

ITEM QUALITY ANALYSIS INSTRUMENT IN THE COLLEGE ENTRANCE TEST IN THE VIEW OF THE EXPERT VALIDATOR

Akbar iskandar¹, Muhammad Rizal H²

^{1,2}STMIK AKBA

¹akbar.iskandar06@gmail.com

²rijakakba@gmail.com

ABSTRACT

The purpose of this study was to determine the quality of the test instrument is based on the views of experts or expert jugmen. This type of research is a kind of ex post facto research. Data collected by the method of observation, interviews and documentation. Mechanical analysis of qualitative and quantitative data. Benefits theoretically is going to contribute a theory of quality analysis test instrument. While the practical benefits to be expected, namely: For researchers, increasing competence in analyzing the quality of the test instrument with a wide range of applications. For the Foundation, obtain information accurate test results and real. As for prospective students, provide information in accordance with the actual situation and self-motivation towards the better. The results of the research instrument test according to the views of experts it appears that the grain instrument does not meet the standards of developing a test because the test results validity by expert tests generated in the category validity of the low and the IT expert in middle category while the results of reliability test was found to be in the category of reliability sufficient so need to do a bit of revision.

Keywords: *instrument, test, Validator.*

INTRODUCTION

One cause of damage to the quality of education is the results of the entrance test inaccurate. To that end, the correct assessment would provide the right information and mendorong in increasing motivation and achievement in student learning. It is also explained by Iskandar (2013: 37) states that the test system and good judgment will encourage students to increase motivation and achievement in learning.

But what often happens in the world of education, we are often faced with the problem of making decisions, whether a student should repeat a particular material, appropriate pass or should not pass. It is not an easy task. It takes careful consideration in order to produce the right decision and the right so as not to harm the students. To that end, the right decision and was strongly influenced by the quality of the instrument used. If the quality of the instruments ugly then the decision is also likely to have a negative impact.

The instrument is a tool used to measure an object, therefore, the instrument must meet the criteria either. Terms instrument is good at least qualify valid and reliable.

Besides qualify valid and reliable also must consider the characteristics of the grain that is the level of difficulty, different power, and the functioning of detractors. This is in accordance with the opinion of Mansyur et al (2015: 30) which says that in order to obtain accurate information it needed a reliable and valid instrument.

Results of preliminary observations in STMIK AKBA found that the questions used in the acceptance of new students not through empirical analysis and asked for the views of experts related item test instrument made. This was revealed by one of the officers executing the test. So that we can be sure the information obtained is biased in the sense that the information is not in accordance with what is desired.

Literature review

Broadly speaking, what is meant by the instrument is a tool that can be used to measure an object. In the field of research, the instrument is defined as a tool to collect data on the variables of research to the needs of research, while in education instrument used to measure students' learning achievement.

Sudrajat said the (2008) assessment (assessment) is the application of a variety of ways and use a variety of assessment tools to obtain information about the extent to which the results of student learning or achievement of competence (circuit capability) learners. So that the results of the assessment process gave birth to decisions relating to students include the placement of students in different educational programs, giving grades to students, guiding and directing the student, the selection of students to participate in educational programs, awards and certificates to the students' competencies.

The test was defined as an instrument or systematic procedure to observe and describe one or more characteristics of the students by using a numerical scale or classification scheme (Nitko & Brookhart, 2007: 7). The test is a way of assessment is designed and implemented to students at a particular time and place and in a condition that meets certain requirements are clear (Sudrajat, 2008).

Sax (1980: 13) says that "*a test may be defined as a task or series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes*". The test was defined as a task or set of tasks that are used to obtain systematic observations were considered to be representative of the characteristics or attributes related to educational or psychological.

On the basis of the above understanding, there is a common thread that seems agreed that the achievement test results of student learning is one way to

explore the capabilities that already possessed by students after participating in the learning process for a certain time. Measurement is a process of giving numbers to obtain a numerical description of the value that is owned by a student against certain characteristics.

Nitko & Brookhart (2007: 7) says that "*measurement is a defined as a procedure for assigning numbers (usually called scores) to specified attribute or characteristic of a person in such a way that the numbers describe the degree to which the person possesses the attribute*". Measurement is defined as a procedure to give a numerical rating (usually called score) for the specification of attributes or characteristics of a person where the value that indicates the degree possessed by that person to attribute being measured. Thus, the measurement includes the determination of a value of the results of a test based on specific rules (for example, calculate the correct answer).

Assessment is often equated with evaluation means. Actually the term ratings is a translation of the term assessment, not a translation of the term evaluation. Both of these terms (assessment and evaluation) actually have similarities and differences.

Both have an understanding assess or determine the value. The difference lies in the context of its use. Assessment is used in the context of a more narrow and usually only carried out internally by the very people involved in the system in question. The evaluation is used in a broader context and usually to a limited level of good programs and a broader level.

Feature assessment can be characterized by an object or program that is assessed and their criteria as a basis for comparison between the reality or what criteria or what it should be, in other words that the assessment is a systematic procedure that includes the activities of collecting, analyzing and interpreting the information to make a conclusion.

The word comes from the Latin test testum, which means a tool for measuring soil. So the test can be defined as the number of questions that need answers or statements which must be given in response to gauge the level of a person's ability or reveal certain aspects of the person subjected to the test.

Mechanical testing is a fact that the man in her life is different between one individual with another individual. No two individuals have the character or the other the same precision, both in terms of physical and psychological terms.

Because to the differences in the individual, it is necessary to create a tool to diagnose or measure the individual circumstances, and gauges that are commonly called tests. With the measuring device that is in the form of the test, then people will succeed in knowing the differences between individuals. Because of the psychological aspect that is different to distinguish one individual to another individual, then later raised also a variety of tests.

Entrance test or selection test in college is often known as "test sieves" or "entrance exam". These tests were performed in the hiring of new students, where the test results are used to select prospective students are classified as the good of the many candidates who took the test.

As a follow up of the results of the test selection, the candidates deemed meet the minimum requirements specified otherwise as participants passing tests and can be accepted as a freshman, in addition to prospective students who did not pass the test, but given the unacceptable as a freshman.

Analysis of test items are generally intended to identify the size of the index level of difficulty, different power and effectiveness index distractor those items are concerned. Analysis of the test can be

done using one of two ways, depending on which test theories used. The theory test can be either classical test theory or modern test theory.

Analysis of the quality of the test is a step that must be taken to determine the degree of the quality of the test, in whole or items that are part of the test. In the assessment of learning outcomes, test samples are expected to describe the behavior and generate value objectively and accurately. If the tests used lecturer unfavorable, the results obtained was certainly not good anyway. This can be detrimental to students themselves, meaning that the results obtained by the students to be not objective.

Therefore, the tests used should be of good quality. Tests should be prepared based on the principles and procedures for the preparation of the test. Once used to be seen whether the test of good quality or not it is necessary to test the quality analysis (Arifin, 2012).

Validity and reliability are the main things that must be met in order to determine the quality of assessment instruments. In a test of validity of the contents of the various ways in which the goal is to see the deal experts in assessing the overall content.

According to Gregory (2007) showed how far the content validity of the questions, tasks or items in a test or instrument capable of representing the overall and proportionally behavior of samples subjected to such treatment. This means that existing instruments reflect the overall content or materials being tested or that should be controlled proportionally. Gregory formula (Column D divided by A + B + C + D) or $V_i = \frac{D}{A+B+C+D}$ with the aid of a 2x2 cross tabulation table as below:

		Less relevant score 1-2	Very relevant score 3-4
Rater 2	Less relevant score 1-2	A	B
	Very relevant score 3-4	C	D

Information:

Vi = Content Validation

A = Both raters do not agree

B = Rater I agree, disagree rater II

C = Rater I do not agree, rater II agree

D = Both raters agreed

Criterion validity of the content:	
0,8 – 1	= Validity is very high
0,6 – 0,79	= High validity
0,40 – 0,59	= Validity was
0,20 – 0,39	= Validity low
0,00 – 0,19	= Validity is very low

Reliability of the instrument indicate the stability, consistency, constancy, reliability of the instrument even though its use in time and under different conditions. The higher the coefficient of correlation between the results of two tests measuring parallel, meaning consistency between the two is getting better and the second measuring instrument called as a reliable measuring instrument Azwar (1996: 180).

The lower limit of reliability coefficient used for an assessment / good test that is equal to 0.70 according to Linn (Mansyur, 2009: 134), while Fleiss (Widhiarso, 2012: 15) categorizes the level of inter-rater reliability into four (4) categories, among others:

$K < 0.4$: *Bad*

$0.4 \leq K \leq 0.60$: *Fair*

$0.60 < k \leq 0.75$: *Good*

$k > 0.75$: *Excellent*

RESEARCH METHODS

This study was included in the category of Ex Post Facto type of research that is systematic empirical research where researchers do not control the independent variables directly because of the existence of these variables has occurred, or because basically these variables can not be manipulated Karlinger (Emzir 2011).

This research was conducted at the college STMIK AKBA in the city of Makassar South Sulawesi Province with the object of study all the answers to the test results of new students in 2014-2016.

RESULTS AND DISCUSSION

Implementation of this research is conducted in accordance with the original plan and assisted by a team of experts as well as the validator is as assessors to test instruments that will be the object of research. Before conducting the first research assessment instrument test device must be made in order to assess the test instrument used in entrance exams for new students in college.

The first step has been carried out to know the quality of the test instrument is used to test the validity of the content and continued with the reliability test is based on the views of experts. Here are described the results of the data analysis has been done.

Descriptive analysis results

Here will be described in descriptive assessment of validator tests and IT experts. More details can be seen in table 4.1 for the clue aspects of each validator.

Table 4.1. Instructions aspect questions

		Freque ncy	Percen t	Valid Percen t	Cumulativ e Percent
Valid	it is not in accordance with	1	50.0	50.0	50.0
	corresponding	1	50.0	50.0	100.0
	Total	2	100.0	100.0	

Based on the analysis above table it appears that of the two validators experts test states of different things that one person (50%) stated that the clue is not

appropriate, and 1 (50%) also validator expert stating the appropriate tests. More details can be seen in the following graph:

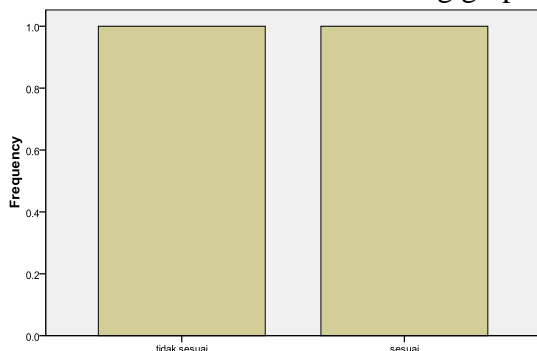


Figure 1. Aspects Instructions matter

In addition to the descriptive analysis of the results of the previous, the next aspect of coverage about where to look for items with the correspondence between the lattice can be viewed as descriptive of each assessor.

Table 4.2 suitability items with grating

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less appropriate	2	100.0	100.0	100.0

Based on the analysis above table it appears that of the two validators experts said the same test that is both validator field test (100%) stated that the items are not in accordance with the lattice. More details can be seen in the following graph:

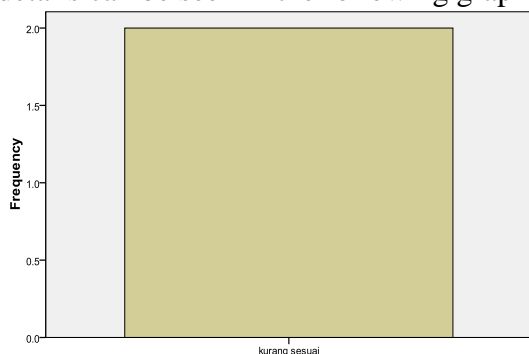


Figure 2. Correspondence between the items with the grating

Furthermore, apart from the results of previous descriptive analysis, further aspects of the coverage about where to look for correspondence between the order of the questions with the material to be tested. From the results of the analysis can be seen as descriptive of each assessor as shown in Table 4.3.

Table 4.3. correspondence between the order of the questions with the material to be tested

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	corresponding	2	100.0	100.0	100.0

It appears from the results of the analysis of the two validators experts said the same test that is both validator field test (100%) stated that the items fit between the order of the questions with the material to be tested. More details can be seen in the following graph:

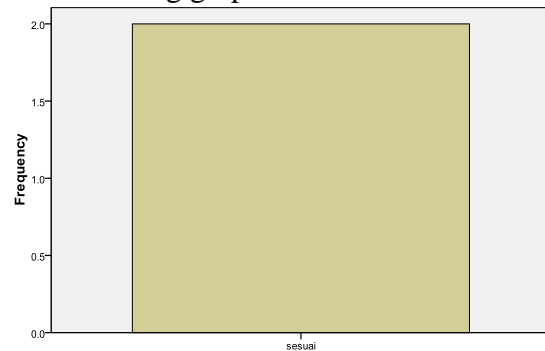


Figure 3. Correspondence between the order of the questions with the material to be tested

Further apart than the previous descriptive analysis results, further aspects of the coverage about where to look for correspondence between questions designed to measure the indicators to be achieved. From the results of the analysis can be seen as descriptive of each assessor as shown in Table 4.4.

Table 4.4. Correspondence between the problems created by the indicators to be achieved

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less appropriate	1	50.0	50.0	50.0
	very appropriate	1	50.0	50.0	100.0
	Total	2	100.0	100.0	

According to the table above analysis it appears that of the two validators experts test states of different things that one person (50%) said problems

created by the indicators to be achieved is less appropriate, and one person (50%) also validator expert tests were expressed very corresponding. More details can be seen in the following graph:

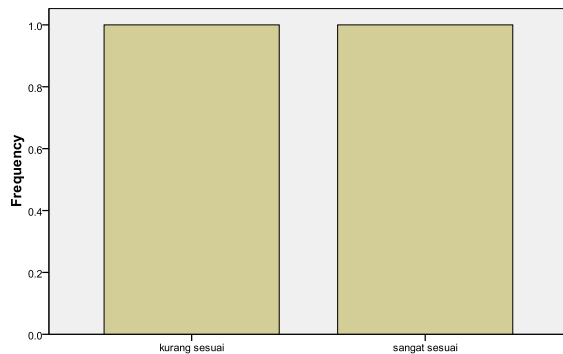


Figure 4. Correspondence between questions that are made with the indicator to be achieved

Much further apart than the previous descriptive analysis results, further aspects of the language in which to see the use of appropriate language to Indonesian rule. From the results of the analysis can be seen as descriptive of each assessor as shown in table 4.5.

Table 4.5. the use of language in accordance with the rules of Indonesian

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Corresponding	2	100.0	100.0	100.0

It appears from the results of the analysis of the two validators experts said the same test that is both validator field test (100%) stated that the use of appropriate language to Indonesian rule to be tested. More details can be seen in the following graph:

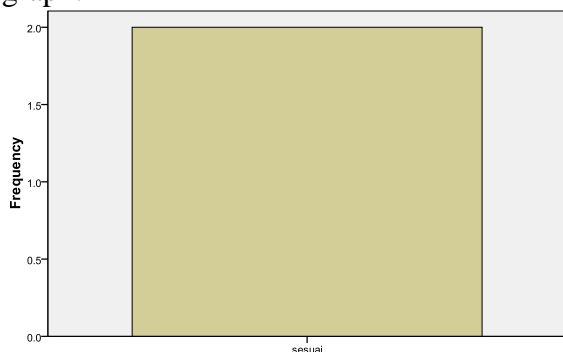


Figure 5. The use of language in accordance with the rules of Indonesian

Then further aspect of the language in which to see the use of language in terms of the formulation of the question is communicative. From the results of the analysis can be seen as descriptive of each assessor as shown in Table 4.6

Table 4.6. The formulation of the question communicative

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under less	1	50.0	50.0	50.0
	Corresponding	1	50.0	50.0	100.0
	Total	2	100.0	100.0	

According to the table above analysis it appears that of the two validators experts test states of different things that one person (50%) stated that the matter created have questions communicative less appropriate, and one person (50%) validator expert test states suitable for about made have questions communicative. More details can be seen in the following graph:

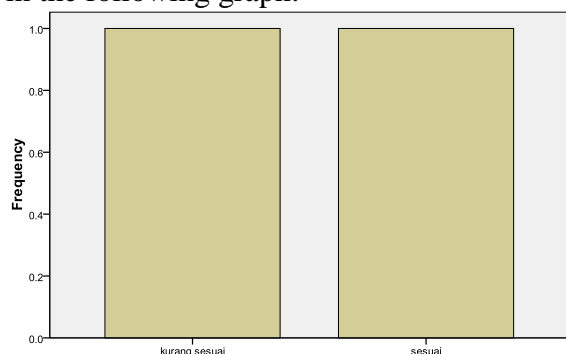


Figure 6. The formulation of the question communicative

Then much further still aspects of the language but who wants to know is related to the use of the phrase is easily understood or not. From the results of the analysis can be seen as descriptive of each assessor as shown in table 4.7

Table 4.7. The use of the phrase in question

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less appropriate	1	50.0	50.0	50.0

Corresponding	1	50.0	50.0	100.0
Total	2	100.0	100.0	

Based on the results table above analysis it appears that of the two validators experts test states of different things that one person (50%) stated that the matter is made to have the sentence is less appropriate in this case is difficult to understand, and one person (50%) validator expert test states suitable for a matter that is made has a question sentence easier to understand. More details can be seen in the following graph:

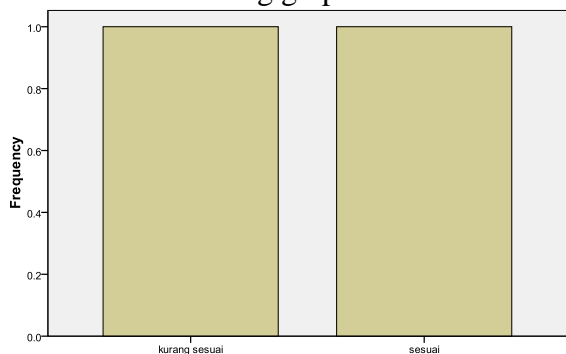


Figure 7. The use of the phrase in question

In addition to descriptive analysis results from previous test experts have also been analyzed by descriptive results of the assessment of an IT expert validator. More details can be seen in table 4.8 for the clue aspects of each validator IT expert.

Table 4.8. Instructions aspect questions

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid it is not in accordance with	1	50.0	50.0	50.0
less appropriate	1	50.0	50.0	100.0
Total	2	100.0	100.0	

According to the table above it appears that the results of the analysis of the two validators IT experts expressed different things that one person (50%) stated that the clue is not appropriate, and one person (50%) validator IT expert who expressed less appropriate. More details can be seen in the following graph:

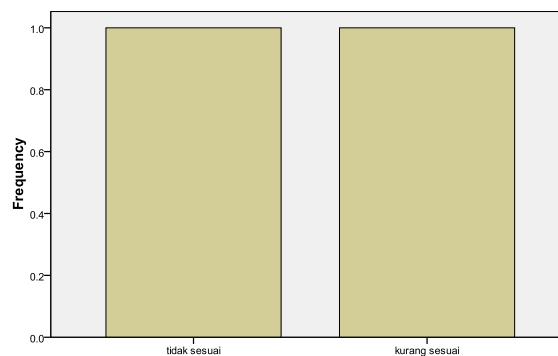


Figure 8. Aspect Hint question

Aside from the previous descriptive analysis results, further aspects of the coverage about where to look for items with the correspondence between the lattice can be viewed as descriptive of each outcome assessors validator IT expert.

Table 4.9. suitability items with grating

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid less appropriate	1	50.0	50.0	50.0
corresponding	1	50.0	50.0	100.0
Total	2	100.0	100.0	

Based on the analysis above table it appears that of the two validators IT experts expressed different things that one person (50%) validator IT experts stated that the items are not in accordance with the lattice. While one person (50%) stated that the validator items in accordance with the lattice. More details can be seen in the following graph:

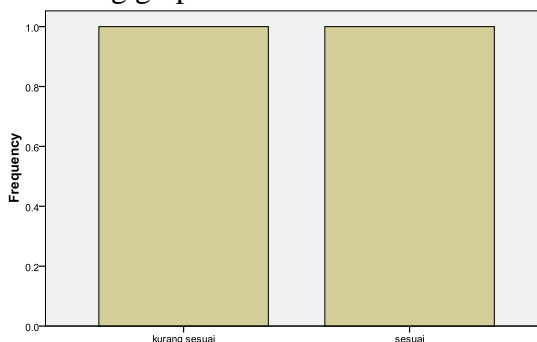


Figure 9. Correspondence between items with the grating

Furthermore, apart from the results of previous descriptive analysis, further aspects of the coverage about where to look for correspondence between the order

of the questions with the material to be tested. From the results of the analysis can be seen as descriptive of each assessor as seen in Table 4:10.

Table 4.10. correspondence between the order of the questions with the material to be tested

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid corresponding	1	50.0	50.0	50.0
very appropriate	1	50.0	50.0	100.0
Total	2	100.0	100.0	

Based on the analysis above table it appears that of the two validators IT experts expressed different things that one person (50%) validator IT experts claim that that the items fit between the order of the questions with the material to be tested. While one person (50%) stated that the validator items fits between the order of the questions with the material to be tested. More details can be seen in the following graph:

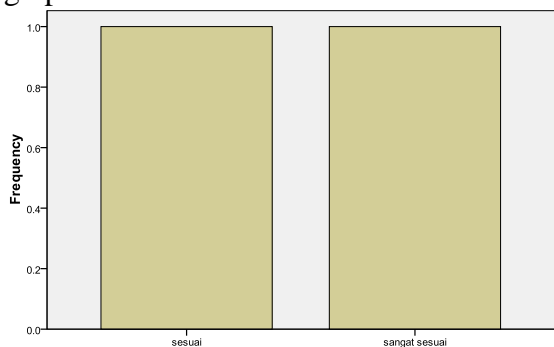


Figure 10. Correspondence between the order of the questions with the material to be tested

Further apart than the previous descriptive analysis results, further aspects of the coverage about where to look for correspondence between questions designed to measure the indicators to be achieved. From the results of the analysis can be seen as descriptive of each IT expert as seen in Table 4:11.

Table 4.11. Correspondence between the problems created by the indicators to be achieved

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid less appropriate	1	50.0	50.0	50.0
corresponding	1	50.0	50.0	100.0
Total	2	100.0	100.0	

According to the table above it appears that the results of the analysis of the two validators IT experts expressed different things that one person (50%) said problems created by the indicators to be achieved is less appropriate, and one person (50%) also validator IT expert stated accordingly. More details can be seen in the following graph:

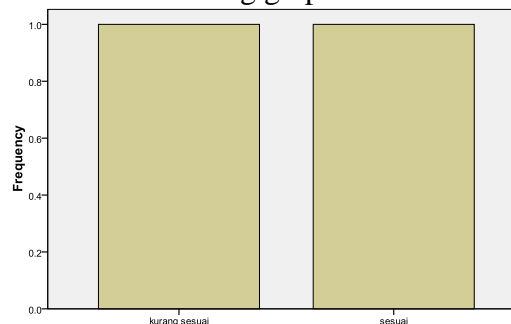


Figure 11. Correspondence between questions that are made with the indicator to be achieved

Much further apart than the previous descriptive analysis results, further aspects of the language in which to see the use of appropriate language to Indonesian rule. From the results of the analysis can be seen as descriptive of each appraiser validator IT expert as seen in Table 4:12

Table 4.12. the use of language in accordance with the rules of Indonesian

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid sesuai	2	100.0	100.0	100.0

It appears from the results of the analysis of the two validators IT experts expressed the same thing that both the IT field validator (100%) stated that the use of appropriate language to Indonesian rule to be tested. More details can be seen in the following graph:



Figure 12. Use of languages in accordance with the rules of Indonesian

Then further aspect of the language in which to see the use of language in terms of the formulation of the question is communicative. From the results of the analysis can be seen as descriptive of each assessor as seen in Table 4:13.

Table 4.13. The formulation of the question communicative

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less appropriate	1	50.0	50.0	50.0
	Corresponding	1	50.0	50.0	100.0
	Total	2	100.0	100.0	

According to the table above analysis it appears that of the two validators IT experts expressed different things that one person (50%) stated that the matter created have questions communicative less appropriate, and 1 (50%) validator expert IT field states appropriate for about made have questions communicative. More details can be seen in the following graph:

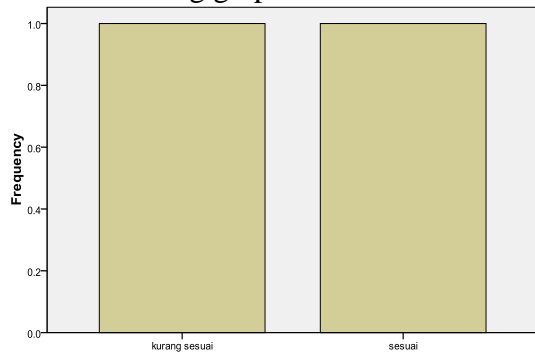


Figure 13. The formulation of the question communicative

Then much further still aspects of the language but who wants to know is related to the use of the phrase is easily

understood or not. From the results of the analysis can be seen as descriptive of each assessor as seen in Table 4:14.

Table 4:14. The use of the phrase in question

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid corresponding	2	100.0	100.0	100.0

According to the table above analysis it appears that of the two validators IT experts say the same thing that is appropriate for a matter that is made has a question sentence easier to understand. More details can be seen in the following graph:

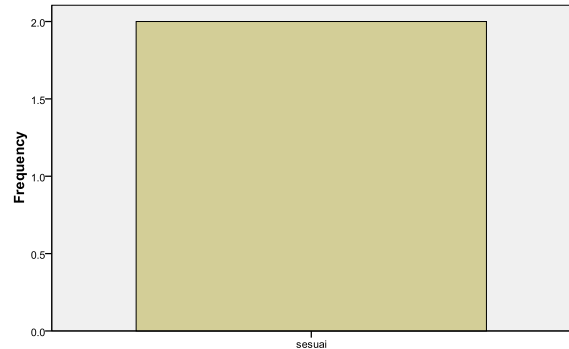


Figure 14. Use of the phrase in question

The Content and Validity of Test Results Test Reliability

In a test of the validity of the various ways that can be used with the aim of testing the quality of an instrument. But the validity test used here is the content validity test which aims to look at the deal from experts in assessing the overall content of the content. These aspects are considered in validating these instruments are: manual, coverage problems, language.

table 4:15 is a summary of the validation test instrument for every aspect of observation. From the results of expert judgments totaling 2 of lecturers as experts in tests, was further supplemented by two IT experts as validator, each of which has the ability in the field of IT, so that the instrument can be used with many revisions.

Table 4:15. Recap result of the assessment of two Expert Tests

Rater 1	Rater 2	From cross-tabulations
1	3	C
2	2	A
3	3	D
4	2	B
3	3	D
3	2	B
2	3	C

Based on the results of a 2x2 cross-tabulations above, are then put into the formula Gregory $V_i = \frac{D}{A+B+C+D}$ then the result $V_i = \frac{2}{1+2+2+2} = 0,28$. Based on the analysis conducted found content validity V_i value of 0.28. So that this instrument meets the criteria of validity of the content in the low category.

Table 4.16. Recap assessment results 2 IT expert

Rater 1	Rater 2	From cross-tabulations
2	1	A
3	2	B
3	4	D
2	3	C
3	3	D
3	2	B
3	3	D

Based on the results of a 2x2 cross-tabulations above, are then put into the formula Gregory $V_i = \frac{D}{A+B+C+D}$ then the result $V_i = \frac{3}{1+2+1+3} = 0,42$. Based on the analysis conducted found content validity V_i value of 0.42. So that this instrument meets the criteria of validity of the content in the medium category.

Furthermore, after the test the validity then continued with the reliability test to see the consistency of the results of the assessment expert validator. Based on the reliability test results it appears that the level of agreement of 4 validator expert in providing an assessment based on an analysis result reliability coefficient of

intraclass correlation coefficient (ICC) and the results were compared with the minimum criteria of reliability. More details can be seen as the result of analysis on a table 4:17.

Table 4.17. Reliability test results by 4 validator expert

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df 1	df 2	Sig
Single Measures	.209 ^b	-.099	.708	2.058	6	18	.110
Average Measures	.514 ^c	-.565	.907	2.058	6	18	.110

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Based on the analysis of reliability of 4 (four) expert validator above it appears that the level of agreement between the validator 1 (one) to validator 4 (four) obtained a value of $K = 0514$. The coefficient of reliability of some assessors (intraclass Correlation Coefficients) this instrument included in the category enough reliable level (fair). The coefficient of reliability of the instrument is smaller than the lower limit prescribed reliability that is equal to 0.70 according to Linn (Mansyur, 2009), so that the instrument does not qualify the criteria of reliable.

Based on the results of previous analysis it appears that the validity and reliability of the test results do not meet the

standard criteria that correspond to the development of tests. Due to the results obtained from the analysis found that the validity of the test results is still in the low category as well as with the reliability test results of the test and IT experts found that the agreement between the four appraisers stated that the existing instruments do not meet the standard criteria for reliability that is equal to 0.70. while the reliability test results obtained amounted to 0.514 in the category of the reliability sufficient.

CONCLUSION

Based on the analysis and discussion it appears that the existing instruments do not meet the standard criteria for the development of tests for the validity of test results generated by test experts are in the category of low validity and IT experts in middle category, while the results of reliability test was found to be in the category of reliability sufficient so need to do a bit of revision, for the further research are expected to developing a test to repair the item instrument that does not meet the criteria for developing a test to follow the suggestions given by experts.

ACKNOWLEDGEMENT

The implementation of this research can be done well thanks to the help of various parties, therefore, on this occasion the author, expressed gratitude and highest appreciation to:

1. Leaders STMIK AKBA who have given the widest possible opportunity to carry out this research.
2. Leaders STMIK AKBA Research Institute which has provided facilities and services and good cooperation during the activity.
3. TIM implementing STMIK AKBA tests that have been willing to accept and permit the implementation of these activities so that research can be done well.

BIBLIOGRAPHY

- Azwar, Saifuddin. 1996. Tes Prestasi. Fungsi dan Pengembangan Pengukuran Prestasi Belajar. Yogyakarta: Pustaka Pelajar
- Arifin, Z. 2012. Evaluasi Pembelajaran Prinsip, Teknik Prosedur. Bandung: PT.Remaja Rosdakarya
- Emzir. 2011. Metodologi penelitian Pendidikan Kuantitatif dan kualitatif. Jakarta: Rajawali Pers
- Gregory, R.J. (2007). *Psychological Testing: History, Principles, and Applications (5th Edition)*. New York: Pearson Education Group, Inc.
- Iskandar, Akbar. 2013. Pengembangan perangkat penilaian psikomotor pada Sekolah Menengah Kejuruan (SMK). Jurnal inspiration jurnal teknologi informasi dan komunikasi. Vol.3/No.1/ 37-46.
- Mansyur, Rasyid. H, & Suratno. 2009. Asesmen Pembelajaran di Sekolah. Yogyakarta: Multi Pressindo.
- Mansyur; Rasyid, Harun & Suratno. 2015. Asesmen pembelajaran disekolah. Panduan bagi guru dan calon guru. Yogyakarta: Pustaka Pelajar
- Nitko, Anthony J, & Brookhart, Susan M. 2007. Educational Assessment Of Students. New Jersey: Pearson Education.
- Sudrajat, Ahmad. 2008. Pengembangan Perangkat Penilaian Psikomotor.([http://akhmadsudrajat.files.wordpress.com/2008/08/penilaian psikomotor.pdf](http://akhmadsudrajat.files.wordpress.com/2008/08/penilaian-psikomotor.pdf), diakses pada 20 Januari 2012).
- Sax, Gilbert. 1980. Principles Of Educational And Psychological Measurement And Evaluation. Second Edition. Belmont: Wadsworth Publishing Company.
- Widhiarso, Wahyu, 2012. Mengestimasi Reliabilitas.(<http://widhiarso.staff.ug>



INTERNATIONAL CONFERENCE ADRI - 5

“Scientific Publications toward Global Competitive Higher Education”

m.ac.id/files/bab_2_estimasi_reliabil
itas_via_spss.pdf, diakses pada 12
Februari 2012).