

# The Effect Number of Replication and the Number of Option Scale toward the Reliability Coefficient of Maximal in the Rubric Assessment of Vocational Learning Outcome

Wardani Rahayu\*, Zainal Abidin

Universitas Negeri Jakarta, Jl. Rawamangun Muka, Rawamangun, Jakarta

\*Corresponding author: [wardani.rahayu@unj.ac.id](mailto:wardani.rahayu@unj.ac.id)

**Abstract** This study aims to determine the effect of number of replication and the number of option scale toward the maximal reliability coefficient in the assessment instrument rubric. This research used experimental method. The population is all students of state vocational high school of food processing technology of Cianjur in the 2014/2015 academic year. This research used t-test analysis. The result of the research shows that maximal reliability coefficient is higher with more and more option scale on instrument of rubric. Number of data replication has an effect on maximal reliability coefficient. The maximal reliability coefficient gets higher with more and more data replication on instrument of.

**Keywords:** number of replication, number of option scale, rubric, maximal reliability

**Cite This Article:** Wardani Rahayu, and Zainal Abidin, "The Effect Number of Replication and the Number of Option Scale toward the Reliability Coefficient of Maximal in the Rubric Assessment of Vocational Learning Outcome." *American Journal of Educational Research*, vol. 5, no. 6 (2017): 645-649. doi: 10.12691/education-5-6-9.

## 1. Introduction

One of the special characteristic of vocational education is productive vocational learning. In a productive learning, students are expected to master a certain skill competencies selected according to their interests. The main activity in the productive vocational learning is vocational practice activities. [1] (Berg, 2002). Vocational productive learning is learning that emphasizes practices to create a product of a certain kind based on a certain skill. Therefore, about 70% of the learning activity in vocational education is focused more in the lab. Vocational productive practices aim to train and prepare students to be skilled in his/her art. So that when the students graduated from vocational schools they are able to work according to their competence skills.

To measure the performance of a learner in conducting a particular practice an instrument called a rubric is used. Rubric is the instrument used to determine the criteria of what should be mastered by the learner. Rubric is a tool used to give score with specific expectation of an assessment. This specific expectation is referred to as the assessment criteria. In a rubric assessment criteria must be made in such a way that it clearly describes the competences of students from the lowest to the highest level. [2] (Stevens & Levi: 2005).

To eliminate the element of subjectivity in assessing the performance of a learner, the instrument of a rubric should

be made as precise as possible. Good instrument is an instrument that has high reliability and validity. Valid means the instrument measures what should be measured, if it measures the student's ability in practice, the instrument must be able to measure such ability instead of measuring others. Reliable means those instruments consistently deliver results that are relatively the same if used to measure other similar subjects.

Etymologically reliable means capable. The term reliability is used to explain how consistent a test in making a measurement. Reliability refers to the scores achieved by the same people when they are tested by repeating the same test on different occasions, or a set of different equivalent items. In other words, reliability is the consistency of the measuring instrument [3,4,5] Reliability is defined as the degree of consistency between the two different measurements. [6] The concept of reliability can be explained as the level of correlation between the scores on a two parallel tests [7]

Maximum reliability coefficient was introduced by Li, Rosenthal, and Tiles which is an extension of Spearman-Brown coefficients in k components. This coefficient is named as the maximum coefficient since the estimation of the reliability is done based on combination of linear sum of a set of items which is optimal in explaining the construction of the measurement. Estimation of the reliability tests is conducted by correlating a linear combination of (X) with a latent construct ( $\eta$ ). [8] By using algebraic manipulations, this concept is eventually derived to an equation to calculate the maximum reliability below [9]

$$\rho_{\max} = \frac{\sum_{i=1}^k \frac{\lambda_i^2}{\theta_i}}{1 + \sum_{i=1}^k \frac{\lambda_i^2}{\theta_i}}$$

Remark:

$\lambda_i$  = Standardized factor loading indicator-*i*

$\theta_i$  = error variance indicator-*i*.

Research related to the reliability of the rubric is the research conducted by Anders Jonsson and Gunilla Svingby [10], Mary Kathryn Thompson et al. [11]. Whereas the reliability discussed in these studies is the reliability that is still commonly used such as alpha reliability, generability coefficient, and Cohens Kappa reliability. So the maximum reliability study on the rubric involving many scale options which was conducted by the researchers is deemed as new and has never been done.

## 2. Method

This research was conducted ta class XII at SMK Karangtengah, SMKN Cikalong, SMKN 2 Cilaku and SMKN Pacet. This research was conducted in the second semester of 2014/2015. This study belongs to experimental research. In this study the average coefficient of maximum reliability on 3 options scale data group is compared to 4 option-scale replication of 20 times, the average coefficient of maximum reliability data group of 3 options scale is compared with 4 options scale on the 30 times replication, the average coefficient of maximum reliability of group of data which is replicated 20 times are compared with data group of 30 times replication on the 3 options scale, the average coefficient of maximum reliability group of data group with 20 times replication are compared to data group with 30 times replication in 4 options scale, in order to obtain four groups of data.

The first independent variable in this research is the number of options of assessment scale, the second independent variable is the number of replication. The number of assessment scale is 3 and 4 options. While the number of replication is 20 and 30 times. In this study, the dependent variable is the maximum reliability coefficient.

The population in this study was all students of public SMK in Products Processing Technology competency skills in Cianjur. The population covered in this study was students who took the performance test. Students are given the performance tests in the manufacture of yoghurt, and they are assessed by the assessment rubric instruments. The total number of students assessed by the two types of rubric instruments are 300 students. The sampling technique adopted in this research is simple random sampling (with replacement) or with returns. This is because the population is considered equivalent. The same applies in the selection of schools in the district level which was also done randomly.

The procedures in this study were as follows: a random sample conducted on a total of 325 students by means of a draw (sampling with replacement). Out of the 325 students, 2 groups were formed, which each group

consists of many as 150 students. Two groups of respondents were given tests of performance and assessment through the assessment rubric. A total of 150 students were assessed using three options assessment criteria and 150 students were assessed using option 4 assessment criteria. Calculation of reliability coefficient scores of both groups with maximum reliability coefficient formula. Replication of data on both groups of as many as 300 samples bootstrapping were done 20 times and 30 times for each group, and then statistical tests by using t test was conducted.

## 3. Research Findings

Description of the data consists of four groups: maximum reliability coefficient of assessment rubrics instruments with 3 and 4 on a scale options with 20 times replication, maximum reliability coefficient group of assessment rubric instruments with 3 and 4 options on the scale with of replication as much as 30 times.

In the boxplot Figure 1, it can be interpreted that: 3 reliability coefficient distribution of options 20 replications is more homogeneous than the distribution of data of reliability coefficient of 4 options with 20 replications, as well as 3 and 4 options with 30 replications; distribution of reliability coefficient of with maximum 3 and 4 option scales with 20 replications shows that most of the data is collected at a high score so that the distribution of data reliability coefficient is negative; while the distribution of the reliability coefficient data for 30 times replication at 3 option scale shows that most of the data collected is of low score so that the distribution of data in the form reliability coefficient is positive; while the distribution of reliability coefficient data for 4 options with 30 Replication times is symmetrically shaped.

The mean of reliability coefficient of 3 options with 20 replication was 0.8906580, the mean of reliability coefficient of 4 option with 20 replications was 0.8951573, the mean of reliability coefficient of 3 options with 30 replications was 0.8972, the mean of reliability coefficient of 4 options with 30 replication is 0.9011 ; whereas the reliability coefficient variant of 3 options with 20 replications is 0.0000586, reliability coefficient variant of 4 options with 20 replication is 0.0000666, reliability coefficient variant of 3 options with 30 replications is 0.0000593, reliability coefficient variant of 4 options with 30 replications is 0.000078.

Research hypothesis testing is conducted by a t-test with independent samples which is performed to determine the maximum reliability coefficient differences between the instrument section with 3 options and 20 times replication, 4 options with 20 times replication, 3 options with 30 times replication, and 4 options with 30 times replications.

Based on the results of independent samples t test calculations, the difference in average maximum reliability coefficient of the instrument section with three options and 4 options is shown and the difference in the average coefficient of reliability on a sample size of 20 and 30 times replication. In reliability coefficient of rubric instrument with 3 dan 4 options and replications of 20

times, based on *t-test*,  $t_{count} = 1.7986 > t_{table} = 1.6859$  and as such  $H_0$  is rejected which means that the average coefficient of maximum reliability of rubric instrument of 4 scale options is higher than 3 scale options with 20 times replications.

In reliability coefficient of rubric instrument with 3 and 4 options and replications of 30 times, based on *t-test* with independent samples,  $t_{count} = 1.8143 > t_{table} = 1$ , hence  $H_0$  which means that the average coefficient of maximum reliability of rubric instrument of 4 scale options is higher than 3 scale options with 30 times replications.

At maximum reliability coefficient of rubric instrument of 3 scale options with 20 and 30 replications, based on paired *t test*,  $t_{count} = 2.9395 > t_{table} = 1.6772$ , thus  $H_0$  is rejected, which means that on average a maximum of 30 replications, the reliability coefficient is higher than 20 replications in the rubric instrument with 3 scale options.

At maximum reliability coefficient of rubric instrument of 3 scale options with 20 and 30 replications, based on paired *t test*  $t_{count} = 2.7003 > t_{table} = 1.6772$ , such  $H_0$  is

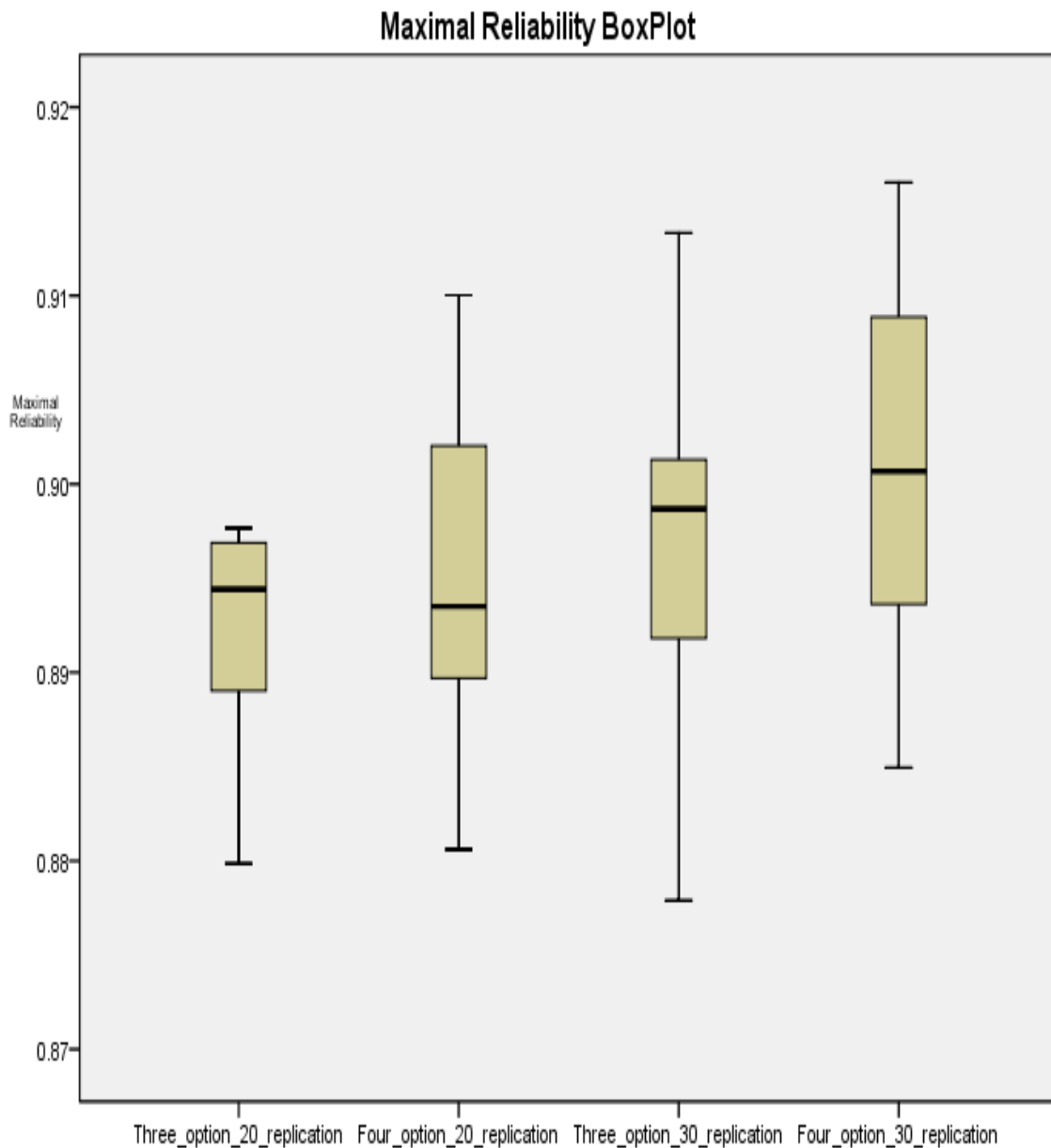
rejected which means that the average maximum reliability coefficient of 30 replications is higher than 20 replications in rubric instruments of 4 scale options.

**Table 1. Recapitulation of Result of *t-test***

No	Group	$t_{count}$
1.	$\rho_{max1}$ and $\rho_{max2}$	1.7986
2.	$\rho_{max3}$ and $\rho_{max4}$	1.8143
3.	$\rho_{max1}$ and $\rho_{max3}$	2.9395
4.	$\rho_{max2}$ and $\rho_{max4}$	2.7003

**Remark**

- $\rho_{max1}$  .Reliability coefficient of rubric instrument with 3 options and replications of 20 times.
- $\rho_{max2}$  .Reliability coefficient of rubric instrument with 4 options and replications of 20 times.
- $\rho_{max3}$  .Reliability coefficient of rubric instrument with 3 options and replications of 30 times.
- $\rho_{max4}$  .Reliability coefficient of rubric instrument with 4 options and replications of 30 times



**Figure 1. Boxplot of Maximum Reliability**

## 4. Discussion

The result of this study indicates that there are differences between the instrument of reliability coefficient of the rubric which has different scale options namely, 3 and 4 option scales. Differences in the amount of replication or duplication of data also affects the reliability coefficient, namely at the replication of 20 and 30 times. Differences in the amount of scale and replication options give the distinction of reliability.

Based on the analysis that has been done, the followings can be presented: the number of scale options on the rubric instrument and the effect of number of replication on reliability. With a number of different options and a lot of different replication generally affects the variance (S<sup>2</sup>) which will affect the reliability. Rubric instrument with 3 options and 20 replications has variance 0.0000586, while the rubric instrument with 4 options and 20 replications has variance of 0.0000666. Rubric instrument with 3 options and 30 replications has variance of 0.0000593, while rubric instrument with 4 options and 30 replications has variance of 0.0000780. Based on these data we can conclude that the general trend is that the higher the variance, the higher the reliability will become. While the mean of reliability coefficient of 3 options with 20 replications is 0.8907, the mean of reliability coefficient of 4 options with 20 replication is 0.8952, the mean of reliability coefficient 3 options and 30 replication 0.8972 while mean of reliability coefficient of 4 options with 30 replications is 0.9011.

Hence, with greater scale options, the number of variance will be higher, with the higher variance, the reliability coefficient will increase. And with more replications, the variance will be higher, with higher variance, the reliability coefficient will increase. The selection of response category in this case is the number of the rubric's scale options namely 3 and 4 scale options which affects the reliability coefficient. The results showed that the reliability of the instrument with four scale options is higher than the 3 options. This is in line with research conducted by Solihin which stated that the reliability of the instrument with six options is higher than the 5 options, this means that more item options will result in the higher reliability coefficient of the instrument. As noted previously by Guilford, increased reliability results from the to increased number of answer options' [12]

The study also found that the maximum reliability coefficient of rubric instruments showed that the greater replication will increase reliability. This strengthens the research conducted by Bambang Suprihatin et al. [13] stating that replication will actually strengthen the degree of reliability (reliability) of a study. That is a regularity (reliability) is high when replication is performed measurements will provide the measurement results are relatively the same. The frequency of replication will generate lower standard error score. The higher the value of replication bootstrap will result in lower standard error for such measurements.

The results showed that 30 replications of reliability coefficient has an average which is higher than 20 replications. This can be due to the standard error on the instrument 3 to 30 replication options are 0.0014 lower

than the 20 replication is 0.0017. Likewise on the instrument 4 options, the standard error at 30 replication lower at 0.0016 compared to 20 replication is 0.0018.

With more frequent replications, the standard error will decrease and reliability will increase. With more replications, standard error will decrease, with the less standard error, the reliability will increase. It is in accordance with what is researched by Breaux and Frey which found that the lower the reliability of the measurement, error will increase. [14]

This study also found that there is significant influence in the use of selection in the midpoint. It can be seen from the difference in mean reliability by using the midpoint of the 3 options by not using the midpoint of the 4 options. Mean reliability coefficient 3 with 20 replication option is 0.8907, the mean coefficient of reliability 4 options with 20 replication is 0.8952, mean reliability coefficient 3 with 30 replication options mean reliability coefficient was 0.8972 4 options with 30 replication is 0.9011. Based on these data it can be concluded there is a significant increase in reliability coefficient on instruments that do not use the midpoint option, is in line with what is disclosed by Rungson in his findings which stated that the option having a midpoint option will affect the variance, where the variance obtained is smaller, since rubric assessor will likely have more midpoint without having to read or think about an instrument that is completed, and thus the reliability decreases.[15] This study therefore agree with the findings of Rungson.

## 5. Conclusion

The results showed that the average maximum reliability coefficient will be higher with more scale options on the rubric instrument. The amount of data replication also affects the maximum reliability coefficient. The more frequent the replication, the maximum reliability coefficient will also get higher on the rubric instrument.

## References

- [1] Berg, G. A. (2002). *Why distance learning? Higher education administrative practices*. 88 Post Road West, Westport: Praeger Publisher.
- [2] Stevens, Danelle D. and Antoni J. Levi. (2005). *Introduction to Rubric*. Virginia : Stylus Publishing.
- [3] Ebel, Robert and David Frisbie. (1991). *Essential of Educational Measurement*. Engelwood Cliffs, New Jersey : Prentice-Hall Inc.
- [4] Anastasi, Anne and Susan Urbina. (1997). *Tes Psikologi, terjemahan Robertus H. Imam*. Jakarta: PT Prenhallindo.
- [5] Wiersma, William and Stephen G. Jurs, *Educational Measurement and Testing*. Boston: Allyn and Bacon, 1990.
- [6] Mehrens, William A. and Irvin J. Lehmann. (1991). *Measurement and evaluation in education and psychology*. Belmont CA: Wadsworth Thomson Learning.
- [7] Azwar, Saifudin. *Reliabilitas dan Validitas*. (2002). Yogyakarta: Pustaka Pelajar.
- [8] Li, H., Rosenthal, R and Rubin, D.B (1996). *Reliability of Measurement in Psychology: From Spearman-Brown to Maximal Reliability*. *Journal of Psychological methods*. Vol. 1, No. 98-107.
- [9] Kamata, A., A. Turhan., E. Darandari. *Estimating Reliability for Multidimensional Composite Scale Scores*. Paper. Presented at the annual meeting of American Educational Research Association, Chicago, April 2003.

- [10] Jonsson, Anders and Gunilla Svingby. (2007). "The use of scoring rubrics: Reliability, validity and educational consequences." *Educational Research Review* 2. 130-144.
- [11] Thompson, M. K., Clemmensen, L. K. H., & Ahn, B. (2013). "The Effect of Rubric Rating Scale on the Evaluation of Engineering Design Projects." *International Journal of Engineering Education*, 29(6), 1490-1502.
- [12] Mueller, Daniel J. (1986). *Measuring Social Attitudes*. New York: Teachers College.
- [13] Bambang, Suprihatin. *et al.* "Estimasi Parameter Bootstrap Pada Proses AR(1)." <http://core.ac.uk/download/pdf/11732680.pdf> (diakses 20 Desember 2015).
- [14] Breaux, Kristina C dan Frances E. Frey, MA. "Reliability Coefficients and Standard Errors of Measurement for Subtest Component Scores." [http:// images. pearsonclinical. Com .../ WIAT3\\_TechRep](http://images.pearsonclinical.com/.../WIAT3_TechRep)", (diakses 20 Desember 2015).
- [15] Chomeya, Rungson. (2010). Quality of Psychology Test Between Likert Scale 5 and 6 Points, *Journal of Socila Sciences*, 6(3), 399-403.