Effectiveness of Flash-Based Summative Assessment Using the Anchor Items on the Learning of Science

Yetti Supriyati
y_supriyati@yahoo.com

Abstract

The main purpose of this research was to study the effectiveness of summative assessment for science using a parallel test that has the anchor items and the test was made using flash animation. The test items that have been developed in this research were two sets of science lesson tests for Junior High School, and each set of test are using anchor items proportion 20% (≠ 8 items), and has been analyzed for their reliability, item difficulty, and discrimination. This research can be classified into a research and development. This research was to develop two point of problems which is based on characteristics of the science-based option animated flash, 40 point in each problem in parallel and equivalent situation. The items has been reviewed based on the concept, expert, and empiric validations. Research indicates that the two parallel tests which are based on flash animation and have made anchor items are effective in the materials attractiveness, theological problem clarity, clarity of choice for use in summative tests on teaching science to junior high school students.

Keywords: assessment, anchor items, flash animation.

1. Introduction

Learning of science, especially physics, generally is a difficult subject to be learned by Indonesian student. The concept that had been taught by teacher is mostly based on basic fundamental of theory and isn’t based on process learning by observation and measurement of natural phenomenon. Whereas based on student competency, student is expected to extend critical observation, creative and logic thinking, and consistent to their daily base life phenomena. Teacher itself is required to make several methods of learning in order to measure their student competency by process and product evaluation. However the deficiency of valuation instrument is one of the big problems where aren’t any valid and reliable instruments that can be used to evaluate learning process of student. Therefore this study is trying to support the teacher on their science process evaluation using flash animation technology and to challenge the student to be logic and critical on doing observation. The instruments are made with anchor items that one group of student will do the different problems with others groups.

The purpose of this research is to define the effectiveness of test modeling which used by flash animation and have anchor items. The development of assessment instrument refers to the accession of basic competence for Junior High School student. This instrument forms with two tests set where between sets have 20 % of anchor items. The content validity and empiric validity of assessment instrument had been done as instrument standardization.

1 State University of Jakarta, Indonesia
Table 1: Elements of Instrument

<table>
<thead>
<tr>
<th>Total</th>
<th>Items</th>
<th>S đủ Indicator</th>
<th>Indicator</th>
<th>Component</th>
<th>Competency</th>
<th>Standard</th>
</tr>
</thead>
</table>

1. Determination of the type of instrument
2. Development of assessment instrument. The steps should be done based on the development of instrument theory as follows:
   a. Identification of the type of instrument
   b. Conceptualization (perspective of instrument)
   c. Development of instrument (perspective of instrument)
   d. Development of instrument (perspective of instrument)
   e. Development of instrument (perspective of instrument)
   f. The study consists of several stages: 1.

2. Methods and Sampling Methods
to notice the suitability between theory and practical observation, and then performed the analysis of the validity and reliability.

B. Population and Sampling Technique

Population of this research is the two groups of students. Coming from the school with different category, that is, "the eminent school" and the non-eminent school, both either private school and also public school. The first set of test (X) is passed to the students coming from the Eminent School (public school and private school); and the second set of test (Y) passed to the student coming from the non-eminent school (public school and private school).

Numbers of test items from the test X and the test Y have the same number, either for the anchor items and also the non-anchor items. The sample technique used is proportional random sampling. In this study, the respondents are Junior High school student (2010/2011). Figure 1 is presented to illustrate two couple of test using anchor items with different proportion. The test X consisting of the test items (X+Z) is parallel with the test consisting of the test items (Y+Z), where X and Y are the test with a non-anchor items and Z is the test with an anchor items.

![Anchor items diagram]

**Fig. 1. the couple of test having anchor items**

To calculate the score equating is done by using the mathematics equation which express the relationship between the score obtained from one set of test and the score obtained from the other one. This relation is a linear relationship, so that there is a coefficient of the score equating (a). Mathematics formula came into use to calculate the score equating in this research is: \( Y = a (X - c) + d \)

<table>
<thead>
<tr>
<th>Description</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = score result of conversion</td>
<td>X = score being converted</td>
</tr>
<tr>
<td>A = coefficient of equating score</td>
<td>C = constant</td>
</tr>
<tr>
<td>D = constant</td>
<td></td>
</tr>
</tbody>
</table>

C. Technique to Analyze the Set of Test Items

Analyzing the set of test items consist of analyzing the reliability of anchor proportion, analyzing the difficulty level of the test using anchor items, analyzing the discrimination power of the test anchor items. Calculating the difficulty level of each the test items starts up analyzing the difficulty level of the test items
C. Research Design

The study design must be calculated after determining the coefficient of correlation to obtain the most suitable coefficient of correlation. After calculating the coefficient of correlation of the various proportions to be obtained, the following formula can be used:

\[
\rho(\rho - 2) = \frac{\text{SS}_a - \text{SS}_b}{\text{SS}_a + \text{SS}_b}
\]

where \( \rho \) is the correlation coefficient of the test scores, \( \text{SS}_a \) is the sum of squares for the first set of test scores, and \( \text{SS}_b \) is the sum of squares for the second set of test scores.

Score by means of the mathematical formula as follows:

Analyzing the variance of each test score is involved with calculating the correlation of each test. The formula for calculating the determination power of each test score is expressed as:

\[
\text{R}^2 = 1 - \frac{\text{SS}_e}{\text{SS}_t}
\]

where \( \text{R}^2 \) is the coefficient of determination, \( \text{SS}_e \) is the sum of squares for the error, and \( \text{SS}_t \) is the total sum of squares.
3. Results

The results of the validation assessment point with flash animation by expert instructional media showed the overall average aspects of the display (graphical) assessment point with flash animation physics of trial data table of media experts demonstrated the value of 3.2 on the good rating. Based on the Likert scale, the scores interpretation for overall judgments for each point through flash animation was in the range 61-80% or 2.81 to 3.40, which means good. These results indicate that assessment with flash animation on science learning in terms of graphics, useable, legibility and efficiency is good.

The Results of reliability test with the formula Kuder-Richardson 20 (KR-20) showed classification as follows:

- 0.91 to 1.00 = very high
- 0.70 to 0.90 = high
- 0.41 to 0.70 = quite
- 0.21 to 0.40 = low
- <0.20 = very low

From the calculation, there are 8 points about the fall, so the reliability coefficient obtained instrument (72 questions), device test X coefficient reliability of 0.89, and device reliability coefficient of 0.88 Y. Instrument reliability is included in the high category, so that the instrument can be used as a measuring tool.

To calculate the distress index instruments using the formula.

Classification index difficulty:

- 0.00 to 0.30 = difficult
- 0.31 to 0.70 = moderate
- 0.71 to 0.90 = difficult

From the calculation, there are 10 questions fall into the category of hard (values between 0.1 to 0.3), 13 questions in the category are (values between 0.45 to 0.7), and 17 questions included in the easy category (values between 0.75 - 1.00).

Distinguishing Power Index Index distinguishing features can determined using the following formula (JW Popham, 1981: 294) with classification as follow:

- 0.00 - 0.20 = poor
- 0.21 - 0.40 = satisfactory
- 0.41 - 0.70 = good
- 0.71 - 1.00 = excellent

From the calculation above, the total of 11 questions included in classification enough. A total of 20 included in the classification good, and 9 questions included in the excellent classification.
4. Conclusion

Based on the result of study on developing assessment instrument on Sains, it can be concluded that there are two valid and reliable instruments. The result of theoretic test by the experts showed there are 72 multiple choice questions that valid. Based on empiric validation, there were 8 points that rejected, therefore obtained 2 set of instruments with each instruments 40 pints with 8 points anchor items in reliability on 0. 88 and 0.89. Based on the result of questionnaire from students, the instruments were interested and they feel fun in finished the test. Moreover, this study produced two set of assessment instruments that have anchor items for each items, which has coefficient equitely stable.

Based on the explanation above, the conclusion is the assessment instruments is valid and reliable, can be used for sains subjects in secondary school, the two sets of assessment instruments can be used to two groups that are equal. Moreover, it can improve students interesting in doing the assessment.

Reference