument will be obtained. A good evaluation tool is determined by several criteria including validity, reliability, objectivity, practicability, economics, implementation, and accuracy of construction [10, 11, 12]. Akker suggests that the quality of development research can be seen from three criteria of validity, practicality, and effectiveness [13].

2. Method
This research was conducted to develop summative assessment instruments in the implementation of the final assessment of the school with biology science topic in semester 1, especially on the skills aspect so as to obtain valid, practical, and effective assessment instruments. A skills assessment instrument developed using the technique of product assessment and the technique of practice assessment. The technique of product assessment was in the form of human skeleton poster while the technique of practice assessment was a test of fat content in food. Both assessment techniques were developed based on indicators on Core Competencies and Basic Competencies of science subjects class 8 Semester 1 Curriculum 2013. This assessment instrument development research using 4-D development model but at the development stage was not done. The objective of the research was the skill aspect assessment instrument tested on 34 junior high school students of SMP Negeri 6 Bontang Semester 1 with an experimental model of "One Shot Case Study" model. The validity of the instrument was assessed by the experts and the readability of the assessment instrument by the teacher. Practicality gained from the implementation of skills assessment instruments. Effectiveness is gained from the results of reliability analysis. Data analysis used qualitative descriptive.
3. Result and Discussion

3.1. Validity

The test is said to have validity when the test measures what should be measured [14]. The validity was the content validity and the construct validity [13]. The result of developing an authentic assessment instrument to measure students’ skill competencies was then validated by the experts. Validation was done by considering the content/material, construction, and language aspects of two experts on authentic assessment instruments to measure students’ skills competencies presented in Table 1.

Table 1. Validation Results of Authentic Assessment Instruments Aspects of Student Skills

<table>
<thead>
<tr>
<th>No</th>
<th>Aspect</th>
<th>O1</th>
<th>O2</th>
<th>Average</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Content</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>VV</td>
</tr>
<tr>
<td>2</td>
<td>Construction</td>
<td>3.00</td>
<td>3.33</td>
<td>3.17</td>
<td>V</td>
</tr>
<tr>
<td>3</td>
<td>Language</td>
<td>3.00</td>
<td>4.00</td>
<td>3.50</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>3.33</td>
<td>3.78</td>
<td>3.56</td>
<td>V</td>
</tr>
</tbody>
</table>

Description: VV: Very Valid; V: Valid; O1: Observer 1; O2: Observer 2

Table 1 shows that the authentic assessment instruments of the skill aspects developed were valid. Reliability analysis results obtained for 93.75%. Instrument development is called reliable if it has a percentage of ≥ 70% [15]. This means that the authentic assessment instrument developed reliably because of the reliability value of 93.75%. The preparation of an assessment instrument can fulfill the validity of the content was in the preparation of instrument items referring to basic competencies, indicators, and learning objectives [16]. The validity of the construct of an instrument must be done through expert review or justification or through the judgment of a group of people who control the substance of the variable to be measured [11]. To determine the validity of the construct of an instrument, a theoretical review of a concept of the variable to be measured, from the formulation of the construct, the determination of dimensions and indicators, to the elaboration and the writing of item items of the instrument. An evaluation tool is said to have construct validity if the question items on the evaluation tool measure the predetermined indicator [17]. Writing test questions is an important step because errors in sentence selection can result in invalidation of a test [10]. When writing an item, the learning objective or outline point for which the item is intended must be relevant to the overall goal of the test. The goal of test development is to cover each construct included in the exam outline, therefore newly written items must specifically map to the outline of the test [18].

The validity of the instruments of authentic assessment of skills aspects is also seen from the legibility of authentic assessment instruments of skills aspects by teachers. The result of the assessment of the authenticity of an authentic assessment instrument to measure the competence of the students' skills by the 4 teachers presented in Figure 1.

Figure 1. Results of Authenticity Assessment Instrument Assessment Aspects of Student Skills
Figure 1 shows that the average teacher states that the developed assessment instrument was interesting and easy to understand. To determine the validity of the test then the test also needs to be tested its readability in terms of the user [19]. The legibility test is performed to obtain display validity by providing an assessment instrument to several people/students to see if the person/student understands what was being asked in the instrument. Ability to predict test readability is useful because it helps educators select appropriate texts for students and authors write texts accessible to the audience they target [20].

3.2. Practicality
The practicality of the authentic assessment instrument was derived from the implementation of the instrument in the conduct of the research with the help of two observers to see and assess the process of implementing the stages of the assessment activity already established. Analysis of the implementation is presented in Table 2.

<table>
<thead>
<tr>
<th>No</th>
<th>Aspect</th>
<th>Observer 1</th>
<th>Observer 2</th>
<th>Average</th>
<th>Criteria (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Preparation</td>
<td>19.00</td>
<td>20.00</td>
<td>19.50</td>
<td>97.50</td>
</tr>
<tr>
<td>2</td>
<td>Implementation</td>
<td>16.00</td>
<td>16.00</td>
<td>16.00</td>
<td>100.00</td>
</tr>
<tr>
<td>3</td>
<td>Time</td>
<td>4.00</td>
<td>3.00</td>
<td>3.50</td>
<td>87.50</td>
</tr>
<tr>
<td>4</td>
<td>Class situation</td>
<td>8.00</td>
<td>8.00</td>
<td>8.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 2 shows each stage of the assessment: preparation, execution, time, and class situation have criteria above 75% so it can be concluded that at each stage of the implementation of the assessment performed very good category (75% - 100%). And overall it has been done very well that was equal to 97.92%. In research, the development of practicality refers to the extent to which users intervene so it will be more interesting and can be used in normal conditions [13]. A practical test was an easy test of its implementation, easy checking, and there were clear instructions that can be given/initiated by others [10]. The test is practical, that means is not excessively expensive, stays within appropriate time constraints, is relatively easy to administer, and has a scoring/evaluation procedure that is specific and time-efficient [21].

3.3. Effectiveness
The effectiveness of assessment instruments was obtained from the analysis of reliability and sensitivity of the test. The effectiveness could be seen from the experience and results of interventions that were consistent with the objectives that have been defined [13]. The reliability of a test was obtained from the consistency of the results of the assessment [22]. Calculation of the coefficient of reliability assessment of competency skills with product assessment techniques (make a poster of the human skeleton) which uses a formula obtained the Spearman-Brown coefficient of 0.72. Calculation of the coefficient of reliability assessment of competence with technical skills practice assessment (test fat content) using a formula obtained the Spearman-Brown coefficient of 0.75. According to Linn and Kaplan minimal limit reliability of an instrument is 0.7 so that an instrument was said to be reliable if it had a coefficient of at least 0.7 [16]. So both of the assessment instruments aspects skills that developed was reliable. Some of the factors that affected the reliability of test scores were: the greater number of items, the more reliable a test; the longer the time of the test, the more reliable; the narrower the range of trouble item, the greater reliable; issues that are interconnected will reduce reliabiltas; more objective scoring, the more reliable; the inappropriateness of the scoring; to answer most questions by guessing; the more homogeneous the material the greater constancy; experience of the examinees;
4. Conclusion
The development of authentic assessment instruments to measure students skills in learning biology grade 8 junior high schools was showed a valid, a practical and an effective. The result of data analysis showed validation of skill assessment instruments was a very valid (3.56), readability of assessment instruments in categories was interesting and easy to understand, implementation of the assessment instruments entered in a very good category with a very high percentage (97.92%), reliability showed a valid, a practical and an effective.

5. Acknowledgments
Our thanks to the principal of Junior High School 6th Bontang who have supported the implementation of this research, science teachers and students for their cooperation so that this research can be accomplished, my beloved wife and children. Hopefully, this research can be useful for education.

6. References

International Conference on Mathematics and Science Education of Universitas Pendidikan Indonesia Volume 3, 2018 | P-ISSN 2655-2361, E-ISSN 2655-3252


[23] Departemen Pendidikan Nasional 2008 Panduan Analisis Butir Soal. (Jakarta: Depdiknas)


[26] Basuki I dan Hariyanto 2014 Asesmen Pembelajaran. (Bandung: PT REMAJA ROSDAKARYA)