Validity of Test Instruments

Sudaryono¹, Untung Rahardja², Qurotul Aini³, Yuliana Isma Graha⁴ and Ninda Lutfiani⁵

¹University of Banten Jaya, Information System, Serang, Banten 42117, Indonesia
²STMIK Raharja, Information System, Tangerang, Banten 15117, Indonesia
³STMIK Raharja, Information System, Tangerang, Banten 15117, Indonesia
⁴STMIK Raharja, Information System, Tangerang, Banten 15117, Indonesia
⁵STMIK Raharja, Information System, Tangerang, Banten 15117, Indonesia

email: dryono10@gmail.com; untung@raharja.info; aini@raharja.info; iz@raharja.info; ninda@raharja.info

Abstract. The two basic principles of the problem in assessment are determining whether a test has measured what is being measured and whether a test has been appropriately used to make a decision about the test taker. It is possible that the test developers argue that math tests for example can predict a person's ability in physics. Validity or validity comes from the word validity, which means the extent of accuracy and accuracy of a measuring instrument in performing its measuring function. In other words, validity is a concept related to the extent to which the test has measured what should be measured. In this study discusses the meaning of validity, types of validity, testing the validity of tests, testing the validity of tests rationally, construct validity, content validity, testing the validity of tests empirically, forecast validity, comparative validity, question validity, and validity testing calculations. With the method of quantitative data analysis, the results of the research from the distribution of questionnaires to 20 (twenty) respondents, 10 (ten) of them were tested for validity and it was found that 7 (seven) items were valid and 3 (three) invalid.

1. Introduction

Validity or validity comes from the word validity which means the extent of the determination and accuracy of a measuring instrument in performing its measuring function. The validity of a test is always divided into two types, namely logical validity and empirical validity [1]. Logical validity is the same as qualitative analysis of a problem, which is to determine whether or not a function is based on predetermined criteria, which in this case is the criteria for material, construction, and language [2]. A test or measurement instrument is said to have high validity if the tool carries out its measuring function [3], or provides a measurement result that is in accordance with the purpose of doing the measurement [4]. This means that the measurement results from these measurements are quantities that accurately reflect the actual facts or circumstances of what is measured [5]. Thus, a valid test for a specific purpose is a test that is able to measure what is being measured [6]. A test that is valid for a specific purpose or a return of a particular decision, may be invalid for the purpose or other decision making [7]. So the validity of a test must always be associated with certain goals or decisions [8]. Entrance tests for example must always be related to how far the entrance test
can reflect the learning achievements of prospective new students after studying later. Test validity needs to be determined to determine the quality of the test in relation to measuring what should be measured [9]. The concept of instrument validity or test can be divided into three types, namely: (a) content validity; (b) construct validity; and (c) empirical validity or criteria validity.

Content validity shows the extent to which questions, tasks or items in a test or instrument are able to represent the overall and proportionate overall behavior of a sample that is the learning objective that will be measured [10]. This means that the test reflects the overall content or material being tested or that should be controlled proportionally. To find out whether the test is valid or not, it must be done through a review of the test grid to ensure that the test questions represent or reflect the entire content or material that should be controlled proportionally [11]. Therefore, the validity of the contents of a test does not have a certain quantity that is calculated statistically, but it is understood that the test is valid based on the review of the test grid.

Therefore, Wiersma and Jurs (1990) state that content validity is actually based on logic analysis, so it is not a validity coefficient that is calculated statistically [12]. Construct validity means that a measuring instrument is said to be valid if it has matched the theoretical construction where the test was made. A test has construction validity if the problems measure aspects of thinking [13]. Construct validity is commonly used for instruments that are intended to measure concept variables, both typical performance characteristics such as instruments to measure attitudes, interests, self-concept, locus of control, leadership style, achievement motivation, etc., as well as performance characteristics. maximum such as instrument to measure talent (aptitude tests), intelligence (intellectual intelligence), emotional intelligence and others [14].

Empirical validity or the validity of the criteria of an instrument or test is determined based on the measurement results of the instrument in question, either through trials or through actual tests or measurements [15]. Empirical validity or criteria validity is defined as validity that is determined based on criteria, both internal criteria and external criteria. Internal criteria are tests or instruments themselves which become criteria, while external criteria are the results of measuring instruments or other tests outside the instrument which are the criteria [16]. Other measures that are considered standard or can be trusted can also be used as external criteria [17]. In the 1940s and 1950 education measurement experts have carried out various studies on how to determine and assess validity. In 1954 The American Psychological Association (APA) through Technical Recommendation for Psychological Test and Diagnostic Techniques proposed four approaches that are often called four-face validity used to determine validity: (a) content validity; (b) construct validity; (c) concurrent validity; and (d) predictive validity [24].

2. Research Methods

2.1. Rational Test Validity Testing

Learning outcomes that have been analyzed rationally have the power to measure accuracy, called learning outcomes tests that have logical validity. Other terms for logical validity are: rational validity, ideal validity, or das sollen validity [19]. Rational validity is the validity obtained on the basis of the results of thought, the validity obtained by thinking logically [20]. Thus, a test of learning outcomes can be said to have rational validity, if after analyzing rationally it turns out that the test of learning outcomes is indeed (rationally) correctly able to measure what should be measured [21]. To be able to determine whether the test of learning outcomes has rational or not validity, it can be traced from two aspects, namely in terms of its contents = content and in terms of arrangement or construction (construct).

2.2. Content Validity

The validity of the content of a test of learning outcomes is the validity obtained after analyzing, tracing or testing the contents contained in the test of learning outcomes. Content validity is the
validity seen in terms of the content of the test itself as a measuring tool for learning outcomes, namely: the extent to which the test of learning outcomes as a tool to measure student learning outcomes, its contents can represent representative of the entire material or learning material that should be tested. Content validity ensures that measurements include a sufficient set of points that represent the concept [22].

Thus according to Guion (1988), content validity is very dependent on two things, the test itself and the process that influences the response to the test. For example a written test prepared for work might not present a valid measurement for the employee's ability to do work, even though it might be a valid tool to measure knowledge about what to do. One way to obtain content validity is to look at the questions that make up the test. If the whole question appears to measure what the test should be used for, there is no doubt that the content validity has been fulfilled.

So, the talk about content validity is actually identical to the talk about population and sample. If only the entire subject matter that has been given to students or has been instructed to be learned by students is considered as a population, and the content of the learning outcomes test in the same subject we consider as the sample, then the test of learning outcomes in these subjects can be said to have content validity, if the contents of the test (as a sample), can be a representative representative (feasible = adequate) for all subject matter that has been taught or has been ordered to be studied (as a population).

2.3. Construct Validity

Construction is something that is related to abstract phenomena and objects, but symptoms can be observed and measured. Gravity for example can be used as an example of how to understand the construct. When apples fall to the ground, a construct about gravity can be used to explain and estimate behavior (fall of apples) observed [24]. However, we cannot see what is meant by the gravity construct itself. The only thing we can see is the apple falling. We can measure gravity and develop theories about gravity.

Construct validity refers to how far a test measures the nature or structure of a particular construct and this validity is important for tests used to assess one's abilities and psychological characteristics [18]. The building approach to understanding the validity aims to establish the building of psychological understanding what is measured by a test and how far the building of understanding can be measured. There are two kinds of validity according to the method of testing, namely internal validity and external validity. Testing an instrument in internal validity is done in two ways, namely: 1) conducting a factor analysis. Factor analysis is carried out if there are similarities, continuities, or overlaps between the factors with one another. Factor analysis is done by correlating factor scores with total scores; and 2) conducting item analysis. To test the validity of each item, the scores on the item in question are correlated with the total score.

Validity is assessed through convergent validity and discriminant validity. Convergent validity is met if the score obtained with two different instruments that measure the same concept shows a high correlation [22]. Discriminant validity is fulfilled if, based on theory, two predicted variables are uncorrelated, and the scores obtained by measuring them are truly empirically proven. Thus, validity can be produced in various ways. Published sizes for various concepts usually report the type of validity that has been generated for the instrument, so that the user or reader can assess the accuracy of the measurement.

2.4. Testing the Validity of the Empirical Test

The empirical validity or validity of the criteria of a test or instrument is determined based on the measurement results of the instrument concerned, either through trial or through actual tests or measurements [8]. Empirical validity is defined as validity that is determined based on criteria, both internal criteria and external criteria. The internal criterion is the test itself which is the criterion, while the external criterion is the result of measuring other tests outside the test which is the criterion. In
other words, empirical validity is validity that originates from or is obtained on the basis of observations in the field.

Starting from that, the test of learning outcomes can be said to have empirical validity if based on the results of the analysis carried out on the observational data in the field, it is evident that the test of learning outcomes has correctly been able to measure learning outcomes that should be revealed or measured through test results [23]. To be able to determine whether the test of learning outcomes already has empirical validity or not, can be traced from two aspects, namely in terms of predictive validity and concurrent validity.

2.5. Validity Testing Techniques

1. Test Validity

If we want to pay close attention, then the test results of learning that are made or compiled by teachers, both teachers, other teaching staff lecturers, are actually a collection of many test items, with items where the test authors want to measure or express learning outcomes that have been achieved by each individual student, after they have followed the learning process within a certain period. The statement contains meaning, that in fact every item in the test of learning outcomes is an integral part of the test of learning outcomes as a totality [24].

The tight relationship between item items and test results as a totality can be understood from the fact that the more item items that can be answered correctly by testee, the higher the total score of the test results. Conversely, the fewer item items that can be answered correctly by testee, the total score of the test results will be lower or decrease.

The statement is an indication that the greater the "support" given by item items (as an inseparable part of the test), to the test of learning outcomes (as a totality), then the test will be able to show more "stability". Conversely, the smaller the "support" that each item gives to the test as a totality, the makates become increasingly "less stable". If the statement we associate with the validity of the item we are talking about, then it can be understood that the validity of the test will be greatly influenced by, or very much depends on the validity of each item that builds the test. The meaning contained in the statement further is that the validity of each item that builds the test will be known by looking at the size of the support given by each item related to the test as a whole [25].

The issue of the validity of the item would not be too urgent to handle, if only based on the validity test it turns out that the test of learning outcomes made by the teacher, lecturer or other teaching staff turned out to have high test validity, so the test of learning outcomes as a totality can already be said reliable and need not doubt the accuracy of measuring it. But the problem will soon emerge, if after testing the validity of the test on the test used as a measuring tool it turns out that the conclusion is that the learning outcome test is very low validity, so that it can be included in the invalid learning outcomes test category.

The purpose of the validity of test items is to determine whether or not a problem can distinguish groups in the measured aspects according to the differences in the group. The validity of the question is the discrimination index in distinguishing between highly capable test participants and low-ability test participants. The validity of the question is the discrimination index of the questions determined from the difference in the proportion of answers from each group [24]. This index shows the compatibility between the problem function and the overall test function. Thus the validity of this question is the same as the distinguishing power of the problem, namely the power to distinguish between highly capable test participants and low-ability test participants.

2. Test Item Validity Test Technique

From the description that has been stated above, it would be quite clear that an item can be said to have high validity or can be declared valid, if the scores on the item in question have suitability or alignment with the total score; or with statistical language: There is a significant positive correlation
between item scores and total scores. The total score here is located as a related variable (dependent variable), while the item score is located as an independent variable (independent variable [26]). If so, then to come to the conclusion that the items that want to know the validity, that is valid or not, we can use correlation techniques as an analytical technique. An item can be declared valid if the item score is proven to have a significant positive correlation with the total score.

Numbers that indicate the magnitude of the validity of the question are called the index of validity of the question whose magnitude ranges from -1 to +1. Negative signs indicate that test participants with low ability can answer correctly while test participants who have high abilities answer incorrectly. Thus the problem of negative validity indicates the reversal of the quality of the test participants. Each question can be seen as a separate part of a test. A question might be able to distinguish the group of test participants well. A question may also not be able to distinguish the group of test participants (eg questions with p = 0 or p = 1). A question also distinguishes the group in reverse, ie the test participants who are unable to answer the questions correctly while the test participants are able to answer incorrectly. One of the objectives of the problem analysis is to look for questions that can measure abilities appropriately.

2.6. Quantitative Data Analysis Method

Data analysis method needed in conducting this research is quantitative data analysis method where data in the form of numbers and by using statistical calculations to analyze a hypothesis in research. Data analysis is the process of systematically finding and compiling data obtained from interviews, field notes, and other materials, so that they can be easily understood, and their findings can be informed to others.

The problem is in choosing and determining the type of correlation technique that is considered appropriate to be used in order to test the validity of the item. As is known, in objective tests there are only two possible answers, namely right and wrong [27]. Each item answered correctly is generally given a score of 1 (one), while for each wrong answer is given a score of 0 (zero). This type of data, namely: right - wrong, yes - no or similar, in the world of statistics known as pure discrete data or dichotomous data [28]. Whereas the total score owned by each individual testee is the sum of each score possessed by each item (for example: 0 + 1 + 1 + 0 + 1 + 0 + 1 + 0 + 0 + 1 = 6) it is continuous data. According to the theory, if variable I is pure discrete data or dichotomous data, while variable II is continuous data, then the correct correlation technique to be used in finding the correlation between variable I and variable II is biserial point correlation technique, where index number Correlations that are given the RSPI symbol can be obtained using the formula:

\[
r_{pbi} = \frac{M_p - M}{SD_t} \sqrt{\frac{p}{q}}
\]

- \( r_{pbi} \) = The biserial point correlation coefficient represents the strength of the correlation between variable I and variable II, which in this case is considered as the Item Validity Coefficient.
- \( M_p \) = The calculated average score is owned by testee, which for the item in question has been answered correctly.
- \( M_t \) = Average score of total score
- \( SD_t \) = The standard deviation of the total score.
- \( p \) = The proportion of testee who answered correctly to the item being tested for the validity of the item.
- \( q \) = The proportion of the testee who answered incorrectly to the item being tested for the validity of the item.
3. Results and Discussion

20 testees were presented with an objective Multiple Choice Item test which presented 10 items, where for each item that was answered correctly it was given a score of 1, while for each item that was answered incorrectly it was given a score of 0. After the test ended, corrected and calculated scores, obtained the test results data as shown in table 1.

Table 1. Dissemination of Test Result Scores Followed by 20 Testee People, by Presenting 10 Item Items in the Form of Multiple Choice Items.

| Testee | Score for item number items: | Total Score |
|--------|-------------------------------|-------------|
|        | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | (X.) |
| A      | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1  | 3   |
| B      | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1  | 7   |
| C      | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1  | 6   |
| D      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 10  |
| E      | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1  | 7   |
| F      | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0  | 3   |
| G      | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 8   |
| H      | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 9   |
| I      | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1  | 5   |
| J      | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 10  |
| K      | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0  | 6   |
| L      | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0  | 5   |
| M      | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0  | 4   |
| N      | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1  | 7   |
| O      | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 8   |
| P      | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1  | 5   |
| Q      | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 9   |
| R      | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1  | 6   |
| S      | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 8   |
| T      | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1  | 4   |

20 | 10 | 12 | 10 | 14 | 13 | 15 | 12 | 16 | 12 | 16 | 130 |

Step 1: Prepare a calculation table in order to analyze the validity of items number 1 through number 10. (See Table 2).
Table 2. Calculation Table in the Framework of Item Validity Analysis

| Testee | Score for item number items: | \( x_i \) | \( X_i^2 \) |
|--------|-------------------------------|----------|----------|
| A      | 0 0 0 1 1 0 0 1 0 1 0 1 0 0 1 | 3 4      | 9 49     |
| B      | (1) 0 1 0 1 0 1 1 1 1 1 1 1 1 1 | (7) 49   | 100 100  |
| C      | 0 0 1 1 1 0 0 1 1 1 1 1 1 1 1 | 6 36     | 3 9      |
| D      | (1) 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 | (8) 64   |          |
| E      | (1) 0 1 1 0 1 0 0 1 1 1 1 1 1 1 1 | (9) 81   |          |
| F      | 0 1 0 0 0 1 0 1 0 0 0 1 0 0 0 | 5 25     |          |
| G      | (1) 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 | (8) 64   |          |
| H      | (1) 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | (9) 81   |          |
| I      | 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 | 6 36     |          |
| J      | (1) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | (10) 100 |          |
| K      | 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 | 5 25     |          |
| L      | 0 1 1 0 0 1 0 1 1 1 1 1 1 1 1 | 4 16     |          |
| M      | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | (7) 49   |          |
| N      | (1) 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 | (8) 64   |          |
| O      | (1) 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 | (9) 81   |          |
| P      | 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 | 6 36     |          |
| Q      | (1) 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | (8) 64   |          |
| R      | 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 | 4 16     |          |
| S      | (1) 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 | (10) 100 |          |
| T      | 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 | 20 = N 10 = \( x \) 12 = \( x \) 10 = \( x \) 14 = \( x \) 15 = \( x \) 12 = \( x \) 16 = \( x \) 16 = \( x \) 130 = \( x \) 934 = \( x \) | 0.5 0.6 0.5 0.7 0.65 0.75 0.6 0.8 0.6 0.8 |

Step 2 : Look for the mean of the total score, that is \( M_i \), by using the formula:

\[
M_i = \frac{\sum x_i}{N}
\]

It is known: \( \sum X_i = 130 \) dann \( N = 20 \). So:

\[
M_i = \frac{130}{20} = 6.5
\]

Step 3 : Look for the total standard deviation, namely SDt, using the formula:

Step 4 : Search (count) \( M_o \) for items number 1 through number 10, which to summarize the discussion, are set out in Table 15.3.
### Table 3. Calculations for Obtaining Mp from Item Item Number 1 Up to Number 10.

| Item Number | Tester whose answer is correct: | The mean of the total score correctly answered ($M_p$) |
|-------------|-------------------------------|----------------------------------|
| 1           | B-D-E-G-H-J-N-Q dan S ($N_p = 10$) | $7\times10+7\times4+9+10+7+8+9+8$ |
|             |                               | = 8,300                           |
| 2           | A-C-D-F-I-J-K-L-M-P-R-T       | $2\times8+1+4+6\times5\times1+5+6$ |
|             |                               | = 5,583                           |
| 3           | B-D-E-H-J-K-L-N-Q dan T ($N_p = 10$) | $5\times10+7\times7+9+10+6+5\times7+9+8$ |
|             |                               | = 7,300                           |
| 4           | C-D-E-G-H-I-J-K-N-O-P-Q-R-S   | $6\times10+7\times5+9+10+6+5\times8+9+8+6$ |
|             |                               | = 7,429                           |
| 5           | B-C-D-G-H-J-K-L-M-O-Q-R-S     | $7\times10+8+9+10+6+5+4+8+9+6$ |
|             |                               | = 7,385                           |
| 6           | A-D-E-F-G-H-I-J-K-N-O-P-Q-R-S | $7\times10+8+9+10+6+5+4+8+9+8$ |
|             |                               | = 6,933                           |
| 7           | B-D-G-H-I-J-K-L-M-O-Q-S-T     | $7\times10+8+9+10+6+5+4+8+9+8$ |
|             |                               | = 7,333                           |
| 8           | B-C-D-E-F-G-H-I-J-M-N-O-P-Q-R-S | $7\times10+7\times3+4+5+10+4+7+8+5+9+6$ |
|             |                               | = 7,000                           |
| 9           | B-C-D-E-G-H-J-L-N-O-Q-S       | $7\times6+8+1+7+8+9+4+3+4+7+8+3+4$ |
|             |                               | = 7,833                           |
| 10          | A-B-C-D-E-G-H-I-J-N-O-P-Q-R-S-T | $5\times7+6+1+7+8+9+5+10+7+8+5+9+6+3+6$ |
|             |                               | = 7,333                           |

To summarize the discussion, the results of the calculation are presented in table 4.
Table 4. Calculations to Know RPPI Correlation Coefficient in the Framework of Validity Test Items Number 1 to Number 10.

| Item Number | $M_p$ | $M_i$ | $SD_p$ | $SD_i$ | $r_{p,i}$ | $r_{i}$ | $r_{p,i} > r_{i}$ | Interpretasi |
|-------------|-------|-------|--------|--------|-----------|--------|-----------------|--------------|
| 1           | 8,300 | 6,5   | 2,11   | 0,50   | 0,50      | 0,853  |                  | Valid        |
| 2           | 5,583 | 6,5   | 2,11   | 0,60   | 0,40      | -0,532 |                  | Invalid      |
| 3           | 7,300 | 6,5   | 2,11   | 0,50   | 0,50      | 0,379  |                  | Invalid      |
| 4           | 7,429 | 6,5   | 2,11   | 0,70   | 0,30      | 0,673  |                  | Valid        |
| 5           | 7,385 | 6,5   | 2,11   | 0,65   | 0,35      | 0,572  |                  | Valid        |
| 6           | 6,933 | 6,5   | 2,11   | 0,75   | 0,25      | 0,355  |                  | Invalid      |
| 7           | 7,333 | 6,5   | 2,11   | 0,60   | 0,40      | 0,684  |                  | Valid        |
| 8           | 7,000 | 6,5   | 2,11   | 0,80   | 0,20      | 0,474  |                  | Valid        |
| 9           | 7,833 | 6,5   | 2,11   | 0,60   | 0,40      | 0,774  |                  | Valid        |
| 10          | 7,000 | 6,5   | 2,11   | 0,80   | 0,20      | 0,474  |                  | Valid        |

Starting from the results of the analysis above, it turns out that out of 10 items tested for validity, 7 items of which have been declared as valid items, namely items number 1, 4, 5, 7, 8, 9 and 10. Whereas 3 other items, items item 2, 3 and 6 are invalid items.

4. Conclusions

1. The two basic principles of the problem in assessment are determining whether a test has measured what is being measured and whether a test has been appropriately used to make a decision about the test taker. Validity or validity comes from the word validity which means the extent of the determination and accuracy of a measuring instrument in performing its measuring function. Rational validity testing can be divided into two types, namely: (a) content validity; and (b) construct validity; and (c) empirical validity or criteria validity. Testing the validity of the test empirically is divided into two types, namely: (a) the validity of the forecast; and (b) comparative validity.

2. The validity of the content of a test of learning outcomes is the validity obtained after analyzing, tracing or testing the contents contained in the test of learning outcomes. Construct validity refers to how far a test measures the nature or structure of a particular construct and this validity is important for tests used to assess a person's abilities and psychological characteristics.

References

[1] Ahiri, Jafar. 2002. “Analisis Faktor Sebagai Alat Evaluasi Validitas Konstruk”, *Matahari: Jurnal Pendidikan dan Manajemen PPs UHAMKA*. Vol.3 No. 1.

[2] Aini, Q., Rahardja, U., Moeins, A., & Apriani, D. M. (2018). Penerapan Gamifikasi pada Sistem Informasi Penilaian Ujian Mahasiswa Untuk Meningkatkan Kinerja Dosen. *Jurnal Informatika Upgris*, 4(1).

[3] Rahardja, U., Lutfiani, N., & Rahmawati, R. (2018). Persepsi Mahasiswa Terhadap Berita Pada Website APTISI. *SISFOTENIKA*, 8(2), 117-127.
[4] Aiken, Lewis R. 1994. *Psychological Testing and Assessment*. Boston: Allyn and Bacon.

[5] Rahardja, U., Moein, A., & Lutfiani, N. Leadership, Competency, Working Motivation and Performance of High Private Education Lecturer with Institution Accreditation B: Area Kopertis IV Banten Province. *Man India*, 97(24), 179-192.

[6] Anastasi, Anne and Susan Urbina. 1997. *Psychological Testing*. New Jersey: Prentice-Hall, Inc Company.

[7] Aini, Q., Rahardja, U., Madiistriyatno, H., & Setiaji, Y. D. M. (2018). Pengamanan Pengelolaan Hak Akses Web Berbasis Yii Framework. *Syntaks Jurnal Informatika*, 7(1), 52-63.

[8] Djaali, Pudji Mulyono, 2004. *Pengukuran dalam Bidang Pendidikan*, Jakarta: Program Pascasarjana Universitas Negeri Jakarta.

[9] Rahardja, U., Lutfiani, N., & Alpansuri, M. S. (2018). Pemanfaatan Google Formulir Sebagai Sistem Pendaftaran Anggota Pada Website Aplit. *or. id SISFOTENIKA*, 8(2), 128-139.

[10] Aini, Q., Rahardja, U., & Naufal, R. S. (2018). Penerapan Single Sign On dengan Google pada Website berbasis YII Framework. *Sisfotenika*, 8(1), 57-68.

[11] Gregory, Robert J. 2000. *Psychological Testing: Principles and Applications*. Boston : Allyn and Bacon.

[12] Aini, Q., Rahardja, U., & Allamiah, F. (2018). Perancangan Sistem Absensi Kinerja Pengabdian Tri Dharma secara Online pada Website Berbasis YII Framework. *Sisfotenika*, 8(1), 13-22.

[13] Gronlund, Norman E dan Linn, Robert L. 1985. *Measurement and Evaluation in Teaching*. New York: McMillan Publishing Company.

[14] Rahardja, U., Aini, Q., & Sari, D. M. (2018). Pemanfaatan Automated Email System (AEMS) Sebagai Media Notifikasi Penilaian Hasil Belajar. *SENSITEK*, 1(1), 49-54.

[15] Kerlinger, Fred N. 2002. *Azas-Azas Penelitian Behavioral*. Terjemahan Landung R Simatupang, Yogyakarta: Gadjah Mada University Press.

[16] Rahardja, U., Aini, Q., & Santoso, N. P. L. (2018). Pengintegrasian YII Framework Berbasis API pada Sistem Penilaian Absensi. *SISFOTENIKA*, 8(2), 140-152.

[17] Nunnaly, Jum C. 1978. *Introduction to Psychological Measurement*. New York: McGraw-Hill Book Company.

[18] Sangaji, E. Mamang, Sopiah. 2010. *Metodologi Penelitian: Pendekatan Praktis dalam Penelitian*, Yogyakarta: Penerbit Andi.

[19] Rahardja, U., Aini, Q., & Faradilla, F. (2018). Implementasi Viewboard Berbasis Interaktif Javascript Charts Pada Sistem Penilaian Perkuliahan. *Jurnal Ilmiah Teknologi Informasi Asia*, 12(2), 91-102.

[20] Wahidmurni, Alfin Mustikawan, dan Ali Ridho, 2010. *Evaluasi Pembelajaran: Kompetensi dan Praktek*. Yogyakarta: Nuha Litera.

[21] Rahardja, U., Harahap, E. P., & Pratiwi, D. I. (2018). Pemanfaatan RinfoSheet Sebagai Media Informasi Laporan Penjualan Barang pada Raharja Internet Cafe. *Jurnal Ilmiah Teknologi Informasi Asia*, 12(1), 65-74.

[22] Sekaran, Uma. 2006. *Metodologi Penelitian untuk Bisnis*, Terjemahan Kwan Men Yon, Jakarta: Salemba Empat.
[23] Spird, L.A. 1993. *Evaluating Test Validity*, Washington DC: American Educational Research Association.

[24] Surapranata, Sumarna, 2004. *Panduan Penulisan Tes Tertulis*, Bandung: Remaja Rosdakarya.

[25] Sudijono, Anas, 2003. *Pengantar Evaluasi Pendidikan*. Rajawali Press: Jakarta.

[26] Rahardja, U., Handayani, I., & Wijaya, R. (2018). Penerapan Viewboard Technomedia Journal menggunakan sistem iLearning Journal Center pada Perguruan Tinggi. Technomedia Journal, 2(2), 78-89.

[27] Sudjana, Nana, 2001. *Penilaian Hasil Proses Belajar Mengajar*. Rosdakarya: Bandung.

[28] William Wiersma dan Stephen G. Jurs. 1990. *Educational Measurement and Testing*. Boston: Allyn and Bacon.