Journal for the Education of Gifted Young Scientists, 7(3), 627-641. September 2019
e-ISSN: 2149- 360X
http://jegys.org

Research Article

Epistemological Philosophy of Development of Assessment Instruments Chemistry Learning Outcomes of Science Class Students

Hadarah HADARAH[1] & Rapiah TULHIKMAH[2]

Received: 3 August 2019, **Accepted:** 27 August 2019

Abstract

The aim of study is that the instrument preparation for assesment of chemistry learning. The process starts from the basic competence in determining the indicators that will be assessed, through the development of the instrument lattice, which will be exposed in detail about the items of the instrument to be prepared. The role of ontology examines the nature of the reality of the object of knowledge, axiology as a value theory related to the usefulness of knowledge. The phenomenon is that there are still many teachers who have not followed standard procedures in the preparation and development of learning instruments, the teacher does not make a grid as a guide for making questions. As a result, students are lazy to learn because the learning outcomes obtained of XI Science Class students at SMA Negeri 5 in Takalar City are not in accordance with the ability level of students. This type of research is the development or Research and Development (R&D) using the Analysis Design Development Implementation and Evaluation (ADDIE) model, to develop and design an instrument for evaluating the learning outcomes of class XI science students in chemistry state senior high school (SMA) 5 Takalar City with a focus on cognitive, affective, and psychomotor dimensions that produce valid and reliable instruments.

Keywords:
epistemological philosophy, chemistry learning, instrument development

To cite this article:
Hadarah, H., & Tulhikmah, R. (2019). Epistemological Philosophy of Development of Assessment Instruments Chemistry Learning Outcomes of Science Class Students. *Journal for the Education of Gifted Young Scientists, 7*(3), 627-641. DOI: http://dx.doi.org/10.17478/jegys.605097

[1] Lecture Institute Agama Islam NegeriSyekh Abdurrahman Siddik, Bangkabelitung, Email: hadarah.rajab@iainsasbabel.ac.id ORCID No: 0000-0003-4295-5350
[2] Lecture Institute Agama Islam NegeriSyekh Abdurrahman Siddik, Bangkabelitung, Email: hikmahwahidpep@iainsasbabel.ac.id ORCID No: 0000-0001-9852-2007

Introduction

Evaluation of learning is not a stand-alone activity, but is built by four interrelated components and is a unity. Evaluation activities must involve the other three activities namely assessment, measurement and test (non-test) ". Learning evaluation requires an assessment process because it is a process of determining the assessment of students.(Erniwati, 2018b)

Assessment as part of evaluating learning often involves measurement. In the process of evaluating learning outcomes, measurement has a very important role, namely to obtain data and information in accordance with the objectives of the assessment concerned. Measurement is quantitative, it is as stated by (Erniwati, 2018a) quality is measured not only quantitatively but also qualitatively, and "measurement is basically an activity of determining numbers for an object systematically". Determination of this number is an attempt to describe the characteristics of an object. "Measurement is a procedure comparing the object to be measured with its size. Measuring instruments used as a means for conducting assessments can be in the form of tests and non-tests (Supratiknya, 2012).

Epistimologystudies illustrate that there are still many teachers who have not followed standard procedures in the preparation and development of instruments, including: the teacher does not make a grid as a guideline for making questions or the teacher usually makes questions first and then makes a grid, no study of the points items that have been compiled so that many errors occur in making questions, test procedures and analysis of test items are not carried out.

So students often complain of dissatisfaction with the acquisition of test results. There are those who feel capable and ready when going to do the test but the results obtained are low, and conversely there are those who do not really master the material but the results obtained are high. As a result, students are lazy to learn because the learning outcomes obtained do not match the ability level of students.

The Role of Epistemological Philosophy in the Development of Science

According to Ahmad Tafsir that science is categorized in two parts' natural science and social science. Natural science includes astronomy, physics, chemistry, earth science and life sciences. While social science including medicine is sociology, anthropology, psychology, economics and politics. What is certain is that science consists of theories which basically explain causal relations. Certainly science is not at the level of giving good or bad value, halal or haram, polite or not sopon, beautiful or not beautiful. But science only gives true or true value is wrong. This fact causes some people to think that science is value-free (Tafsir & Tafsir, 2009)

The explanation above, it can be understood that there is a need to control the value of science through actions that are suspected to prevent the occurrence of unexpected symptoms or symptoms that are expected from the development of science (science) through the process of supervision, regulation, and control. The task of supervision and regulation is closely related to the control device.

Aspects of epistemological philosophy of study in development stage, then there is a relationship with moral ontology in science, whereas in the application stage the concept is related to scientific axiology. Ontology is defined as a study of the nature of the reality of the object studied in producing knowledge, while axiology is defined as a theory of values relating to the usefulness of the knowledge acquired. As it is known that every knowledge, including scientific knowledge has three basics namely ontology, epestimology and axiology. In this study Epestimolgi Philosophy is devoted to discussing ways to obtain knowledge which in scientific activities is called the scientific method (Mustofa, 2016) as a method of deepening aspects of Instrument developmeng for Evaluation of Student Learning Chemistry Learning Outcomes.Related to the objectivity of science, it has become a general provision and accepted by various circles that science must be objective. One factor that distinguishes between scientific statements with general assumptions lies in objectivity (Hamdi & Bahruddin, 2015). A scientist must look at empirical reality by ignoring an ideological, religious and cultural awareness. A scientist must be free in determining the topic of study and his research, is free to conduct scientific experiments. Therefore, a scientist works, as far as possible focused on the process of scientific work and the aim so that the study works well. Objective value is only the main goal, do not want to be bound by subjective values (Fitrah, 2018).

Research Problem

What is the epistemological philosophy of developed instrument of learning achievement assessment in chemistry subjects in class XI for odd semester examinations in the form of a multiple choice written test, self-assessment questionnaire, and performance sheets. The developed instrument is expected to meet the valid and reliable instrument criteria (Susilana, Si., & Riyana, 2008) as conducted by Rapiah Tulhikmah on SMAN 5 in Takalar Regency, South Sulawesi Province?

Method

Research Model

This type of research is the development (R&D) by using the Analysis Design Development Implementation and Evaluation (ADDIE). This model which aims to develop and design an instrument for evaluating the learning outcomes of students of class XI Science in SMA Negeri 5 Takalar, Takalar City, on the cognitive dimension, affective, and psychomotor produce valid and reliable instruments. Data collection techniques used in this study test techniques, in quantitative item analysis, item validity test, and instrument reliability test.

Sample/Participants

The object of the research object is students as users of the instrument for evaluating learning outcomes in Chemistry subjects at SMA Negeri 5 Takalar for the odd semester. The test and non-test targets are students in class XI IPA in Takalar State High School 5 Takalar City. It takes odd semester years, namely class XI IPA2, XI

IPA4, XI IPA5, XI IPA6 as many as 158 students. Respondents who had received the whole learning process tested on Chemistry learning outcomes in the form of tests and non-tests. Where for instruments in the way of tests measuring cognitive dimensions and non-tests measure effective and psychomotor dimensions.

Data Collection Tools

For the validity of internal criteria, calculating the biserial correlation coefficient (RBIs) with the validity analysis criteria proposed by Widoyoko (2011). It states "items having a score greater than the minimum standard (0.3) can be declared valid".

Correlation of items is used to determine the internal consistency of each item, useful for analyzing the validity of items, completing the good and poor-quality items. Items with high correlation are high-quality items.

The way to determine correlation is to use the Point-Biserial correlation formula for dichotomy the correlation coefficient obtained by correlating the score of one test item to the overall test score.

Development of instruments based on effective and psychomotor

Briefly the initial research, found data based on the results of observations and interviews with chemistry teachers in making assessment instruments with these three domains, namely cognitive, affective, and psychomotor to see the learning outcomes of students of SMAN 5 Takalar, Takalar Regency. South Sulawesi Province, and the authors feel the need to conduct further studies under the title "Philosophy of Development of Student Chemistry Learning Outcomes Assessment Instrument"

Description of Procedure for Developing Chemistry Learning Outcomes Instruments

Regarding test and non-test instruments, aspects or criteria were formulated to be assessed in the assessment of learning outcomes. Assessment of learning outcomes is seen as an assessment of outcomes and learning processes to measure one or several aspects of the ability of students, but in the sense that it must measure all aspects of the ability of students in the learning process. So that the students' ability assessment of the material that has been taught by the teacher or studied by students gets a number of previous material.

Data Analysis

The data analysis techniques in the form of descriptive analysis qualitatively are as follows.

Qualitative descriptive analysis

a. Qualitatively analyze items

The validity of the items qualitative in the two valid learning instruments and added thirteen reviewers. In this way, reviewers have given a grid of item items to be examined, the format of the reviewers/assessments, and the guidelines for the reviewers/reviewers. In the initial stages, reviewers explain how to examine each

aspect to be assessed from each item of questions, then the next stage for reviewers working independently in different places. The reviewers are welcome to improve directly on the text of questions and statements, and provide comments and values on each item with good criteria, repaired/revised or replaced. The results of the analysis of qualitative items given to reviewers based on aspects assessed in terms of material, construction, and language.

b. Analysis of face validity

Before the trial used, first face validity test must be done to 2 validators plus thirteen reviewers. The form of the instrument that examined from the face validation was in the form of the accuracy of the instrument measuring dimensions, cognitive, affective, and psychomotor in the chemistry learning instrument.

Quantitative descriptive analysis

a. Analysis of validity, reliability and item analysis

Data analysis of test results of test instruments with internal criteria validity calculates the biserial correlation coefficient (RBIs). The production moment correlation coefficient (a) and the reliability test using the KR-20 formula and also the degree of difficulty. The different power and effectiveness of options used with the help of the Microsoft Excel program.

b. Cognitive dimension

1. Validity analysis criteria
2. For the validity of internal criteria, calculating the biserial correlation coefficient (RBIs) with the validity analysis criteria proposed by Widoyoko (2011). He states "items having a score greater than the minimum standard (0.3) can be declared valid".
3. Correlation b Criteria for reliability analysis. To determine the reliability coefficient criteria are high must limit. A test is said to have a steady nature (consistent and stable is 0.70 (Linn, 1989) proposed by Mansyur et al. (2015) (Mansyur, 2017). Internal consistency method determines the formula the reliability coefficient of the consistency of combined items by calculating the score of dichotomous items can use the Kuder-Richarson formula known as KR-20 (Djaali, 2000).
4. But it is used to find out the internal consistency of each item, useful for analyzing the validity of items, completing the good and poor quality items. High-correlated items are quality items.
5. Reliability analysis criteria to determine the reliability coefficient criteria are high because of the minimum. The limit of a test is said to have a steady nature consistent and stable is 0.70 (Linn, 1989) proposed by Mansyur et al. (2015).

c. Affective dimension

Analysis of test results of non-test instruments with empirical validity of the formula (rit), product-moment correlation coefficient (rα) and reliability testing using the Alpha Crombach formula used with the help of the Microsoft Excel program.

d. Psychomotor dimensions

Analysis of the results of the test results of the performance appraisal instruments with the validity of internal criteria calculating the product-moment correlation coefficient (a) and the reliability test using the Cohen's Kappa coefficient (ICC), with the help of Microsoft Excel / SPSS 20.0

From the results of research on the development of the instrument of chemistry learning outcomes in SMA Negeri 5 Takalar theoretically. In the cognitive dimension studied by two experts and 13 PPS UNM students. The face validity analysis revision was in the top category, and the internal consistency coefficient for multiple-choice packages A and B was 0.81 and 0.82. On the affective dimension, the internal consistency coefficient for the self-assessment scale was 0.87. In the psychomotor aspect, the internal consistency coefficient for performance sheets is 1,00.

From the results of the development of instruments, empirically analyzed the validity and reliability. The cognitive aspects of the 50 items tested then obtained 46 valid details with security for multiple-choice test package A 0.77 while for numerous choice test package B are 0.73.

For cognitive aspects at the level of difficulty of the instrument questions for multiple choice package, A shows that 8 test items in the severe category. 32 test items in the medium category and 9 test items in the natural category whereas for multiple-choice package B indicates that 6 test items in the severe category, 34 test items in the medium category, and 10 test items in the natural category.

For the cognitive aspects of the difference in power, the instrument questions for multiple choice package. A indicates that 10 test items are in the top categories, 22 test items are in an outstanding grade, 11 category test items need to be fixed, and 7 test items are in the less or discarded category to choose double package B shows that 10 test items are in an outstanding class, 22 test items are in the top division, 9 test items in the group need to fix, and 9 test items are in the less or discarded category. So for the cognitive aspects of the effectiveness of the option has been functioning correctly. For the affective aspects of the 24 items tested, all things were valid, and reliability was 0.87. For the psychomotor aspects of 30 subjects tested all obtained accurate details with security, one was received.

Results

The Results of Descriptive Qualitative Analysis of Theoretical Data

The results of the descriptive qualitative analysis can be described as follows.

A. Description of the Results of Face Validity Analysis

Cognitive Dimension

Before the multiple choice test instrument was tested, advance face validation was assessed by two validators and 13 PPs UNM students. The shape of the instrument that was examined was the accuracy of the items measuring the indicators, language

clarity and graphics and pictures on the multiple choice test instrument can be seen in table 1.

Table 1.

Distribution of Percentage of Face Validity on the Likert Model Scale Multiple Choice Test Instrument

Component	SA	A	D	DA	SD	Total	Percentage
Provision of items to measure indicators	0	31	19	0	0	50	
	0	62	38	0	0		100 %
Component	**SA**	**A**	**D**	**DA**	**SD**	**Total**	**Percentage**
Language Clarity	0	45	5	0	0	50	
	0	90	10	0	0		100%
Component	**SA**	**A**	**D**	**DA**	**SD**	**Total**	**Percentage**
Graphics and Image	0	24	26	0	0	50	
	0	48	52	0	0		100%

Source: Results of Validity Analysis of Multiple Choice Test Instrument Advance.

(Note=SA: strongly agree, A: Agree, D: doubtful, DA: disagree, SD: strongly disagree).

The results of the analysis of qualitative test items given to reviewers based on aspects assessed in terms of material, construction and language with the number of questions 50 test items of chemical learning instruments for students of SMA Negeri 1 Galesong, Takalar district. As many as 50 item items, tests that did not undergo revision were 34 questions and 16 item items were revised on multiple choice test instruments in class XI in SMA Negeri 5 Takalar, Takalar Regency.

Affective Dimension

Before the questionnaire instrument on the attitude statement item is used first the face validity is assessed by two validators and 13 students of PPs UNM. The form of the instrument that has been validated face is the accuracy of the instrument grain measuring the indicators and language clarity used in each instrument item in the statement sentence from the assessment instrument attitude can be seen in Table 2.

Tabel 2.

Distribution of Percentage of Face Validity on the Likert Model Scale Attitude Assessment Instruments

Component	SA	A	D	DA	SD	Total	Percentage

Provision of items to measure indicators	0 0	18 75	6 25	0 0	0 0	24	100 %
Component	**SA**	**A**	**D**	**DA**	**SD**	**Total**	**Percentage**
Language Clarity	0 0	20 83,3	4 16,7	0 0	0 0	24	100%

Source: Results of analysis of validity of instrument questionnaire.
(Note=SA: strongly agree, A: Agree, D: doubtful, DA: disagree, SD: strongly disagree).

The results of the assessment of two validators and 13 pps UNM students contained 14 unrevised items and 10 items that were revised by the questionnaire instrument for students of SMA Negeri 5 Takalar, Takalar Regency.

Psychomotor Dimension

The instrument of performance evaluation was carried out an analysis of the face validity which was assessed by two validators and 13 PPs UNM students. The form of the instrument that was reviewed from face validation in the form of the accuracy of the instrument measuring indicators and clarity of the language used in each item of the performance appraisal instrument can be seen in table 3.

Tabel 3.

Distribution of Percentage of Face Validity on the Likert Model Scale Performance Appraisal Instruments

Component	**SA**	**A**	**D**	**DA**	**SD**	**Total**	**Percentage**
Provision of items to measure indicators	0 0	8 66,7	4 33,3	0 0	0 0	12	100 %
Component	**SA**	**A**	**D**	**DA**	**SD**	**Total**	**Percentage**
Language Clarity	0 0	9 75	3 25	0 0	0 0	12	100%

Source: Results of validity analysis of multiple choice test instrument advance.
(Note=SA: strongly agree, A: Agree, D: doubtful, DA: disagree, SD: strongly disagree).

Furthermore, from the results of the analysis of qualitative test items given to reviewers based on aspects assessed in terms of material, construction and language with a total of 12 items performance appraisal instruments for grade XI science students of SMA Negeri 5 Takalar, Takalar district.

B. Description of the Results of Content Validity Analysis

Cognitive Dimension

After analyzing the validity of the face and studying the items qualitatively on the multiple choice test instrument, then the validity analysis is assessed qualitatively by

the validator against the instrument tests that have been made in the form of multiple choice test instruments.

The relevance of the two experts is the content validity provides a model to determine the overall internal consistency coefficient (overal) based on judgments of exhibits, namely in the form of an internal consistency coefficient. 75%), then the relevance of the two experts is said to be high, meaning that the measurement of the instrument was declared valid.

The main criteria for determining whether an instrument is appropriate or not depends on the results of the validity of the two validators. To find out whether the instrument being developed meets the level of relevance, the instrument developed needs to obtain correlation and input from the validator about the instrument being developed. -the contents have been written based on indicators. From the indicators developed instrument items.

The results of the analysis between the evaluations of the two validators showed that the internal consistency coefficient value on the cognitive indicator in the form of multiple choice for package A was 0.92 percentage 92% and for package B was 0.82 percentage 82% due to the internal consistency coefficient> 75%, the test instrument Multiple choice for package A and B of the students of SMA Negeri 1 Galesong, Takalar Regency is Valid. If the index V> 0.75, the instrument is declared reliable based on internal consistency.

Based on the evaluations of the two validators, there are a number of remedial notes covering the contents of the questions, sentence structure and item incorporation. Note the improvement of items by the validator on the items of test results of learning outcomes.

Affective Dimension

After analyzing the face validation and analyzing the items qualitatively on the questionnaire instrument, the analysis of the content validity assessed by the two validators was qualitatively related to the assessment of the instrument that had been made in the form of a questionnaire instrument.

The relevance of the two experts is the validity of providing a method of determining the overall content validity (overal) based on judgments of experiments, namely in the form of an internal consistency coefficient. %) Then it is said that the relevance of two experts is high, meaning that the measurement of the instrument made is valid.

The main criteria for determining whether an instrument is appropriate or not depends on the results of the validity of the two validators. To find out whether the instrument developed meets the level of relevance, the instrument developed needs to get correlation and input from both validators about the instrument being developed. For the development of the instrument begins with develop a grid that has been written based on indicators. From the indicators developed instrument items.

The results of the analysis between the evaluations of the two validators showed that there was an internal consistency coefficient value of 1.00 percentage of 100%, because the coefficient of internal consistency was 100%> 75%, then the measurement of the questionnaire instruments for the students of SMA Negeri 5 Takalar in Takalar District was valid.

Discussion

Data on the results of research on the philosophy of epistimology on the development of chemistry learning instrument results in the type of research and development that aims to produce instruments (Nurwahida, Danial, & Mansyur, 2018) states that "The criteria include valid and reliable. The valid and reliable nature is considered by the high reliability of the measurement results of a test ". Chemical learning outcomes are valid and reliable, complemented by supporting learning activities will provide an attitude of confidence and motivation towards students in providing an objective assessment (Jennings, Sycara, & Wooldridge, 1998) Chemistry learning instrument that is suitable for use by teachers in schools is very helpful in the assessment process. This can be described as follows.

Description of the Results of Data Analysis in terms of Theoretical

Epistemological, the process of measuring tools for the assessment of learning outcomes by following the stages that are structurally and systematically through instructions based on what goals and objectives will be measured as (Susilana, Si, & Riyana, 2008) dan (Nurwahida et al., 2018) offers another perspective on validity. Validity is described as "an integrated evaluation of the degree of empirical evidence and theoretical basis that supports the coverage and appropriateness of actions with conclusions based on test scores or other modalities of assessment".

In the opinion of the expert, it is also similar that states that validity has many dimensions, in this study will discuss three aspects, namely the validity of the content (content validity), the validity of the construct (construct validity), and the instructional validity (instructional validity) (Susongko, 2010).

Developing and designing chemical learning outcomes instruments conceptually through processes of face validity, item validity qualitatively and instrument validity from the results of systematically and philosopied development produce good instruments and are able to measure the cognitive, affective and psychomotor dimensions of students in participating in chemical learning. For the purposes of developing representative representational points, the development of those points must be based on a grid plan. Validation testing is carried out by examining the appropriateness of items measured with plans as outlined in the grid. The criteria on which the validation test is based are planned grille. Valuation items are declared valid, if after examining the contents of the items written have shown compliance with the grid (Kinyua & Okunya, 2014).

Content validity also has a very important role with a theoretical rational analysis that provides an illustration of how an instrument is validated with consideration by both validators carried out in the following manner. First, the two validators are asked to carefully observe all items in the instrument to be validated (Stiggins, 2002). Secondly, the expert validator is asked to correct the interpretation of the items that have been made. Third, at the end the expert's consideration of all aspects to be measured has gone through interpretation of statement items or questions in the instrument. In other words, which has been reflected as the purpose of the instrument.

The results of the cognitive, affective, and psychomotor dimensions have been assessed by two validators and 13 PPM UNM students. The results of the analysis show that the formulation of indicators for each item of multiple choice test instruments, questionnaires, and performance appraisal of 50, 24, 30 items are found in R (Doubtful) and T (Right), Language clarity is found in R (Doubtful) and J (obviously). As for the compound compound for each item of the multiple choice test instrument, and the assessment found in R (Hesitate) and B (Good) can measure the ability and skills of students of class XI IPA in SMA Negeri 5 Takalar, Takalar Regency.

Based on the results of the study showed the three components of the multiple choice test instrument, questionnaire, and performance appraisal, both components are in the good category meaning the instrument of Chemistry learning outcomes in class XI of SMA Negeri 5 Takalar, Takalar Regency on the cognitive, affective, and psychomotor dimensions are in the good category.

Furthermore, the results of the qualitative analysis of items given to reviewers based on aspects assessed in terms of material, construction, and language with multiple choice test instruments and performance assessment of 50 items, 24 items, 30 items were revised.

Determine the overall internal consistency coefficient of a test through expert judgment on multiple choice test instruments. This is to evaluate the test items assessed by two experts, whether these items are included in the high content validity coefficient. Ellis & Levy(Ellis & Levy, 2009) suggested that internal consistency is one way to show reliability, according to Ruslan if the results of the content validity coefficient (V> 75%), then it can be stated that the measurement of the instrument carried out is valid (Ruslan, 2018).

Based on the results of the inter-appraisal analysis by the two validators on the multiple-choice test instrument, questionnaire and performance appraisal showed that the internal consistency coefficient was 0.92; .82; 0.87; 1,00. So that the measurement of the instrument in terms of cognitive, affective and psychomotor dimensions on the instrument of chemistry learning outcomes of students of SMA Negeri 5 Takalar, Takalar Regency can be declared valid.

Based on the description above it can be interpreted that the results of the instrument in terms of the cognitive, affective, and psychomotor dimensions of the chemistry learning instrument of class XI students of SMA Negeri 5 Takalar Takalar Regency can meet valid criteria.

The Description of the Results of Data Analysis is Reviewed Empirically

A good instrument is an instrument in which the teacher and students play an equally important role in the class, (Sudjana, 1995) in an effort to improve and enhance learning in the classroom, a teacher who is professional in teaching must have a measuring tool that is ready to provide students in terms of instruments of cognitive, affective, and psychomotor dimensions, so that improving the learning process will be even better, if an educator prepares everything that will be the goal and target in improving the quality of education in the subject of chemistry in class XI IPA of SMA Negeri 5 Takalar, Takalar Regency.

The results of research on attitudes on the components of cognition, affection and conation as indicated on the results of the questionnaire instrument class XI students in SMA Negeri 5 Takalar, Takalar Regency. The results of the instrument using the validity of the correlation coefficient used criteria (> 0.30) and the value of the reliability coefficient is thermally high, because the minimum limit of a test is said to have a steady nature (consistency and stability) is 0.70 (Nurwahida et al., 2018).

The questionnaire instrument on the attitude statement was tested as many as 50 items to 158 students of class XI in SMA Negeri 5 Takalar, Takalar Regency. The last version of scale shows 46 valid items and 4 items that were wasted. Based on the results of the study on the performance evaluation questions sheet and performance assessment instruments were tested on 40 students of class XI IPA2 in SMA Negeri 5 Takalar, Takalar Regency. Performance appraisal instruments for agreement measures with significant use ($\alpha = 0.05$) were proposed by Landis & Corh (2012). Performance appraisal instrument as many as 30 items questions, thus one-time trial was conducted with the results of the assessment of the two observers collected analyzed using statistical methods Cohen's Kappa SPSS 20.0 application assistance program.

Furthermore, the results of the study used the analysis of the ICC Kappa method on the performance research question sheets and performance appraisal instruments with the calculated results using the class correlation coefficient method (Intraclass Correlation Coefficients)(Fleiss & Cohen, 1973) obtained a reliability value of 1,000, meaning that the performance appraisal instrument could meet the criteria declared reliable.

To find the level of item difficulty item discrimination and biserial point of correlation provide useful information with each question reducing how the questions on the test relate to each other and the overall problem (Adams & Wieman, 2011).

Sudijono stated that, the items of the learning achievement test items could be stated as good items, if the items were not too difficult and not too easy in other words the degree of difficulty of the items was easy, moderate, and sufficient (Sudijono, 1998).

Based on the results of research with the analysis of the level of difficulty items on multiple choice test instruments with a total of 50 items there are 50 items in the medium category, 0 items in the sufficient category and 0 items in the easy category. As an explanation (Mardapi, 2004) that the value of the distinguishing items is less than 0.3. The number 0.3 is the minimum criteria number of a good item raised by (Nurwahida et al., 2018) meaning that the number 0.3 is a limit that has a good differentiating power can distinguish groups of participants who have high abilities and groups of students who have low ability.

The results of further research on the distractor (deception) the effectiveness of the option that the distractor is declared to have been able to function properly if the distractor has been chosen at least by 5% of all test takers stated in Sudijono, (Sudijono, 1998) based on the results of research data analysis of the percentage of effectiveness option on multiple choice test instruments out of 50 test items that are included in the good category as many as 46 test items and the unfavorable category 4 items. Based on the results of research on the effectiveness of options on multiple choice test instruments students have functioned well (Kolstø et al., 2006).

Data analysis of the percentage of option effectiveness on multiple choice test instruments from 50 test items included in the good category totaling 46 test items and the unfavorable category 4 items. Based on the results of research on the effectiveness of options on the multiple choice test instrument students in class XI SMA Negeri 5 Takalar, Takalar Regency.

The results of the study presented by Bahia & Natel, explain developing a measurement instrument to measure customer information satisfaction using reliability testing, content validity, criterion related validity, convergent validity, discriminatory validity, and nomological validity.(Bahia & Nantel, 2000).

Relevant research results were presented by several experts who showed that highly experienced teachers provide high validity and reliability testing. In the same way, the teachers aim at the construction test and the test analysis is more or less valid and reliable. According to Stiggins (Stiggins, 2002) and (Kinyua & Okunya, 2014) the level of teacher education generally influences the teacher in making test testing and validity.

Conclusion

The learning achievement test in chemistry subjects in class XI for the SMAN 5 odd semester exam is in accordance with development procedures that guarantee that a good quality of learning outcomes can be obtained, so that the measurement results can accurately reflect the learning outcomes achieved by students after participating in learning activities. The decision on the assessment of learning outcomes achieved

by students after participating in learning activities. The decision to evaluate learning outcomes is arranged according to the criteria for making tests. Therefore it is necessary to increase insight and skills in developing learning outcomes test instruments that pay attention to the objectives of chemistry subjects as well as compliance with assessment standards. Preparation and development of tests intended to obtain a standardized test.

The results of the research on the multiple choice test instrument using the internal consistency method in calculating the reliability coefficient used the KR-20 formula with a value of 0.92 for multiple choice package A and 0.82 for multiple choice of package B with the criteria for reliability coefficient (> 0.70 So 0.92 and 0.82>0.70 means that it can meet the criteria is reliable, then the results of the questionnaire instrument using the Alpha reliability coefficient formula of 0.814 and 0.812 with the Alpha reliability coefficient criteria (0.814> 0.70) for package A and the criteria Alpha reliability coefficient (0.812> 0.70) for package B means the questionnaire multiple choice test instrument can fulfill the stated reliable criteria.

Biodatas of Authors

Dr. Hadarah, M. Ag Lecture of IAIN Syaikh Abdurrahman Siddik Bangka Belitung, Research Focus on Sufism Philosophy of Saince. Sufism Teaching, Akhlak Tasawuf Teaching.
Affiliation: IAIN Syaikh Abdurrahman Siddik Bangka Belitung, Indonesia
E-mail : hadarah.rajab@iainsasbabel.ac.id

Rapiah Tulhikmah, M. Pd Lecture of IAIN Syikh Abdurrahman Siddik Bangka Belitung, Research Focus on research and educational evaluation. Statistic Method Teaching, Micro Teaching.
Affiliation: IAIN Syaikh Abdurrahman Siddik Bangka Belitung, Indonesia
E-mail : hikmahwahidpep@iainsasbabel.ac.id

Acknowledgements

Finally, I am indebted to my parents, my beloved son Muhammad Nur Idham Khalid, my precious daughter Siti Aida Khadijah Khalid for their prayers and also support and encouragement for my pursuit. I want to thank the leader of IAIN Sas Bangka Belitung, and all the participants that had given me their energy and mind for the realization of this journal. In particular, thanks to the editors, and reviewers who helped with this journal.

References

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312.

Bahia, K., & Nantel, J. (2000). A reliable and valid measurement scale for the perceived service quality of banks. *International Journal of Bank Marketing*, *18*(2), 84–91.

Ellis, T. J., & Levy, Y. (2009). Towards a Guide for Novice Researchers on Research Methodology: Review and Proposed Methods. *Issues in Informing Science & Information Technology*, *6*.

Erniwati, E. (2018a). *No TitleEvaluasi pembelajaran bukanlah kegiatan yang berdiri sendiri, melainkan dibangun oleh empat komponen yang saling terkait dan merupakan suatu kesatuan. Menurut Mansyur dan Suratno (2009) bahwa "kegiatan evaluasi harus* (p. 8). p. 8.

Erniwati, E. (2018b). *Pengembangan Instrumen Penilaian Hasil Belajar Ilmu Pengetahuan Alam Pada Siswa Kelas V Sekolah Dasar Inpres Loka Kabupaten Bantaeng*. Universitas Negeri Makassar.

Fitrah, M. (2018). *Metodologi penelitian: penelitian kualitatif, tindakan kelas & studi kasus*. CV Jejak (Jejak Publisher).

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619.

Hamdi, A. S., & Bahruddin, E. (2015). *Metode penelitian kuantitatif aplikasi dalam pendidikan*. Deepublish.

Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, *1*(1), 7–38.

Kinyua, K., & Okunya, L. O. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal*, *2*(2), 61–71.

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., … Ulvik, M. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education*, *90*(4), 632–655.

Mardapi, D. (2004). Penyusunan tes hasil belajar. *Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta*.

Mustofa, I. (2016). Jendela Logika dalam Berfikir; Deduksi dan Induksi sebagai Dasar Penalaran Ilmiah. *EL-BANAT: Jurnal Pemikiran Dan Pendidikan Islam*, *6*(2), 1–21.

Nurwahida, N., Danial, M., & Mansyur, M. (2018). *Pengembangan Perangkat Pembelajaran Kimiaberbasis Discovery Learningpada Materi Larutan Penyangga Di Kelas Xi Sma*. Universitas Negeri Makassar.

Ruslan, Z. A. (2018). *Pengembangan Lkpd Kimia Berbasis Pembelajaran Islam Terpadu Melalui Model Pbl Di Smait Ibnu Sina Makassar Pada Materi Pokok Larutan Elektrolit Dan Nonelektrolit*. Universitas Negeri Makassar.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765.

Sudijono, A. (1998). *Pengantar evaluasi pendidikan*. PT RajaGrafindo.

Sudjana, N. (1995). *Penilaian hasil proses belajar mengajar*. PT Remaja Rosdakarya.

Supratiknya, A. (2012). Penilaian hasil belajar dengan teknik nontes. *Yogyakarta: Universitas Sanata Dharma*.

Susilana, R., Si, M., & Riyana, C. (2008). *Media pembelajaran: hakikat, pengembangan, pemanfaatan, dan penilaian*. CV. Wacana Prima.

Susongko, P. (2010). No Titlekesahihan memiliki banyak dimensi, dalam kajian ini akan membahas tiga aspek, yaitu kesahihan isi (content validity), kesahihan konstruk (construct validity), dan kesahihan instruksional (instructional validity. *Cakrawala: Jurnal Pendidikan*, *4*(8), 1–7.

Tafsir, A., & TAFSIR, A. (2009). *Filsafat Ilmu: Mengural Ontologi, Epistemologi dan Aksiologi Pengetahuan*. Remaja Rosdakarya.

Zaini, M. (2017). Kontrol Nilai Terhadap Sains. *Substantia*, *19*(1), 37–52.