Authentic assessment instruments for performance in mathematics learning in elementary schools

R P Yudha\textsuperscript{1,\*}, D S Anggara\textsuperscript{2} and O Zulaeha\textsuperscript{3}

\textsuperscript{1}17 August 1945 Cirebon University, Indonesia
\textsuperscript{2}Pamulang Univesity,Indonesia
\textsuperscript{3}STKIP Kie Raha Ternate, Indonesia

*Corresponding author: Rivoyudha@yahoo.co.id

Abstract. Products in the form of assessment instruments are packed in book form. Guidance on the use of authentic assessment instruments of the performance of elementary school students that contain; 1) assessment purpose; 2) the division of the syllabus; 3) lattice; 4) observation sheet; 5) rubric. This research is an R & D development model using Borg and Gall analysis that modifies researchers into three stages including: (1) introduction; (2) product development (development); (3) Presentation of product development result. Research subjects at SD Nurrusshidiq city Cirebon and SDN Karanganyar Cirebon City. Validity uses expert validation from four assessors with expert valuation techniques. Performance assessment comes from three evaluation experts, a mathematical instrument and one senior math teacher. (1) An overview of existing instruments in Cirebon Elementary School in Mathematics learning is that teachers still use the description of the tests in the classroom assessment process. (2) The stage of development of this study consists of the initial stage that contains the needs analysis, grid making, the manufacture of performance appraisal instruments, development stage consisting of expert validity test, small-scale test, and stage presentation consisting of large-scale experiments and test the practicality of the instrument. (3) Experimental results of the assessment.

1. Introduction

Traditional assessment (TA) refers to multiple-choice tests, fill-in-the-blanks (supply type), true-false, matching, short-answer type and the like. Students typically select an answer or recall information to complete the assessment. These tests may be commercially available or teacher-made. Performance assessments call upon the examinee to demonstrate specific skills and competencies whilst traditional assessment requires the examinee to recall facts to help select an answer from a set of options.

Authentic performance tasks call for students to engage in complex, problem-solving processes with multiple peers over an extended period, and such tasks have the potential to support and foster adolescents' cognitive development, in addition to the obvious ways such collaboration works to enhance the natural social development of young adolescents. The complexity of authentic performance tasks reduces the likelihood that one group member will possess all the cognitive capabilities to complete the task independently [1]. Rather, group members learn that their success relies on all members assuming responsibility for sharing their knowledge and problem-solving processes [2], thereby working to ease some of the isolation that too often characterizes the middle grades. Throughout debate and discussion, students have opportunities to recognize, clarify, and modify inconsistencies in their own and in each other's knowledge or to fill in gaps in understanding [3].
Authentic performance tasks have five common features [4]. First, the prompt is a real-world problem or scenario intended to increase and sustain students' engagement. Second, students engage in authentic complex processes such as mining research, establishing a plan of action, and evaluating competing ideas. Third, authentic performance tasks promote the use of higher-order thinking skills by including activities that encourage students to analyze, synthesize, and evaluate information. Fourth, students construct an authentic product or performance such as a wall display, book, poem, song, demonstration, or presentation. Finally, student performance is assessed using detailed rubrics to evaluate processes and products. Rubrics are provided to students with the problem statement so that students are also able to self-assess their work over the course of the project.

Conventional assessment of student achievement historically has focused on the reproduction of factual and procedural knowledge from students [5]. The items on such assessments typically measure recall of discrete facts, retrieval of given information, and application of routine computational formulas or procedures [6]. But while 'snapshot' conventional assessment results give a partial picture of students’ performance at a given moment, performance assessment depicts a comprehensive view of the student’s performance at a given time [7].

Most elementary school mathematics teachers here are particularly uninterested and do not want to use authentic assessment or performance appraisal. It is generally thought that doing authentic judgments is a waste of time and energy and too expensive, let alone an authentic assessment needs to be well designed. That particular opinion is incorrect. Assessing performance with a written test is certainly not valid, as it does not measure what it wants to assess, Performance needs to be assessed as the activity is in progress. If the performance appraisal is done to a number of students and not designed first or done carelessly, of course the result can not be accounted for because it is not consistent. Thus we may be unfair to some students in assessing their performance. Designing and executing performance appraisals is very efficient, because it is steady or consistent (read reliabel), not expensive and wasted no time. Standards can not be created without performing performance-based assessments [8].

The problem that often faced by primary school mathematics teacher next is in performing assessment of performance lies in validity and reliability of measuring instrument used. The preparation of student performance tests is still very limited to the teacher's knowledge and understanding of simulated tests. Assessment results are often influenced by teacher objectivity as rater because in conducting self-assessment without involving other teachers as collaborators.

The most frequently used format of teacher-made tests for formative and summative purposes in the classroom is the multiple-choice type. The over dependence on the use of objective tests in the classroom has affected more than the form of subject-matter knowledge. Objective-type tests make use of items at the knowledge level rather than more cognitively complex levels. Therefore, a broader range of assessment tools are needed to capture important learning goals and processes and to connect assessment to ongoing instruction. Instruction and the use of performance assessment tasks. When these changes are effected it will offer the students an opportunity to reason critically, to solve complex problems and to apply this knowledge in real-life situations. Performance assessment in mathematics is primarily concerned with connecting classroom learning to the real world applications of mathematical concepts. However, with the influx of question and answer.

The view that performance assessment should enhance the validity of measurement by (a) representing the complete range of desired learning outcomes, (b) preserving the complexity of disciplinary knowledge domains and skills, (c) representing the contexts in which knowledge must ultimately be applied, and (d) adapting the modes of assessments to enable students to show what they know [9].

Based on the above description can be concluded that the application of validity assessment instruments that are valid reliable, practical and can be used repeatedly on different performance tasks, can help teachers to perform assessment of student performance at the lab in laboratory. Product development of assessment instrument of laboratory laboratory in physics field, before decided to be
used by teachers need to be conducted research. The goal is to obtain information whether the product developed is valid, reliable, and practical.

2. Method

This research is a development research with the aim of producing product in the form of performance assessment instrument. The development model used in this research is the R & D development model. Procedures in this development include: (1) Introduction (Define); (2) Planning (3) Developing the initial product (Develop)¹, (4) Preliminary trial; (5) First revision; (6) major field trials; (7) second product revision; (8) Test of operational product; (9) Final Product Revision; (10) Presentation of the final product (Deliver) [10].

Based on the ten steps of research and development developed by Borg & Gall, in this research on the implementation process adaptation that refers to the approach model. The development adaptation of this research outline consists of three main stages of activity, namely: (1) introduction; (2) product development (develop) and (3) presentation.

The introduction stage consists of three main activities covering the needs analysis activities, the design of the grille and the manufacture of performance assessment instruments. Needs analysis activity aims to reveal the real condition of mathematics teachers in performing assessment of student performance, especially in current mathematics learning. Need analysis in the form of survey result by using interviews conducted on elementary school mathematics teacher. As the subject of research in the requirement analysis phase, it is found that the constraints of the assessment of the performance of mathematics learning activities in primary schools, one of the obstacles is that teachers still do not understand the scanning guidelines in unclear instruments that are difficult to use, the components considered difficult to observe, tend to be ignored.

Development stage (develop), at this stage the instrument has been designed consulted to supervisor lecturer. Expert or Validation test, carried out by respondents of model or product design experts. This activity is carried out to review the initial product providing input for improvement, this validation process is called Expert Judgment. The generated instrument is evaluated, whether the resulting format is feasible or not, and how the content of the learning assessment material matches. If the instrument is not feasible, then the instrument is revised again so that the instrument becomes feasible to be tested. Prior to the trials, validation of the instrument was performed by four evaluation experts, then the instrument was tested to assess the learner's performance during small-scale learning, ie, on the fifth grade students on a small scale. It aims to find out whether the instrument is feasible to use or not to know how the performance of learners. The results of trials on the use of instruments in the fifth grade are used as references for the further development and improvement of the instruments.

The presentation stage (deliver), at this stage is tested more widely in different schools with the same class that is class V and on a larger scale. Real product trials are conducted to assess the performance of class V learners at the time of learning. the result of this phase is the conclusion of success or failure of product design developed for the benefit of users and of the team involved.

Instrument data used in this research are as follows: (1) Rating scale (Rating Scale), scale scale used to see indicator of psychomotor skill in conducting practicum activity. This multilevel scale contains about student activities in the form of skills to be observed. Psychomotor skills that are observed include the preparation of tools and materials, drawing space nets, shaping space, and smoothing tools and materials. This multilevel scale is filled by observers who observe all students; (2) Rubric, Rubric is used as a scoring guide that describes the criteria that teachers want in assessing or assigning levels of student work. The rubric contains a list of desired characteristics that need to be demonstrated in a student's work along with a guide to evaluate each of these characteristics. The purpose of the rubric assessment is that students are expected to clearly understand the basic assessment that will be used to measure a student's performance. Both parties (teachers and students) will have clear shared guidance on expected performance demands; (3) Teacher Interview Guidelines, interview guidelines in this study were used to collect deeper data on the practicality and effectiveness
of developed assessment instruments, either in relation to the instrument used, the assessment exercise technique, or other matters not revealed through the scaling and rubric; (4) Testing Instruments, as well as interviews, tests also provide samples of individual behavior, only in stimulus tests that respondents respond more standardized than interviews. Standardized test forms help to reduce the biases that may arise during the performance appraisal process. The responses usually can be changed in the form of scores and quantitative analysis. It helps the supervisor to understand the respondents. The scores obtained are then interpreted according to the norm; (5) Observation Observation Sheet, Observation Guidance in performance assessment on mathematics learning is used to observe and evaluate each student participant of the performance test by using a rating scale and its assessment weight. Arrangement of observational instrument constructs in the form of scale of assessment based on material that reflects the skills to be measured. Next is determined the scale of assessment for each material. In this study used a scale of four, namely one to four (1-4).

3. Results and Discussion
First, in the analysis needs, it is found that the constraints of performance appraisal of the performance of mathematics learning activities in primary schools, which are derived from interviews using interview sheets of ten elementary school mathematics teachers in Cirebon city. One obstacle is that teachers still do not understand the scanning guidelines in unclear instruments that are difficult to use, those that are considered difficult to observe, and thus tend to be ignored.

The results of interviews from ten teachers related to the assessment method to measure the students' competence in the building materials used by teachers include: Performance tests, observations, written tests. Based on his experience, in fact every teacher has different priorities on the seven methods of assessment. The teacher's response to the assessment method is summarized in the following table.

| No | Question | Answer |
|----|----------|--------|
| 1  | Is the process of learning mathematics, teachers use performance appraisal? If not explain why! | Not yet using, not too understand the procedure of making the appropriate performance assessment instruments |
| 2  | How often do teachers perform performance appraisals in learning? Explain! | Never because doing an authentic judgment makes time and energy and too expensive, let alone an authentic assessment needs to be well designed. |
| 3  | What material often uses performance appraisal in mathematics learning? | Programe lineair, aljabar, geometry |
| 4  | Where do teachers get the performance appraisal guidance used? | Adaptation for package book and RPP |
| 5  | What aspects are assessed in the learning activities of mathematics? | If done normally, if using appraisal performance only skill aspect |
| 6  | How does the teacher assess learners' performance in mathematics learning? | Observation |
| 7  | Is there a performance appraisal instrument as a reference that teachers can use to assess learners' performance in math learning? • If yes, go to no. 9 • If not, go to no. 12 | Yes there is, performance assessment instruments just still do not understand in detail |
| 8  | What the assessment instrument used is specific to each subject or is still general | Not because the components considered to be difficult to observe |
| 9  | What is the assessment instrument used? | The instrument is in the form of a statement description only, the |
instrument has not been using clear guidelines. Scoring only by teacher's subjectivity.

10 If using an assessment instrument scale, is it accompanied by a rating rubric? Already

11 Do you think it is necessary to develop an authentic assessment instrument for performance on mathematics learning? It is necessary, in order to facilitate the teacher in doing the assessment especially to assess the skill aspect.

12 What is the expected form of assessment instrument? Expected instruments include observable and realistic aspects of skill when used in the field.

Second, the next step is the preparation of the instrument grille, evaluation tool developed in this research is the performance assessment instrument, to assess the performance skills of students in learning mathematics Space waking. Preparation of the grid of this class performance appraisal instrument referring to the Core Competence (KI), Basic Competence (KD), further can be seen in table 2

Tabel 2. Latticework of Instrument of performance assessment

| Basic competencies | Assessment indicators | Assessment Technique | No. Item |
|--------------------|-----------------------|----------------------|----------|
| Identify the properties of simple space builds | 1. Draw geometry from the given object properties | Geometry | 1 |
| | 2. | | 2 |
| Determining the webs of various simple geometry | 2. Make the webs of geometry | Geometry | 3 |
| | 4 |

Third, after the grating is made the next step is the preparation of performance instruments. Assessing performance instruments, teachers should prepare at least 2 documents, namely: 1) Problem 2) Observation instrument or observation sheet is a checklist or rating scale. The observation sheet here is an instrument used to observe the appearance of observed aspects of performance skill. The observation sheet here is a rating scale. The rating scale is a list of questions or statements to assess the quality of implementing observed skill-ranging aspects 1-4.

1. On the paper you draw a cube ABCD EFGH with the following provisions:
   a. The length of the ribs is 5 cm.
   b. The ABFE side is visible up front
   c. All the ribs are visible
   d. ABCD side in shading
2. On the paper you picture the PQRS TUVW beam with the following:
   a. PQ side 5 cm, and the side of PT 2 cm, while the QR side arbitrary
   b. PQUT side appears in front
   c. All the ribs are visible
   d. PQRS side is shaded
3. Make a space-building net.
   a. Look at the following spatial forms!
   b. Make a net for every wake up space at least 3 shapes and a maximum of 8
   c. Do it thoroughly!
4. A Block made of cardboard with length 5 cm, width 4 cm and height 3 cm. Draw two different webs of blocks of that size.
### Tabel 3. Rubric of number 1 dan 2

| No. | Step          | Indicator                                                                 | Aspects of Assessed / Developed                                                                 | Scoring guide                                                                 |
|-----|---------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| 1   | Preparation   | Draw plane figure and solid figure from the given properties             | • Provide tools and materials according to what is required to draw a plane figure and solid figure | Tools: Pencils, rulers, erasers, colored pencils, pens Material: Striped or striped folio paper |
|     |               |                                                                          | • Take the appropriate tool to make a plane figure and solid figure                          | Pencils, erasers, colored pencils                                               |
| 2   | Implementation| Draw plane figure and solid figure from the given properties             | • Drawing is done according to the order of stages of the provisions                       | 4 cm long edge, all edge appear, abcd shaded side, side ABCD shaded, coloring cube |
|     |               |                                                                          | • The image process is done individually                                                  | Drawing is done alone without the help of teachers or friends.                   |
|     |               |                                                                          | • The image process is done well and correctly                                            | The drawing process is in accordance with the terms of the task                  |
| 3   | Product Results | Draw plane figure and solid figure from the given properties            | • The results of the drawings are completed on time                                         | The image process is dismounted exactly 10 minutes                                |
|     |               |                                                                          | • The image results are neatly arranged and true to the specified size                      | The results are neat and in accordance with the provisions                      |

### Tabel 4. Scoring guidelines of number 1 and 2

| Step          | Description                                                                 | Score                                                                                           |
|---------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Preparation   | Preparation of materials / tools, step / step work, taking the tool used    | 4 = if carrying the complete tools and materials and picking up the equipment to suit the needs |
|               |                                                                             | 3 = if only carry and use 5 tools                                                                |
|               |                                                                             | 2 = just carry and use 3 tools                                                                    |
|               |                                                                             | 1 = carry less than 3 and do not use tool that suits tool needs                                   |
| Step               | Description                                                                 | Score |
|--------------------|-----------------------------------------------------------------------------|-------|
| Implementation     | Specific abilities needed to complete tasks, such as drawing accuracy and implementation process | 4 = if acting in accordance with the order of provisions, done alone and all according to the provisions<br>3 = if the sequence is done according to only 4 sequences done a-e sequence, and done with the help of the teacher, and only 4 provisions are made<br>2 = if in order but only 3 sequences and assisted by friends, and only 3 provisions are made<br>1 = if the execution is not in the order specified, and assisted by the teacher and friends, therefore not in accordance with the prescribed conditions. |
| Product Results    | The accuracy and completeness of aspects to be assessed, such as accuracy of results, timeliness, and neatness | 4 = Completed the product is less than or exact.<br>3 = Completed more than ten minutes or less than or equal to 15 minutes<br>2 = Completed more than 15 minutes<br>1 = product not completed |

| Score              | 0-10 |

**Tabel 5. Rubric of number 3**

| Step               | Description                                                                 | Score |
|--------------------|-----------------------------------------------------------------------------|-------|
| Chart              | Images and descriptions of the graph presented are correct                  | 4     |
|                    | Most of the images and postings given are correct                           | 3     |
|                    | Some pictures are presented and some are true                               | 2     |
|                    | The images and postings provided are very limited and only partially true   | 1     |
| Specification      | All specifications provided are correct according to the terms              | 4     |
|                    | All the specifications given are almost all true                           | 3     |
|                    | All of the specifications are partially true                               | 2     |
|                    | The specs given are generally incorrect                                     | 1     |
| Total              | Make 10 nets or more made each Solid Figure                                | 4     |
|                    | Create 6-9 specified nets                                                  | 3     |
|                    | Make 3-5 specified nets                                                    | 2     |
| Step | Description | Score |
|------|-------------|-------|
|      | Make less than 3 specified nets | 1     |

**SCORE**

### Tabel 6. Scoring guidelines of number 3

| Step  | Description                                                      | Score |
|-------|------------------------------------------------------------------|-------|
| Chart | Images and descriptions of the graph presented are correct       | 4     |
|       | Most of the images and postings given are correct                 | 3     |
|       | Some pictures are presented and some are true                     | 2     |
|       | The images and postings provided are very limited and only partially true | 1     |
| Spec  | All specifications provided are correct according to the terms    | 4     |
|       | All the specifications given are almost all true                  | 3     |
|       | All of the specifications are partially true                      | 2     |
|       | The specs given are generally incorrect                          | 1     |

Fourth, after the authentic assessment instrument of performance is made, the next step is the expert validation stage. At this stage the author is able to know the validity of the developed instrument. Because the instrument developed is an instrument that measures the performance and instruments including non-test, the validity test is validity of the construction validity. Test the validity of the constructs of the author using his opinion Purwanto stating that the test of construct validity can be done by requesting expert judgment, by using instrument validation sheet [11].

The use of non-test instruments that collect data in the form of narrative or nominal enough done with the validity of the contents or constructs. Content validity is intended to know the content and a measuring tool (the material, the topic, the substance) whether it is representative or not. In mathematics education research, especially related to learning activities in school, non test instrument used can be considered valid at least if it has fulfilled the validity of content obtained through expert judgment.

The results of the analysis becomes a requirement to test the product. Analysis in this research is by doing expert judgment by using Delphi technique to grid and grain of math problem to be developed. Delphi technique to determine the validity of the contents of the instrument to be developed. The next qualitative technique is done when the researcher performs the analysis of the data obtained through the interview activity, that is when the researcher performs the needs analysis activity, discuss with the teacher when determining the solution alternative, and test the feasibility of the model of performance assessment instrument developed.

The Delphi technique, the Delphi method is a systematic method of gathering opinions from a group of experts through a series questionnaire, where there is a feedback mechanism through the 'round' / round question being held while maintaining the anonymity of the respondents' responses.
The Delphi method is a structured communication technique, originally developed as an interactive forecasting method that relies on a number of experts [12]. Some formulas for obtaining instrument validity empirically include inter-rater reliability test through calculation of agreement coefficient between observer (rater), also called coefficient of concordance. The coordinate of the concordance is sought by the formula Ebel [13]. The concordance coefficient is acceptable at 3% significance level if the probability of error is < 0.03 (which is commonly used in social research, educational research). If the probability of error is greater than that provision, which means that between observers there is no matching observation, then the items assessed should be aborted and should not be used as research analysis material. In other words the item is not valid. In addition to the concordance coefficient.

At this stage the instrument has been designed to be consulted to experts. The generated instrument is evaluated, whether the resulting format is feasible or not, and how the content of the learning assessment material matches. The following instruments of expert judgment using delphi techniques and validation results from 4 evaluation experts are on table 7.

### Table 7. General expert instruments data

| No. | Expert | Work Experience (Year) | Position | Education |
|-----|--------|------------------------|----------|-----------|
| 1   | I      | 27                     | Practitioners and Academics | S3        |
| 2   | II     | 25                     | Practitioners and Academics | S3        |
| 3   | III    | 23                     | Practitioners and Academics | S3        |
| 4   | IV     | 30                     | Math Teacher              | S1        |

Based on table 7 the researchers selected 4 experts from different perspectives and with different criteria based on the researcher's wish but homogeneous according to their interests and relevance to the variables they wish to validate from academics, practitioners, and contents, to find the selected variables. Of the 4 experts will get comments or suggestions in the form of variable sentences research, addition and reduction of the number of variables, data processing, and so forth. Here are the experts who qualified researchers.

### Table 8. Summary of validation results of the performance appraisal instrument on mathematics learning

| No. | Rated aspect                        | Assessment Criteria | Expert Judgment | Scores | Quality     |
|-----|------------------------------------|---------------------|-----------------|--------|-------------|
| 1   | Conformity Aspects of Performance Assessment With Existing Indicators | 1 5 5 5 5 | 20 | Very Good |
|     |                                    | 2 4 5 4 5 | 18 |
| 2   | Writing                            | 3 5 4 4 4 | 17 | Very Good |
|     |                                    | 4 5 5 3 5 | 18 |
|     |                                    | 5 4 5 4 5 | 18 |
| 3   | Language                           | 6 5 5 4 4 | 19 | Very Good |
|     |                                    | 7 5 5 4 5 | 19 |
|     |                                    | 8 5 4 3 4 | 16 |
| 4   | Physical Appearance                | 9 5 5 3 4 | 17 | Very Good |
|     |                                    | 10 5 5 5 5 | 20 |
The results of the assessment of the four assessors of the authentic performance assessment instruments are summarized in Table 8. Aspects of assessment include: the appropriateness of performance assessment aspects with existing indicators, aspects of conformity with the indicators, writing, language aspects, and aspects of physical appearance.

From the experiments conducted to 3 instrument experts and 1 mathematician then taken the average percentage of test results that is for performance tasks class V, Percentage Ideal assessment instrument 91%, 95% Conformity Test, 88% Writing Test, 90 Language Test %, Test Physical appearance 92.5%.

In addition to the percentage of idealization of the general validation results of the four assessors of the assessment assessment instrument that can be seen from the value of expert judgment, who get a maximum score of 20 on the aspect of conformance indicators and physical appearance. In addition to the data presented in the table also obtained written data derived from the column notes and verbal data transcribed from the results of interviews with experts and practitioners about some of the main points of input of the four assessors are: (1) Procedure of writing the language still not quite correct eg merging or segregation of sentences; (2) The sentence on the instrument should be straight to the root of the problem, not long-winded; (3) The instrument's appearance is still less attractive: (4) Instruments should measure the specific competencies expected to emerge in the lesson; (5) The assessment rubric should be one if the same type of problem is to be more efficient.

Fifth, Implementation Small-scale trials were conducted with the help of 4 rater taken from math teacher. The stages in this first trial are as follows: (1) one day before the experiment, the researcher gives the instrument to the rater and explains the intentions contained in the indicator points; (2) Each rater receives a copy of the instrument and is requested to complete the assessment item with the member 1,2,3,4 which is the result of the assessment. This is done so that when the rater perform the assessment can avoid the error of interpretation of the point of assessment; (3) the rater conducts an assessment of the rater, and each rater evaluates or observes the number of students (3) the researcher holds discussions with the rater and asks for feedback on the assessment instruments used. Item performance instruments that have been prepared and used in this research, each of them as much as 4 performance test questions in which there are 4 rubric assessment has been tested the validity of its content and reliability using the coefficient of reliability assessment instrument performance assessment on learning mathematics material wake up space in Primary schools are conducted using the Genova coefficient (Generalizability of variance). the assessment includes: the appropriateness of performance assessment aspects with existing indicators, aspects of conformity with the indicators, writing, language aspects, and aspects of physical appearance.

From the experiments conducted to 3 instrument experts and 1 mathematician then taken the average percentage of test results that is for performance tasks class V, Percentage Ideal assessment instrument 91%, 95% Conformity Test, 88% Writing Test, 90 Language Test %, Test Physical appearance 92.5%.

In addition to the percentage of idealization of the general validation results of the four assessors of the assessment assessment instrument that can be seen from the value of expert judgment, who get a maximum score of 20 on the aspect of conformance indicators and physical appearance. In addition to the data presented in the table also obtained written data derived from the column notes and verbal data transcribed from the results of interviews with experts and practitioners about some of the main points of input of the four assessors are: (1) Procedure of writing the language still not quite correct eg merging or segregation of sentences; (2) The sentence on the instrument should be straight to the root of the problem, not long-winded; (3) The instrument's appearance is still less attractive: (4) Instruments should measure the specific competencies expected to emerge in the lesson; (5) The assessment rubric should be one if the same type of problem is to be more efficient.
The instrument reliability test using the most appropriate formulation observation sheet is by using Inter Rater Reliability, reliability involving rater is usually called by inter rater agreement or inter rater reliability. If in the case of self-report reliability indicated by the internal consistency seen from between one grain and the other grain has a high correlation, then in case of reliability between rater tested consistency is raternya. So the grain position is replaced by the person's position (rater). To assess the reliability between two or more observers, as well as test-retest reliability. In this subchapter the reliability of the inter rater is calculated using Inter rater Reliability showing the comparison between the variations attributed to the attributes as measured by the overall measurement variation. The level of agreement (reliability) among the four assessors can be explained by calculating the reliability coefficient of the Assessor using the Intraclass Correlation Coefficient coefficient. The result of the calculation (using the banten program SPSS version 16).

Sixth, the practicality of a test is an indicator of the quality of a measuring tool is good or not. The result of analysis practicality of performance appraisal instrument based on data obtained by giving questionnaire questionnaire to four assessors (teachers) who tested the use of instruments. Empirically, the seven assessors are asked to fill out the questionnaire with five types of questions in the form of rubrics with a range of values 1 to 5, which relates to the practicality of use. The data of the respondents' answers are then analyzed statistically by using the Score T formula. Based on the analysis of data can be interpreted that through standardized test score (standard) with T score on the practicality of the use of performance appraisal instrument on learning mathematics seen in table 9.

| Person   | X  | x̄  | (x-  x̄)^2 | Z=(x-  x̄)/SD | Skor T | Kriteria |
|----------|----|-----|------------|----------------|--------|----------|
| Rater I  | 49 | -1,33 | 1,77       | -1,15          | 49     | Practical |
| Rater II | 51 | 0,67  | 0,45       | 0,58           | 54     | Practical |
| Rater III| 51 | 0,67  | 0,45       | 0,58           | 54     | Practical |
| Jumlah   | 151|      |            |                | 157    | Practical |
| Mean     | 50,33 | 2,6 |            | 52,33          |        | Practical |
| SD       | 1,15  |    |            |                |        |          |

Seen from Table 9 in general teachers assess the Performance instruments have good subjectivity, systematics, construction, language and practicality. This is reflected from the T score of each rater I of 49, rater II obtained a T score of 54, and from rater III obtained a T score of 54. Thus according to the criteria of practicality of Glicman instruments Performance instruments can be said to be generally considered practical by the teacher (rater) in assessing the quality level of practicality of performance instrument of elementary school students on learning mathematics.

4. Conclusions
Based on the results of research development that has been done, can be drawn conclusions as follows: (1) assessment of performance that has been there is the elementary school teachers in the city of Cirebon still use the write test description and duty sheet; (2) Validation of assessment instruments is done through expert test and empirical validity. The result of the assessment obtained from the validation of the expert states that the assessment of this performance is appropriate to be used as a form of assessment.

The authentic assessment instrument of performance on the developed Mathematics learning has fulfilled the requirements of validity, reliability and practicality, as an evaluation tool that can be used further by Mathematics teachers in Cirebon elementary school.

References
[1] Cohen A 1994 Classr. Restruct.: Cond. Product. Small Group. 64
[2] Ce Cohen BP, and Cohen EG 1991 From group work among children to R & D teams: Interdependence, interaction, and productivity. In EJ Lawler, B. Mark-ovsky, C. Ridgeway, & HA Walker (Eds.), Progress in group process (Greenwich, CN: JAI)

[3] Cobb P, Yackel E and Wood T 1992 Study Educ. Math. 23 99

[4] Chun M 2010 Change 42 22

[5] Moss P, Girard B J and Haniford L C 2006 Res. Rev. Educ. 30 109.

[6] Newmann F M, Lopez U S and Bryk U S A 1998 Quality of intellectual work at Chicago school: A basic report (Chicago: Chicago School Research Consortium)

[7] Rochex J 2006 Res. Rev. Educ. 30 163

[8] Wiggins G and McTighe J 2005 Understand by design (2nd ed.) (Alexandria, VA: Association for the Supervision and Development of the ASCD Curriculum)

[9] Shepard L A, Flexer R J, Hiebert E H, Marion S F, Mayfield V and Weston T J 1996 Meas. Educ.: Probl. Pract. 15 7

[10] Borg and G 2007 Educational Research, an introduction (New York & London Longman Inc. From NetLibrary)

[11] Sugiyono 2010 Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif, dan R&D (Bandung: Alfabela)

[12] Guilford J P 1956 Fundamental Statistic in Psychology and Education 3rd Ed (New York: McGraw-Hill Book Company, Inc)

[13] Harold A and Linstone M T 1975 Delphi Method: Technique and Application (Massa: Adison-Wesley)