

The Estimation of Standard Error Measurement of Physics Final Examination at Senior High Schools in Bima Regency Indonesia

Syahrul Ramadhan^{1,*}, Djemari Mardapi¹, Chuduriah Sahabuddin², Rudy Sumiharsono³

¹Department of Evaluation, Yogyakarta State University, Indonesia

²Universitas Al Asyariah Mandar, Indonesia

³Institut Keguruan dan Ilmu Pendidikan (IKIP) PGRI Jember, Indonesia

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This study aims to find out reliability and estimating the standard error measurement of the questions which have been designed by the teacher. The data analysis was based on the response of 635 students towards the test instrument of Physics Final Exam of grade XII in six schools, which were located in Bima Regency. Test instrument reliability was analysed by using Itean 4.3 software. Standard Error Measurement analysis of the six test instruments was done by using Compound Binomial Method, Anova Method and Itean. Index of test instrument reliability in State Senior High School 1 and State Senior High School 3 is less than 0,7, while the test instrument in State Senior High School 2 Sape, MAN, Muhammadiyah Senior High School and PGRI Senior High School are more than 0,7. Standard error measurement extremely depends on the error, which is caused by the test instrument, in this case, the reliability of some tests. The result of this study shows the teachers' need for training in designing proper test questions.

Keywords Estimation, Error Measurement, Test Instrument

1. Introduction

In carrying out the evaluation, measurement is a step that must be taken. When an evaluation process will be carried out, it should be preceded by a careful measurement process. Educational measurement is an activity to quantify symptoms or objects. Symptoms or objects can be in the form of motivation, achievement, confidence or achievement, all of which are expressed in the form of numbers [1]. The numbers obtained from the measurement

results will help the researcher/teacher know the level of success of the learning process carried out. Of course, the information obtained through measurement is the result of the process of analysing data with certain techniques. Then the teacher conducts an intensive assessment to see the success of the learning process, whether the numbers obtained through measurements have met the standards of success or vice versa [2, 3].

The success of the learning process can only be seen through the measurement process. To carry out measurements, a good instrument (measuring instrument) is needed, which can reflect the actual ability of the student. A good measuring instrument will provide accurate information about the success of learning. On the contrary, a measuring instrument that is not good will produce a standard error of measurement.

Making good questions is one of the abilities that must be possessed by a teacher [4]. Good questions are important for teachers to get the true results of the success of the learning process. Questions / Instruments that are not good will make it difficult for teachers to get information about the success of the learning that has been done and cause a standard error of measurement.

Reliability is an important requirement that must be owned by the test kits. Reliability shows the extent to which a test kit is consistent or steady in getting information when used again at different times [5].

Knowing reliability and magnitude of standard error measurement is important for an educator. This is because someone's ability can be seen from the magnitude of the standard error measurement that occurs in the testing process. The smaller the standard error measurement, the more visible the score closer to the actual score, on the contrary, the greater the standard error measurement, the score that appears to be farther away from the actual score [3, 6].

Table 1. Number of items and research samples

No	Statistics	Schools						Total
		SSHS 1 SAPE	SSHS 2 SAPE	SSHS 3 SAPE	MAN SAPE	PGRI SHS	Muhammadiyah SHS	
1	<i>N of Items</i>	30	30	35	30	30	25	180
2	<i>N of Examinees</i>	202	98	70	85	120	60	635

1.1. Statement of the Problem

Semester Final Examination (UAS) is an educational evaluation conducted every six months. For teachers of grade XII at Senior High School, the Semester Final Examination is expected to be able to see the readiness of children in facing the National Examination (UN), because the material of grade XII is 50% of the total UN question material in senior high school. In order for the results of the National Exam to be able to describe the actual results, the question instruments used must meet all the requirements as a good and tested measuring instrument in various aspects. The problem that arises now is whether UAS questions are a good measurement tool that can reflect the abilities of students. Based on this, a research is needed to see the magnitude of the standard error in measurement.

This study aims to estimate the reliability of the test equipment and estimate the error in measuring the questions designed by the teacher. In general, the teachers in Bima District did not carry out the analysis of reliability and standard errors in the measurement process they did. So, with this research, it is expected to be able to inform the teacher about the quality of the questions they make. This result can also be an input for improving education in Bima Indonesia.

2. Research Method

2.1. Type of Research

This research is categorised as Ex post facto research, because this research describes the events that have occurred, to know the reliability and how much of the standard error of measurement that occurs in the testing process using the *Binomial Compound Method*, Variant Analysis Method and IteMan.

2.2. Research Sample

This research was carried out in Bima District in December 2014 until January 2015. The object of this study was all students of grade XII answer sheets in Semester Final Examination of Senior High School at Physics subjects designed by teachers in the 2014/2015 school year. The students were students of Science major. The student answer sheets used in this study were obtained from 6 senior high schools in Sape Sub-district, Bima

District, NTB Province, which had different test kits for each school.

2.3. Data Collection Techniques

The data collection technique used in this study is document analysis. Researcher directly met Physics subject teachers who know about the questions and student answer sheets that have been tested in the 2014/2015 school year. The data taken was student answer sheets and questions in the Semester Final Examination in the 2014/2015 school year.

2.4. Data Analysis

Reliability analysis was done by using the IteMan 4.3 computer program. The value of received reliability is at least 0.7. Analysis of the standard error measurement with IteMan was based on the output results from the IteMan 4.3 software. Data from each school were analysed. The magnitude of the Standard Error Measurement (SEM) value can be obtained in the output section, which then became the basis of the standard error measurement values for each instrument made by the teachers.

Analysis of the standard error measurement with the Binomial Compound Method emphasises the selection of items by dividing the test into several parts. The process of selecting the items is based on the parameters of the items' level of difficulty of the questions, which is equivalents. The level of difficulty is sought using the help of a computer program IteMan 4.3. After dividing the test items into parts, the next step is to estimate the standard error of measurement based on the Binomial Compound Method by using the following formula [7-9].

$$S_{E(i)} = \left[\sum_{h=1}^c \frac{X_{ih}(k_h - X_{ih})}{k_h - 1} \right]^{1/2} \tag{1}$$

Information:

$S_{E(i)}$: Measurement error for test taker i

X_{ih} : Scores obtained by the test taker i in the clusters or groups of test items on the h category on a test kits specification.

k_h : Number of items on the h category

Standard error measurement analysis using the Variable Analysis Method is done by looking at the question score, which is considered as a two-way factorial design without replication [7, 9]. Each item is considered as a treatment. The first step is to find the mean square of interaction

between test takers and items. Syaifuddin Azwar [10] suggests a simple formula in determining the magnitude of the standard error of measurement as follows:

$$MS_{Sxl} = \frac{\left(i - \frac{\sum X^2}{k} - \frac{\sum Y^2}{n}\right) + \frac{\sum i^2}{nk}}{(n-1)(k-1)} \tag{2}$$

Information:

MS_{Sxl} : Means Square interaction between test takers and item questions

X: The value obtained by each test taker on the whole item

Y: The overall score of the test takers on one item

Next is estimate the standard error measurement with the Analysis Method Variance by using the following formula

$$S_E = [k(MS_{Sxl})]^{1/2} \tag{3}$$

Information:

S_E : Measurement Error

K: Number of Item Questions of a test kits

MS_{Sxl} : Means Square interaction between test takers and item questions

3. Results

This study discusses the estimation of reliability and the standard error measurement of UAS questions at the physics subject of grade XII in six schools in Bima District. The results obtained from each estimate are very diverse. The six test kits are designed by different teachers. Some teachers are civil servants and certified with a service period of more than ten years, such as in Senior High School 1 Sape and Senior High School 2 Sape. The several other school teachers, such as Senior High School 3 Sape, MAN, PGRI Senior High School and Muhammadiyah Senior High School, are honorary teachers who have work experience which is still less than five years and not yet become a civil servant.

The first step is to calculate the reliability estimate and proceed with estimating the magnitude of standard errors measurement of UAS questions at Physics subjects of grade XII in the six schools. The results of the reliability analysis and estimation calculation of the standard error measurement will be presented in full as follows.

3.1. Reliability Analysis

Reliability analysis is done using the help of Iteman 4.3 computer program. The analysis results are the following.

The test kits used by Senior High School 1 Sape and Senior High School 3 Sape has a reliability value of 0.680 and 0.695 while the value of reliability received is at least 0.7, which means that the test kits of these two schools have not met reliability requirements. The test kits for Senior High School 2 Sape, MAN Sape, PGRI Senior High School and Muhammadiyah Senior High School have

fulfilled the element of reliability because they have values above 0.7.

Table 2. Reliability of the test kits

School	Reliability	Criteria	Decision
State Senior High School 1 of Sape	0.680	> 0.7	Not Reliable
State Senior High School 2 of Sape	0.750		Reliable
State Senior High School 3 of Sape	0.695		Not Reliable
State Madrasah Aliyah Sape	0.718		Reliable
PGRI Senior High School	0.805		Reliable
Muhammadiyah Senior High School	0.793		Reliable

3.2. Analysis of Standard Error Measurement

The estimation of standard error measurement of Senior High School Semester Final Examination test kits at Physics subject is done based on three different estimation methods, namely the Binomial Compound Method, Variance Analysis Method, and SEM Iteman results. These three methods produce variable standard error values that vary. The variation in the estimation value of standard error measurement in the methods is due to differences in the steps and the basis of the formula used in calculating the estimation value of standard error measurement. The results of the calculation of the estimation value of standard error measurement can be seen in Table 3.

Table 3. SEM Analysis Results

Schools	Analysis Model of Standard Error		
	Compound binomial	Varian Analysis	Iteman
SSHS 1	0.927	2.370	2.364
SSHS 2	1.035	1.853	2.042
SSHS 3	0.974	2.052	2.580
MAN	1.159	2.058	2.296
PGRI	0.988	1.997	2.237
Muhamadiyah	1.028	2.192	2.172

The estimation of standard error measurement of Semester Final Examination Physics test kits based on Iteman is quite varied, where the smallest to largest values are sequentially from Senior High School 2 Sape, Muhammadiyah, PGRI, MAN, Senior High School 1 Sape and Senior High School 3 Sape. Based on the method of variance analysis, the smallest to largest value starts from Senior High School 2 Sape, PGRI, Senior High School 3 Sape, MAN, Muhammadiyah and Senior High School 1 Sape. Based on the Binomial Compound method, the smallest to largest standard error is from Senior High School 1 Sape, Senior High School 3 Sape, PGRI, Muhammadiyah, Senior High School 2 Sape and MAN. A good instrument is expected to have a minimum standard

error measurement, which means that it is getting closer to the results of the actual abilities of students.

4. Discussion

Assessment is part of teacher responsibility that is needed in learning. The development of students in schools must be measured to see what competencies students need to have and they have gotten through the learning process. Campbell & Levin [11] states that assessment in learning is an important strategy for connecting learning strategies and classroom practices to needs individuals, progress, and student learning results.

Based on the results of the analysis, it shows that the test kits used by Senior High School 1 Sape and Senior High School 3 Sape have a reliability value of 0.680 and 0.695 while the value of reliability received is at least 0.7, which means that the test kits of these two schools have not met reliability requirements. The consequences of these results are the results of the assessment carried out by the teacher not fulfilling the requirements because the assessments are not able to measure the true abilities of the students due to the inaccuracy of the instrument's measurement used. Furthermore, these two schools are public schools, the questions kits at Senior High School 1 Sape are prepared by teachers who are civil servants with tenure of more than ten years. Whereas the question kits at the Senior High School 3 Sape was prepared by honorary teachers whose work experience was still less than five years and not yet a civil servant status. These results indicate that the teacher's ability to arrange questions does not always correlate with the teacher's status and the length of service. These results also indicate that teachers need to be given training and coaching to improve their ability to develop questions.

Another concept that also reflects the consistency of a test is the standard error measurement. The standard error measurement is an estimation of how a researcher expects an error from a test that has been made. The estimation of standard error measurement of the UAS Physics test kits based on Iteman is quite varied, where the smallest to largest values are sequentially from Senior High School 2 Sape, Muhammadiyah, PGRI, MAN, Senior High School 1 Sape and Senior High School 3 Sape. From these data, it can be seen that the two schools, which have the largest SEM are the schools with the lowest reliability. These results are by the statement from Miller [5] which states that "*SEM is a quantitative expression of the magnitude of error in the test score based on the test reliability*". The statement indicates that the standard error in measurement depends on the error caused by the test tool itself, in this case, the reliability of a test. The higher the coefficient level of the reliability of the test score, the smaller the standard error of measurement, on the contrary, the lower the level of the reliability coefficient of the test score, the greater the standard error of the measurement.

The estimation of standard error measurement of the UAS Physics test kits based on the variance analysis method, the smallest to the largest value starts from Senior High School 2 Sape, PGRI, Senior High School 3 Sape, MAN, Muhammadiyah and Senior High School 1 Sape. Based on the Binomial Compound method, the smallest to largest standard error is from Senior High School 1 Sape, Senior High School 3 Sape, PGRI, Muhammadiyah, Senior High School 2 Sape and MAN. The differences in each of these methods may be due to differences in the concepts of each method and the factor causing the occurrence of standard error measurement. Standard error measurement (SEM) has two categories, namely random error measurement and systematic error measurement. Random error measurement can be caused by poor physical and mental conditions. Systematic error measurement is a measurement error caused by the measuring instrument itself [1, 12].

Errors in measurement are a natural thing. This is because to make a truly reliable test or measuring instrument is difficult. Kubiszyn & Gary [13] state that the Notion of error in testing is very similar. No test measures perfectly, and many tests fail to measure as well as we would like them to. In line with the statement by Kubiszyn & Gary, Biggs [14] states that making an error is quite natural because nothing can be labelled as 'complete' or 'perfect' itself. However, if it is referring to the results of this study, the needs of teachers for training in the preparation of good and correct questions is a must.

5. Conclusions

The results of this research show that the test instrument used by State Senior High School 1 Sape and State Senior High School 3 Sape still do not meet reliability requirements. Test instruments for State Senior High School 2 Sape, MAN Sape, PGRI Senior High School, and Muhammadiyah Senior High School have fulfilled the components of reliability because they exceed the score 0,7. Standard error measurement extremely depends on the mistake caused by the test instrument itself, in this case, is the test reliability. Based on the research results, the researchers recommend Department of Education in Bima Regency to conduct training for improving teachers' competence especially training of making good questions, which are reliable and have standard error measurement as small as possible. The principal is expected to direct the teachers to always improve their pedagogical ability as the support in the learning process by following the training of competence improvement. The teachers are expected to follow the training to improve their pedagogical competence to make reliable questions and have standard error measurement as small as possible. Teachers' knowledge about the estimation of standard error measurement is still really less. Therefore, it is an

opportunity for the researcher to investigate different respondents in the province or other regencies.

Classroom application and practice," ed: Hoboken, NJ: John Willey & Sons, 2010.

[14] J. B. Biggs, Teaching for quality learning at university: What the student does: McGraw-hill education (UK), 2011.

Acknowledgements

This work was financially supported by Lembaga Pengelola Dana Pendidikan (LPDP). We would like to express our institution, Yogyakarta State University that has provided many experiences in term of research. Our deepest gratitude to students who participated in our experiment research and also who have contributed to the collection of our empirical data. Thankfully to the teacher of senior high school who helped this research

REFERENCES

- [1] D. Mardapi, Assessment measurement and educational evaluation. Yogyakarta: Nuha Litera 2012.
- [2] J. Gardner, Assessment and learning. Los Angeles: Sage Publications Ltd 2012.
- [3] C. R. Reynolds, R. B. Livingston, V. L. Willson, and V. Willson, Measurement and assessment in education: Pearson Education International Upper Saddle River, 2010.
- [4] R. J. Wright, Educational assessment: Tests and measurements in the age of accountability: Sage Publications., 2007.
- [5] P. W. Miller, Measurement and teaching: Patrick W. Miller & Associates, 2008.
- [6] W. A. Mehrens and I. J. Lehmann, Measurement and evaluation in education and psychology: Holt, Rinehart and Winston New York, 1978.
- [7] L. S. Feldt, M. Steffen, and N. C. Gupta, "A comparison of five methods for estimating the standard error of measurement at specific score levels," Applied psychological measurement, vol. 9, pp. 351-361, 1985.
- [8] R. L. Brennan and W.-C. Lee, "Conditional scale-score standard errors of measurement under binomial and compound binomial assumptions," Educational and Psychological Measurement, vol. 59, pp. 5-24, 1999.
- [9] S. Ramadhan, "A literature study: The comparison of standard error estimation methods," in Seminar nasional evaluasi pendidikan II, 2015.
- [10] S. Azwar, Reliability and Validity. Yogyakarta: Pustaka Pelajar 2014.
- [11] C. Campbell and B. Levin, "Using data to support educational improvement," Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), vol. 21, p. 47, 2009.
- [12] D. Mardapi, The technique of test and non-test instrument design. Yogyakarta: Mitra Cendikia 2008.
- [13] T. Kubiszyn and G. Borich, "Educational and measurement: