Development of evaluation results information system learn based on website

Y Sastrawijaya*, P W Yunanto and K Jami’s
Fakultas Teknik, Universitas Negeri Jakarta, Jakarta, Indonesia

*yuliatri_s@yahoo.com

Abstract. SIP-Eval uses the concept of a decision support system and is developed with a SDLC model by designing using UML. SIP-Eval uses three tests that is assessment testing, usability testing and user accepted testing. Testing do to determine the quality of the item’s product by SIP-Eval and to know the usefulness of the SIP-Eval and to know SIP-Eval can function properly. The test results conducted at Vocational High School 7 Jakarta and Vocational High School 26 Jakarta involving 3 Server Administration subject teachers and 27 students of the Computer Network Engineering expertise program show that SIP-Eval has a good ability in assisting users in completing the evaluation of learning outcomes process and can function very well for educators and students in the implementation evaluation of the learning outcomes in schools with 73.75% obtained in usability testing and 94.70% in UAT testing. SIP-Eval produces 40 items that are declared valid. Difficulty analysis states that 5 items are too difficult, 34 items are moderate and 1 item is too easy. Differentiating power analysis states that 2 items have a very good, 19 items have well, 7 items are good enough for their power, 3items are not good for their power and 9 items are not very good at their power. Test equipment (assessment) used in the test subjects of server administration class XI even semester has a very high reliability scale with reliability test results obtained at 0.896.

1. Introduction
Evaluation of learning outcomes or learning process is an evaluation carried out to determine the level of mastery of students towards specific objectives to be achieved and to know the level of achievement of students towards learning objectives. Measurement in evaluation requires a tool to assess student learning outcomes. According to Sudijono, tools for evaluating learning outcomes must have valid, reliable, objective and practical characteristics so that the test equipment to be used must be tested first using testing techniques for validity, reliability and item analysis [1]. Based on the research that Rokhayati had done by carrying out tests on the validity and reliability of tests made by Indonesian Language teachers in Lino District, it was found that from 220 items analysed only 82 items were valid, 138 questions were invalid [2]. Seeing these results, it can be concluded that in the preparation of the question items, many teachers did not pay attention to the rules for forming questions and curriculum. In an effort to resolve the obstacles faced by the teacher when carrying out evaluation of learning outcomes, the media or information system was chosen as a media or evaluation tool by providing grid compilation facilities, validating items, test reliability. Media or tools in the implementation of evaluation of learning outcomes will be developed in the form of a website-based learning outcome evaluation information system.
Development of learning outcomes evaluation information systems is based on educational evaluation rules in the scope of evaluation of learning outcomes with the use of summative tests as an evaluation tool. According to Sudijono, there are basic principles in the preparation of learning outcome tests [1], namely: (1) Learning outcomes tests must be able to measure clear learning outcomes (learning outcomes) that have been determined in accordance with the objectives of learning, (2) Items of learning result test questions should be a representative sample of the population of learning material that has been studied, (3) The form of questions issued in the learning outcomes test must be varied, (4) Tests of learning outcomes must be designed according to their use to obtain desired results, (5) Learning outcomes tests must have reliable reliability and (6) Tests of learning outcomes in addition to being able to be used as a measuring tool for learning success of students and must be used as a tool to find information that is useful for repairs How to learn students and how to teach teachers. The cognitive competency level uses a model developed by Bloom [3], the level of cognitive competence of students starts from the lowest to the highest, namely: (1) Knowledge, (2) Understanding, (3) Application or application, (4) Analysis, (5) Synthesis and (6) Evaluation. Compilation of Grid Tables. The first step in preparing the grid table is to determine the number of items to be tested. In this development the items are limited to 40 items. Furthermore, determining the number of balances of the items seen from the indicators to be achieved by determining the content of the material to be tested, determining the percentage of content components and determining the number of items in each subject, while the calculations performed as follows:

Determine the number of questions for each content of the material and determine the balance of the number of items seen from the level of competence that will be tested in each subject matter content with the percentage set according to Sudijono as follows [1]: Knowledge Level 35%, Level of Understanding 20%, 10% Analysis Level, 5% Synthesis Level, 5% Evaluation Level. Number of items for each level of competency from each content.

1.1. Validity testing techniques, item analysis and reliability
Tests carried out in evaluating learning outcomes consist of testing validity, item analysis and reliability. Validity Testing Validity testing is based on the rules of writing. The question review format is used as the basis for analysing item questions. The format developed by Nurgiantoro examines the items based on the rules of writing items in terms of material, constructs, language / culture and the truth of the key answers / scoring guidelines conducted by several reviewers [4]. To find out the results of testing rational validity can be known by using the Content Validity Coefficient - Aiken's V. Aiken formulating Aiken's V formula to calculate the content-validity coefficient based on the results of the expert panel as many as "n" of an item in terms of the extent to which the item represents the construct measured by Validity Testing Techniques, Item Analysis and Reliability [4]. This is in accordance with the objectives of the study, namely to provide alternative tools or media to educators in conducting the evaluation process of learning outcomes by providing facilities for designing, testing, processing, calculating and analysing validity and reliability that can be done in one system. In addition, the results of using the system can give an idea to the energy question quality educators that have been used in the exam so that educators, especially teachers, can improve the questions that will be used in the next exam.

2. Method
Tests conducted in evaluating learning outcomes consist of testing validity, item analysis and reliability. a. Validity testing is based on the rules of writing. The question review format is used as the basis for analysing item questions. The format developed by Nurgiantoro examines the items based on the rules of writing items in terms of material, constructs, language / culture and the truth of the key answers / scoring guidelines conducted by several reviewers [4]. To find out the results of testing rational validity can be known by using the Content Validity Coefficient - Aiken's V. Aiken formulating Aiken's V formula to calculate the content-validity coefficient based on the results of the expert panel as many as "n" of an item in terms of the degree to which the item represents the measured construct. empirical from the item in question [5]. Empiric data obtained from the questions that have been tested. Analyzing
the items can be assessed in terms of the degree of difficulty of the item, the item distinguishing power and the function of the distractor. Analysis of Distinguishing Power of Questions Questioning the important items as one of the bases held for compiling the items in the test of learning outcomes is the assumption that the ability between one testee and another differs, and that the items must be able to provide test results that reflect differences in abilities that exist in the testee reliability. Reliability is done to answer the question of how far measurements are made repeatedly on the same subject or group of subjects, giving results that are relatively unchanged. Calculation of reliability uses the Single Test-Single Trial approach using the Kuder-Richardson Formula, where the KR20 formula is applied in developing Information Systems Website-based Learning Outcomes (SIP-Eval) uses the SDLC (Life Cycle Development System) model.

2.1. System testing
Usability Testing Usability testing is done to know the capabilities of the software to help users complete a task. Products can be used by users based on effectively defined goals and levels of satisfaction in their use. Nielson states that usability testing is based on five components, namely learnability, efficiency, memorability, and satisfaction [5]. Usability testing instruments are created using the Linkert scale. This type of measurement scale will get answers in the form of gradations from very positive to very negative in the form of words. The use of the Linkert scale is used in this study to get overall answers to a topic, opinions or experiences from the use of products that have been developed. The usability testing instrument was adapted from the Test Script developed by Krug [7]. b. Testing of UAT (User Accepting Testing) User Acceptance Testing (UAT) is the final stage of the testing carried out to determine whether there are still defects in the developed system. The UAT process is based on mutually agreed requirement documents. UAT testing instruments are made using the Guttman scale. This type of measurement scale will get firm answers, namely "yes-no", "right-wrong", "never-never", "positive-negative" and others. According to Sugiyono the use of the Guttman scale was used in this study to get a firm answer to the problem being asked [1]. In UAT testing refers to functions. 3.4. Analysis Techniques The test results are carried out by calculating the Rating Scale questionnaire to be able to interpret or interpret each number given in the alternative answers on each instrument, Riduwan explains that the calculation can be done using the percentage formula as follows: The number of ideal scores can be analysed by calculating the average answer based on the score of each answer from the respondent [8]. The ideal number of percentage scores is interpreted by the scale of the category according to the results obtained in processing. Arikunto provides a description of the percentage through intervals made with categories that match the scores from the analysis of percentage data [9]. The following is a table of percentage intervals tailored to the assessment category based on:

Table Percentage and Category Intervals:

- 0% < P ≤ 20% very Bad
- 20% < P ≤ 40% Not Good
- 40% < P ≤ 60% Good enough
- 60% < P ≤ 80% Good
- 80% < P ≤ 100% very Good

3. Results and discussion
Development of Data Processing Information Systems Learning Outcomes Evaluation with Validation Testing Techniques and Summative Test Reliability, abbreviated as SIP-Eval, uses an engineering research methodology with a system development model for SDLC (System Development Life Cycle). In the development of SIP-Eval with the SDLC model there are five stages implemented, namely: (1) Planning by looking at potential problems that occur in the field related to the process of evaluating learning outcomes in schools, (2) Analysis to find teacher needs in the process of evaluating learning outcomes in the form of documents functional requirements with planning steps starting from preparing Server Administration syllabus for class XI Computer and Network Technology expertise fields
followed by preparation of specification tables / problem grids, testing validity tests rationally by reviewers (Server Administration subject teachers), improvements items based on the results of testing rational validity, conducting summative server administration tests online, testing the degree of difficulty analysis, distinguishing and reliability, (3) design using UML (Unified Modelling Language) in the form of use cased diagrams, activity diagrams, sequence diagrams and classes diagram, (4) Implementation or stage of making SIP-Eval (coding) with website-based, and (5) Testing or testing to see the results of SIP-Eval feasibility by looking at the usefulness and functioning of SIP-Eval using usability testing and User testing. Acceptance Testing (user acceptance test). Based on the stages that have been implemented, SIP-Eval was successfully developed into a product in the form of a website-based information system. SIP-Eval is a system created as an alternative media that helps to implement evaluation of learning outcomes at the level the use of SIP-Eval involves three users, namely administrators, Educators and Students. SIPEval can be accessed with the url address of sipeval education url address.000webhostapp.com

3.1. Analysis of question items
The review involved 3 Server Administration subject teachers conducted State Vocational High School 7 Jakarta and State Vocational High School 26 Jakarta. The study was analysed based on the suitability of the questions based on material, construction and language / culture. Analysis uses the calculation of Ratting Scale.

3.2. UAT analysis
UAT testing is done to educators and students. Testing to teaching staff involves 3 Server Administration subject teachers State Vocational High School 7 Jakarta and State Vocational High School 26 Jakarta. Testing to students involved 27 XI class participants. Based on the results of tests conducted, it is known

![Figure 1. Display of the SIP-Eval website.](image-url)
that SIP-Eval is in the Very Good category with \( P = 98.24\% \) and \( P = 94.70\% \) so that SIP-Eval can function very well for educators and students.

3.3. **Analysis of degree of difficulty**
Degree of difficulty analysis is done to find out the quality of the item from the degree of difficulty where good items are not too difficult or too easy. The index number of difficulty items is obtained by using the porposi formula

3.4. **Distinguishing power analysis**
Distinguishing power analysis is done to find out whether an item can distinguish between high-ability students and low-ability students. The differentiating power can be known by using the discriminatory power formula

3.5. **Reliability test results**
Reliability testing is done to answer the question of how much measurement is done repeatedly on the same subject or group of subjects, giving results that are relatively unchanged. Reliability testing using a single test single trial approach using the kuderrichardson formula (KR20) obtained results of 0.896

4. **Conclusions**
Design and development of Learning Outcome Evaluation Data Processing Information System known as SIP-Eval successfully developed into a product in the form of a website-based information system.

The results of the tests conducted at State Vocational High School 7 Jakarta and State Vocational High School 26 Jakarta involving 3 Server Administration subject teachers and 27 students of class XI Computer Network Engineering expertise program showed that SIP-Eval has good ability in helping users in completing the evaluation process results learning and can function very well for educators and students in the implementation of evaluation of learning outcomes in schools with 73.75\% obtained at usability testing and 94.70\% on UAT testing. So that SIP-Eval greatly facilitates teachers in the implementation of the evaluation process of learning outcomes. SIP-Eval produced 40 items that were declared valid with the highest validity value 0.91 and the lowest validity value was 0.86. Degree of difficulty analysis states that 5 items are too difficult, 34 items are medium and 1 item is too easy. The differentiating power analysis stated that 2 items had very good distinguishing power, 19 items were good in terms of power, 7 items were quite good in terms of the power of reasoning, 3 items were not good in terms of power and 9 items in the problem were not very good at dissecting power. The assessment used in the even semester XI server administration subject exam has a very high reliability scale with the results of reliability testing of 0.896.

**References**

[1] Sugiyono 2015 *Research and Development Research & Development Methods* (Bandung: Alfabeta)
[2] Rokhayati R 2016 *Validity and Reliability of Summative Tests Made by Indonesian Teachers in Limo District* (Jakarta: Indraprasta PGRI University)
[3] Bloom B S 1956 *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook Cognitive Domain* (New York: Longmans)
[4] Nurgiyantoro B 2011 *Language Learning Assessment* (Yogyakarta: BPFE)
[5] Aiken L R 1989 Three Coefficients for Analytical Reliability and Validity of Ratings *Educational and Psychological Measurement* 45 pp 131-142
[6] Neilsen J 2003 *Usability 101: Introduction to Usability* [Online] Retrieved from: http://useit.com/alterbox/20030825.html Accessed on: January 2019
[7] Krug S 2006 *Don't Make Me Think! A Common Sense Approach to Web Usability* (California: New Riders)
[8] Riduwan 2007 *Scale Measurement of research variables* (Bandung: Alfabeta)
[9] Arikunto S 2005 *Basics of Educational Evaluation* (Jakarta: Bumi Aksara)