Developing performance instruments assessment of 5th grade elementary school students in integrated thematic learning

Ifan Awanda1*
Alben Ambarita2
Rochmiyati3
Dwi Yulianti3

1,2,3 Master’s Program in Teacher Training for Elementary School Teachers, University of Lampung, Indonesia

ABSTRACT
This study aims to develop an instrument for assessing student performance in elementary schools. The type of research and development used refers to the R&D theory of Borg and Gall. This study was conducted in three schools in Bandar Lampung. The sample data of this research were taken from the fifth-grade students of SD Negeri 1 Perumnas Way Halim. For data collection, observation, questionnaires, and documentation were employed, with descriptive analysis of the percentage of expert and practitioner validation, followed by instrument analysis for validity and reliability. This finding indicates that the assessment performance instruments of students in elementary schools were developed with theoretical and practical feasibility. Suggestions and limitations for future research are also discussed.

KEYWORDS
Instrument assessment; student performance assessment; R & D

Introduction
Learning activities consist of educators, students, objectives, materials, facilities, strategies and assessments. Student performance assessment instruments are needed in 21st-century learning because they require students to solve problems according to their environment and become an ideal generation capable of facing all the challenges of the 21st century. In addition, they are solving problems according to their environment and becoming a perfect generation capable of meeting all the challenges of the 21st century. The assessment process carried out so far also emphasizes the mastery of concepts needed with the objective and subjective written tests as measuring tools. This is supported by research by (Rustaman et al., 2005), who stated that the tests carried out so far only measured mastery of the material and covered low-level cognitive domains. This kind of situation is one of the reasons why educators are reluctant to carry out learning activities that focus on developing children’s attitudes and process skills.

The assessment component is an integral part of the learning process. Because the assessment of educators will measure the mastery of student competencies, it can also be used as feedback for educators to improve the learning process to be more effective. In addition, educators can also use the assessment results as a consideration for making decisions about students. According to (Miller et al., 2009), assessment is a general term that includes various procedures used to obtain information about students. According to (Kunandar, 2014), performance assessment assesses actions that can effectively collect different information about the forms of behavior or skills expected to appear in students. (Kartowagiran & Jaedun, 2016) state that performance assessment is an alternative assessment that provides multidimensional assessments in real situations and is authentic.

(Kurniawan, 2014) states that performance appraisal is an assessment of the quality of the work process in completing the learning tasks that are done. Furthermore, (Kurniasih & Sani, 2013) explain that performance appraisal is an assessment that asks students to perform a task in an actual situation that applies knowledge and skills. (Kunandar, 2014) explains that performance assessment is an assessment that asks students to demonstrate and apply knowledge in a context that is established criteria. Performance assessment is done by observing the activities of students in doing something. Meanwhile, (KomalaSari, 2013) revealed that performance appraisal is an
assessment carried out by observing the activities of students in doing something. Meanwhile, (Asmawi & Nasution, 2001) emphasizes the need for performance assessments to measure aspects other than cognitive, namely the seven basic abilities. According to (Gardner, 1999), it is impossible to judge only in usual ways. The seven essential skills are: visual-spatial, (2) bodily-kinesthetic, (3) musical-rhythmic, (4) interpersonal, (5) intrapersonal, (6) logical-mathematical, (7) verbal-linguistic.

There are five research results related to the variables in this study. The research conducted by Siti Fatonah et al. with the title "Developing and Authentic Assessment Model in Elementary School Science Teaching" in this study met the criteria as a good instrument in terms of reliability. The instrument for observing the performance of class IV has a trustworthiness of 0.794. This instrument is considered a reliable category. The mechanism for class V performance observation has a trustworthiness of 0.834; This is also considered a dedicated category. For the instrument of observing students' character, it has a reliability of 0.82, including reliability. The scientific attitude questionnaire instrument has a reliability of 0.94, so it is a reliable criterion. Therefore an authentic assessment model for teaching science in elementary schools (SD / MI) can be used.

Research conducted by (Irwanysyah et al., 2013) with the title of developing an instrument for assessing student performance at SMA Negeri 17 Makassar based on the 2013 curriculum in the theoretically correct category of 13 performance rubrics. In terms of format, content, language, benefits of performance assessment tools, and time rationality. Development of performance assessment instruments for SMA Negeri 17 Makassar students based on the 2013 curriculum in the empirically correct category as many as 12 performance rubrics from 13 performance rubrics with a reliability value of 0.83.c. The results of the teacher's response to the instrument developed were included in the very high category with a percentage value of 81.25%.d. at the time of the limited trial by the researcher and 48.21 at the time of the little trial by the teacher, which was included in the high category.

Next research conducted by (Pantiwati, 2013), with the title "Authentic Assessment for Improving Cognitive Skills, Critical Creative Thinking and Meta-Cognitive Awareness". This study indicates that authentic assessment is inseparable from integration with learning activities and contributes to students. Accurate estimate improves thinking skills because of the principles of developing meta-cognitive awareness. One of the principles of authentic assessment is continuous assessment; learners are continuously encouraged to monitor and control their studies. Learners are trained to develop an understanding of meta-cognitive thinking continually, and this is to improve their thinking skills.

Research conducted by (Al-nouh et al., 2014) results shows that educators can carry out alternative assessments. However, educators still find it challenging to carry out alternative checks, so educators still need workshops and training on alternative reviews. The ability of educators who can carry out alternative checks shows that there are differences in the level of education (bachelor), age and experience.

Furthermore, research conducted by (Almuflichan & Tjalla, 2016) The results showed that the theoretical validity test on the instrument was carried out by three experts and 30 panellists, which was intended to test the suitability between variables, dimensions, indicators, and descriptors. The results of this revised expert study obtained 20 items with three sizes and nine hands and 100 performance standards developed from descriptors which will then be carried for the second stage of theoretical testing. Furthermore, after the draft instrument was revised, an academic validity test was conducted on 30 panellists. In the educational validation activity, the statement items in the instrument are assessed based on five criteria, namely: score 1 with terrible measures (wrong concept or concept has no relationship with descriptor), score 2 with poor standards (correct idea, incorrect language formulation, not core performance descriptor), score 3 with sufficient measures (right concept, wrong discussion formulation, core descriptor performance), score 4 with good standards (correct concept, correct language formulation, descriptor core performance) and cut 5 with perfect criteria (right concept, correct language formulation, core performance descriptor). The results of the panellist test that 20 items with three dimensions and nine indicators, and 100 performance standards, show that all items are valid and can be used for empirical testing. In the first practical test with 20 articles, three dimensions and nine indicators and 100 performance standards.

Based on the description above and a preliminary study conducted at a school in Bandar Lampung, researchers see the importance of government policies on standards and educator problems, namely that there are no instruments according to the development of 21st century students, so researchers are encouraged to develop instruments that are appropriate to 21st century developments. an assessment instrument that includes grids, assessments, and rubik's according to the development of students in the 21st century.

To see the development of student performance assessment instruments, the objectives in this study are 1) Produce performance assessment instruments for fifth-grade elementary school students that are theoretically feasible in integrated thematic learning. 2) Produce practical performance assessment instruments for fifth-grade elementary school students in integrated thematic learning.

**Methods**

The development procedure in this study refers to the Borg and Gall theory, namely, research. The development of the Borg & Gall model has the following steps: 1) research and information collecting, 2) planning, 3) develop a preliminary form of product, 4) Desk Evaluation, 5) primary product revision, 6) main field testing, 7)
By the ten steps of implementing research and development, in this study, the researcher only carried out steps one to seven, namely research and information collecting (introduction) to Operational product revision (final product based on input from the primary field test). Steps eight to ten were not carried out because of time constraints and expensive costs for developing research products, and research standards for thesis requirements indeed carried this out. Population and Research Participants

This research was conducted in the fifth grade of SD Negeri 1 Perumnas Way Halim, Indonesia. The research subjects were 30 students. The researcher determines the sample in this study using the Cluster sampling technique.

Data collection technique

This study uses basic data collection techniques in the form of questionnaires and observations. Questionnaires were used to collect material experts, evaluation experts, linguists, media experts, and fifth-grade teachers. At the same time, observations were used as performance assessment data for fifth-grade students. Data collection activities were obtained from questionnaires and observations based on the developed guidelines, based on aspects that have been observed operationally based on indicators of student performance assessment.

The analysis can be assessed from the practicality of the instrument, which includes three aspects, namely attractiveness, convenience and usefulness. The study of the practicality test was obtained from a questionnaire that educators and product validation tested. This analysis involves several stages, including recapitulation of statements obtained from the validator, finding the average level of validity for each criterion, calculating the average score for each aspect, calculating the total average, and matching the total average with the validity category. Categories that have been determined, as can be seen in table 1.

Table 1. Criteria for the validity of material experts, evaluation experts, linguists, media experts, and educators

<table>
<thead>
<tr>
<th>Criteria Interval score (%)</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>82% - 100%</td>
<td>Very good</td>
</tr>
<tr>
<td>63% - 81%</td>
<td>Well</td>
</tr>
<tr>
<td>44% - 62%</td>
<td>Not good</td>
</tr>
<tr>
<td>25% - 43%</td>
<td>Not good</td>
</tr>
</tbody>
</table>

Source: Anas Sudijono (2008)

Based on the calculation of descriptive analysis, the assessment developed is declared feasible if it gets a score of > 62%. Measurement of validity for the significant group test was carried out using the Rasch application. First, a sample adequacy test was carried out by looking at the Kaiser Mayer Olkin (KMO) value. The KMO value is obtained through analysis with the help of IBM SPSS 20 software. If the KMO value is more than 0.5, the variables and samples are used to allow further research (Farhan & Retnawati, 2014).

The data analysis technique for estimating the reliability of large group test instruments uses an internal consistency estimation technique with the Chronbach-alpha formula assisted by the Rasch application. Suppose the value of Chronbach’s Alpha is 0.60 and less than 1. In that case, the instrument’s weight has a high or reliable correlation, whereas if the value of Cronbach’s Alpha is below 0.50 and below, then the instrument has a low or unreliable correlation.

Results

This study obtained an assessment instrument for fifth-grade elementary school students theoretically in integrated thematic learning. They resulted in practical performance instruments for fifth-grade elementary school students in integrated thematic education. The research results are presented in tabular form in stages. Previous research has been carried out in assessing integrated thematic learning for fifth-grade elementary school, which was carried out with a field survey stage in Bandar Lampung. This is done to collect information.

Furthermore, a plan for developing appropriate instruments is made for fifth-grade elementary school students. After the initial design was made in the form of student performance assessment instruments, product tests and revisions, field tests and valid student performance assessment instruments, this development is carried out through literature study to learn concepts related to the instrument. The literature study stage was obtained by preparing product drafts and (Sukmadinata, 2016).

The 5th-grade elementary school student assessment instrument developed has been validated by material experts, evaluation experts, linguists and design experts. Development refers to the Borg and Gall model and is limited to step 7. After four experts validated the student performance assessment instrument, the following scores were obtained. In the first stage of the preliminary study, the initial data results showed that most of the educators had not used the instrument according to the 2013 curriculum guidelines, which contained grids, core competencies, essential competencies, indicators, and objectives. In addition, all educators have not used the performance appraisal instrument as described above, which refers to the 2013 curriculum.
The next stage is planning the development of performance assessment instruments in the form of curriculum analysis and themes, core competencies, essential competencies, and performance task indicators. The prototype results are in the form of product covers that are by the themes and subthemes that will be examined in the development of performance instruments, then in the form of KI and KD mapping and indicators product grids for student performance assessment instruments.

The third stage is the initial product development. Appropriate themes are “Clean Air for Health” and “how the body processes clean air.” The instrument for assessing student performance from lessons one to six consists of; Core Competencies, Theme Networks, Authentic Assessment Instruments Grid (Performance/Performance), Steps for Student Performance Assessment Activities, Performance Assessment Scoring Guidelines.

The next stage is the initial trial stage, in the form of expert assessment. Expert assessment is carried out by submitting the product of the student performance assessment instrument. Then the expert was asked to assess the tool's suitability with the indicators on the validation sheet. Aspects set during this initial field trial included expert validation by lecturers (evaluation experts, materials, language) and the responses of educators and students in small group trials. Boba test data can be seen below.

Table 2. Recapitulation of results Validity of material experts, evaluation experts, linguists, media experts

<table>
<thead>
<tr>
<th>Value Validator (%)</th>
<th>Material Expert</th>
<th>Evaluation Expert</th>
<th>Linguist</th>
<th>Media Expert</th>
<th>Average</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>81%</td>
<td>87%</td>
<td>83%</td>
<td>85.25%</td>
<td>Very Worthy</td>
</tr>
</tbody>
</table>

Source: Data Analysis

Based on expert analysis calculation analysis, the assessment developed is declared feasible if it gets a score > 62%. The validation results by experts who assess the product design of the student performance assessment instrument met the criteria with an average score of 85.25 and got very decent standards. During the small group trial, the requirements were efficient, with an average percentage of 88.75. Then continued with the response test of significant group educators obtained data such as table 3 below.

Table 3. Recapitulation of the results of the questionnaire analysis of the practicality of educators

<table>
<thead>
<tr>
<th>No</th>
<th>Rated aspect</th>
<th>Educator 1</th>
<th>Educator 2</th>
<th>Educator 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>attractiveness</td>
<td>11</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Convenience</td>
<td>10</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>Usefulness</td>
<td>15</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Total score</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Maximum Score</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Score Percentage</td>
<td>90.00%</td>
<td>87.50%</td>
<td>87.50%</td>
<td></td>
</tr>
<tr>
<td>Average Percentage</td>
<td>88.33%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Practical Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Research analysis

This shows that the results of the significant group test of educator responses for the practicality test of educators obtained an average score of 3 educators with a percentage of 88.33%, including the efficient criteria.

Furthermore, in the fifth section, the initial product revision is carried out. The revision includes material, evaluation, language, and design. After being improved from the previous step, the main field trial was carried out in the sixth part. This instrument was tested at the target school with a total of 30 participants. This primary field trial was given treatment in the experimental class to find out the instrument was valid and reliable in assessing student performance.

Data Validity Test

The validity test of the performance appraisal data consisted of 30 items from observations of student performance assessments in large classes. The validity test was carried out to determine the validity, validity and level of confidence of the instrument. The data were analyzed using explanatory factors, but first, a sample adequacy test was carried out by looking at Kaiser Mayor Olkin (KMO) using SPSS 20.
Table 4. Major Olkin Kaitser Value (KMO)

<table>
<thead>
<tr>
<th>KMO and Bartlett's Test</th>
<th>KMO and Bartlett's Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</td>
<td>.065</td>
</tr>
<tr>
<td>Bartlett's Test of Sphericity</td>
<td>Approx. Chi-Square</td>
</tr>
<tr>
<td>df</td>
<td>700,151</td>
</tr>
<tr>
<td>Sig.</td>
<td>.000</td>
</tr>
</tbody>
</table>

Source: Results of primary data processing

The KMO value in the collaboration skills data shows a value of 0.625 > 0.5. This means that the assessment items observed on the observation sheet are good to use. The value of Bartlett's Test of Sphericity with Chi-Square 700,151 > df 435 and Sig. 0.000, it can be concluded that the assessment items observed on the observation sheet are valid.

Table 5. Reliability Test Results

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s Alpha</td>
<td>Instruments</td>
</tr>
<tr>
<td>Performance rating,69</td>
<td></td>
</tr>
</tbody>
</table>

Source: Results of primary data processing

Based on the table above, the Cronbach’s alpha reliability test results on the collaboration skill instrument show 0.69 > 0.5. It can be concluded that the instrument for assessing performance is reliable or consistent. The last part is the improvement of the product as the result of testing the data obtained is reliable. So the effect of the developed student performance assessment instrument is not revised and is feasible to be implemented.

Discussion

Theoretically Feasible

The theoretical feasibility of the student performance assessment instrument is seen from the results of the assessments of four experts, namely material experts, evaluation experts, linguists, and design experts. Based on the results of the evaluations of four experts, this assessment instrument is theoretically feasible because each validation expert obtains, among others, the validation of the material expert receives a value of 90%, where the aspects of the instrument follow the basic competencies to be assessed, the material on the instrument presented is following competence (urgency, relevance, continuity, and usability in everyday life). Evaluation experts get a score of 80% where the aspects of the indicators assessed have met the criteria to be achieved, using clear instructions on how to evaluate the performance to be assessed so that it is easy to use. The instrument presented contains a process and product scoring rubric.

Meanwhile, linguists got a score of 87% where the specific aspect where the accuracy of sentence structure, sentence effectiveness, and word choice had met the criteria, in the communicative element, it was stated that the readability of the message, the accuracy of the use of language rules on the instrument was easy to understand, in the writing aspect where the composition of the font size was in writing has met the criteria. Finally, the design expert scored 83%, where the images used in the instrument, the shape of the letters, and the colour gradations used were fascinating. The average expert score as a whole is 85.25 in very decent criteria.

The assessment used in the 2013 Curriculum emphasizes the facts that students do when learning takes place or commonly called authentic assessment. Authentic assessment (authentic assessment), according to the Curriculum Center in (Safitri et al., 2019), namely: Authentic assessment (authentic assessment) is a process of collecting, reporting and using information about student learning outcomes by applying the principles of assessment, continuous implementation, precise, accurate, and consistent evidence as public accountability. Authentic assessment emphasizes the ability of students to demonstrate accurate and meaningful knowledge. Assessment activities do not just ask or tap knowledge already known to the learner but perform from the knowledge and skills mastered. Based on the description above, it can be stated that the scope of authentic assessment includes attitude, knowledge and skill competencies. The authentic assessment target used in this study is the original assessment target based on the preparations of Krathwohl, Anderson, and Dyers, which have been listed in the Attachment Copy. Researchers in conducting this study limit only authentic assessment of psychomotor competence, namely in students' performance assessment.

(Kunandar, 2014) the steps in carrying out a performance appraisal are: 1) Delivering the rubric before implementing the assessment to students and providing understanding to students about the assessment criteria. Deliver assignments to students. 2) Checking the availability of tools and materials used for performance tests. 3)
Assess the planned time. They are comparing student performance with the assessment rubric. Record the results of the assessment. 4) Documenting the results of the assessment. The steps in carrying out the skills competency assessment through the performance assessment are used as guidelines by researchers in conducting research.

In line with researchers who develop products for assessing student performance, Performance/Performance/Practice Assessments can be used to determine students’ understanding, ability to apply, inquiry ability and ability to inform students. According to (Kurniawan, 2014), performance appraisal is an assessment of the quality of the work process in completing the learning tasks that are done. Meanwhile, (Komalasari, 2013) revealed that performance appraisal is an assessment carried out by observing the activities of students in doing something.

**Practicality**

The practicality of the performance assessment instrument is seen from the results of small group trials and large group trials. This practicality assessment uses a practicality response questionnaire by educators and students consisting of attractiveness, convenience, and usefulness. The results of the practicality response of educators by five educators, in small group educators totalling two educators, get an average score of 88.75 and in prominent group educators totalling three people get a score of 88.33, and from the calculations of prominent group educators and small get an average percentage of 88.54% and meet the efficiency criteria. While the results of the practical response of students in which a small group of 10 people with an average value of 90.83 meet the efficiency criteria. The reactions of educators and students to the questionnaire stated that the assessment instrument developed had benefits, especially with assessing student performance.

**Conclusion**

It can be concluded from this research that the assessment instrument developed theoretically and empirically is feasible to measure the knowledge of fifth-grade elementary school students. The expert test stated that the test instrument developed was in the "perfect" category. In addition, this assessment instrument also meets the criteria of empirical validity in the validity and reliability test.

Measurements using theoretically and empirically feasible observation sheets can help educators measure essential competencies in the curriculum that the government has set. The implication is that learning in elementary schools where the research is conducted has not yet led to performance appraisals. This happens because educators do not understand the assessment of student performance. So that in the end, the instruments made by educators cannot train students in honing their abilities and creativity, and it is complicated to distinguish between students who can and those who cannot.

This research is limited to only being carried out in three elementary schools in Bandar Lampung, so a more significant similar research is needed, perhaps in one sub-district, so that future research results can be considered in determining the policies of a government.

**Acknowledgements**

The author would like to thank the respondents of high school teachers throughout Way Kanan district who have taken the time and fully supported this research.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**References**


