International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 12, Issue 8, 2020

Examination of the Validity of Instruments used for Students in Vocational High Schools

Soeprijanto[a*], Gaguk Margono[b], Rivo Panji Yudha[c], [a,b]Universitas Negeri Jakarta, Jakarta, Indonesiaa, [c]Universitas 17 Agustus 1945 Cirebon, West Java, Indonesia, Email : [a*]soeprijanto@unj.ac.id

This study aims to examine the validity and reliability of the instrument used as a result of learning. Instrument learning outcomes of vocational high school student engineering practices consist of 5 indicators and 20 descriptors, each descriptor there are 5 observation items (checklist). Standardization of the items of observation was carried out through the Expert Panel and Field Test (empirical test), through a Factor Analysis Test sample of 298 vocational students. The results found that the inter-panel panel showed consistency with rkk value = 0.87 ≥ 0.5. The conclusions obtained through the study are according to use as a scale rubric on the results of learning the technical practices of vocational students. The results of the empirical tests showed that the value of the goodness of the test produced an index of 208,823 with a degree of freedom of significance and 0,000. The contributions of this study include: (1) providing an alternative way to test the quality of student performance assessment instruments in learning in Engineering Field Practice; (2) adding a more objective form of rubric in assessing students using observation sheets.

Key words: *Items observations, validity, and reliability, learning results of practices.*

Introduction

Vocational education includes skills-based education and there are levels of Vocational High Schools and a Diploma program. According to Presidential Regulation No. 8 of 2012 concerning the Indonesian National Qualification Framework, Vocational High Schools are prepared to produce graduates with level 1 qualifications and able to work as operators and or technicians (President of the Republic of Indonesia, 2012). In order to achieve competence with these qualifications, a vocational education institution requires a good curriculum, complete learning media, sophisticated practice equipment, and a reliable evaluation system (Torrance, 2012). The vocational education evaluation system at the vocational level is carried out through competency skills testing at the end of the education period. In the implementation of competency tests, a system and a reliable appraisal tool are needed (Wahba, 2010).

Traditionally, vocational educators rely on performance-based assessment strategies to assess student mastery of specific job-specific skills (Hodgman, 2014). There are two types of tests seen from the material tested on competency tests: vocational theory tests and vocational practice exams. A Vocational Theory Test is carried out in the form of written tests that can be either objective or subjective tests. Whereas for practice exams conducted in the form of skills tests or performance tests (Vleuten & Sluijsmans, 2017), the successful implementation of vocational practice exams is determined by instrument quality, the measurement method used in the test of vocational practic, and the readiness of the person carrying out the practice exam (UNESCO, 2012). Instruments that are widely used for skills tests and performance tests are in the form of observation sheets such as a checklist or scale (Learning and Teaching Center, 2010)

Important things in the development or use of instruments for measuring learning outcomes are that they be valid and reliable instruments and likewise to measure the results of practical learning in Vocational Schools. As well as instruments in general, indicators for quality key practical learning outcome measurement tools also require validity and reliability. Reliability is an estimate of the reliability of instrument stability measurements including internal consistency, and inter-related reliability of instrument scores (Kimberlin & Winterstein, 2008). Validity and reliability are two important factors to consider when developing and testing any instrument. Validity refers to the extent to which an instrument accurately measures what it wants to measure (Mohajan & Mohajan, 2017)

Testing Validity and reliability of instruments for measuring learning outcomes was achieved by testing for the validity and reliability of the Testing Test Instrument in the form of multiple choice test options. There are only a few examples of the validity and reliability tests and for the checklist observation sheet.
The background for this research is its' aim to examine the validity and reliability of the instruments of learning outcomes of engineering practices. The research problem was formulated as follows: How to test the validity and reliability of the instrument of the results of practical learning for junior high school students.

Given the extent of fields in vocational education, it is necessary to limit the fields developed and to get the accuracy of the problem the formulation of the problem is described as a more detailed sub-problem. In this study, the instruments developed were limited to basic engineering practices. The problem formulation is translated into three, namely:

1. What is the level of suitability of indicators and descriptors of developed practical skill instruments?
2. How is the validity of the practice skill instrument developed?
3. How is the reliability of the practice skill instrument developed?

Literature Review

Test Instrument Learning Result of Engineering Practical Teaching

Learning practice or practicum is the process of training student skills in certain fields of vocational education. According to (Munawar, Berman, 2014: 308) Practical Learning is an educational process that serves to guide students systematically and directed so that they can perform a skill. According to Leighbody and Kidd (1966: 21). The process of implementing practical teaching includes the opening, presentation, application closure and evaluation. Learning outcomes are what students will know and be able to do at the end of the course or at the end of the learning program, learning outcomes refer to knowledge of attitude and attitude (Shirley Lesch, 2012: 1-2). Learning outcomes of engineering practices include students' knowledge of the technical objects practiced, practical work assignment skills given by the teacher and the work attitudes shown by students during the practical tasks. In school activities, many results of student practices show the results of the teacher's assessment of student performance during the engineering practice assignment (Bonney, 2015).

The instrument model commonly used to measure students' practical abilities is the rating scale and observation sheet. (Wood, 2010) When the observed behavior is a process and the work of a student during a pratic lesson then the observation is a form of the performance test. As Leighbody stated, (1968) that the observation process when students practice the skills they learn is a form of the performance test. Learning assessment engineering practices can be focused on processes and outcomes (Jones, 2005), and form of instruments in the form of peformanace skill observation sheets (Milanowski, Prince, & Koppich, 2007).

Validity and Reliability Instrument

Validity is an instrument quality that shows the extent to which an instrument is able to measure what should be measured. Very rarely, it was found that the instrument was one hundred percent valid, therefore to find out the quality of the instrument required to test the validity and reliability. (Nor Hasnida, 2016). The term validity turns out to have a diversity of categories. Robert Ebel (1991) divides validity into concurrent validity, construct validity, face validity, factorial validity, empirical validity, intrinsic validity, predictive validity, content validity, and curricular validity.

Relief can be defined as the proportion of test score variance due to true score differences, (Cecil R. Reynolds, 2009: 96). There are four estimates commonly used to determine instrument reliability, namely (1) inter-related / observer reliability; (2) Test-Retest Reliability; (3) Alternate-Forms Reliability; and (4) Internal consistency Reliability. Estimating inter-related reliability is a fairly straightforward process. If the scoring of an assessment relies on subjective judgment, it is important to evaluate the degree of agreement when different individuals score the test, this is referred to as inter reliability, (Cecil R. 2009). Test-retest reliability is nothing but a degree that shows the consistency of the results of a test from time to time. Test-retest shows the variance score obtained from the implementation of an evaluation test conducted twice or more, as a result of measurement errors. Test-retest reliability is very important, especially when used to determine the ability test predictor. The ability test will not be useful if it shows results that always change significantly when given to respondents.

Internal Reliability Consistency is one type of reliability based on constancy in each test item (Nurosis 2003). In a Classical test theory, the subject's response to a particular item is the sum of the correct score and the error of Internal Reliability-Consistency mainly reflects errors related to sampling content. This estimate is based on the relationship between items between a test and is derived from a single test set. Internal consistency reliability consists of split reliability, Alpha reliability and Kuder Richardson's reliability (Cecil R. 2009). Alpha Cronbach can be interpreted in several ways. This is a correlation between scale and all tests or other possible scales of a hypothetical (Norusis 2003).

Some important terminology is used to see the quality of the instrument, among others correlation coefficient, validity coefficient and. Reliability Coefficient. Correlation coefficients are a statistic that indicates the degree of association between any two sets of measures obtained from the same group of individuals. The validity coefficient is a correlation coefficient that indicates the degree to which we predict predictions or estimates performance on some measures measure. Norman E, Gronlund, 1985: 88).

Research Methods

In general, the purpose of testing instrument validity and reliability is to determine a quality measuring instrument that is reliable. The scale used in this instrument is a branch scale with the optional option 1 s.d. 5. Options are given based on the number of descriptors or statement items that appear at the time of observation. When 1 descriptor appears, it means option 1, so if 5 descriptors appear, the choice is on obsi 5. Casting is done by marking (checklist) the descriptor of each indicator of the instrument item than in the number. The highest value for each indicator of the test item amounting to 5 and the lowest value 1. For indicators that are not filled due to the nature or characteristics of practice that does not use the indicator is given a score or score of 0.

Testing instrument validity is carried out in two stages, the first stage is through the panel. The second stage has empirically tested the validity and reliability of the instruments that have been developed. Validation process first tested theoretical construct suitability. The points assessed by the panellists are: (1) Suitability of dimensions with indicators developed and (2) suitability of statement of items with indicators. The steps to test the validity and reliability are as follows.

Dimension Suitability Test and Indicator

Panellists assess suitability every indicator use a scale of 1 to 9. The level of validity of each indicator (item) is determined based on the Median value obtained from the ten panellists. The criteria used to make decisions based on the level of validity of indicators can be seen in Table 1 below.

Table 1: Indicator Assessment Criteria (item)

Score	Assessment
1 s.d.3	Replaced
4 s.d. 6	Repaired
7 s.d 9	Valid

Test to determine the inter-related reliability (rater) used Ebel's formula as found in Guilford (1954: 135-196). Results of calculation of reliability level were used to determine the level of agreement of the ten panellists. Acceptance level limit is rkk $\geq$ 0.50.

Suitability Test Descriptor (checklist item) With Indicator

Assessment of item statements using a scale of 1 to 5. Test the validity of the descriptor each indicator is performed using the Median value from the assessment of the ten assessors. Interpretation of the median value is done by referring to the assessment criteria for the stated test item statement, namely: aborted if the median score is ten assessors 1, corrected language if the Median score of the ten assessors is 2 s.d 3 and is immediately used (valid) if the median score of the ten assessors is 4 s.d. 5.

Test with Factor Analysis

The next empirical construct validity test is performed by factor analysis. Through factor analysis, it is expected to find dimensions, indicators, and solid points that form the construct of the variables being tested. In addition, through this factor analysis is expected to be found a set of new variables that are fewer in number than the previous variables. According Supranto (2004: 114-115) factor analysis aims to: (1) recognize or identify the underlying dimensions or factors that explain the correlation between indicators, (2) recognize or identify a new set of less correlated indicators to replace a correlated set of original variables, (3) recognize or identify an important set of indicators from a larger number of indicators.

Factor analysis is also used to test whether the tested variables indicate that among these indicator indicators are fellowship factor. According to Gable (1986: 39), the number of samples or respondents for factor analysis ranges between 6 to 10 times the number of items tested. This instrument has 20 indicators, this means a minimum sample size of 100 people. The number of samples in this study was 220 people students.

The calculation of statistics needed to analyze data is done with the help of the SPSS program. The approach used in this factor analysis is the Principal Component Analysis (CPA) method. As a test, it is also tested by using the Extraction method: Maximum Likelihood obtained the following results: The number of factors is determined by the extraction method and orthogonal/varimax rotation method.

The factor that can be used as an indicator has an eigenvalue greater than or equal to 1.0. Tests for analysis requirements were conducted with Kaiser Mayer Olkin regarding the Measure of Sampling Adequacy (KMO-MSA). According to Supranto KMO-MSA measures the adequacy of sampling (sampling adequacy), which is the index that compares the magnitude Correlation coefficient is observed with the magnitude of the partial correlation coefficient. The small KMO MSA value indicates that the correlation between pairs of variables cannot be explained by other variables and the factor analysis is not appropriate. The KMO size is determined based on the opinion of Kaiser (974) in Neurosis (2003: 400) which states that KMO MSA ≥ 0.90 is very good; 0, 90 > MSA ≥ 0.80 good; 0.8 > MSA ≥ 0.7 medium; 0.7 > MSA ≥ 0.5 enough; 0.50 > MSA is bad or unacceptable. So, the level of acceptance used in this study is the Measure of Sampling adequacy (MSA) 0,5 0.50.

Testing metrics that are formed, whether a unit metrics or identity metrics as defined by Barlett's Test of sphericity (X2) was to test whether the matrix formed is a unit matrix or identity matrix. Hypothesis Ho: p = Ivxv versus H1: p ≠ Ivxv with formula $X^2 = \{1/6\ (2v + 5) - (n-1)\}$ Ln [Mvv]. Where v = number of variables or items, n = number of samples and

[Mvv] is reflected in the correlation matrix with degrees of freedom dk = ½ v (v-1) (Suryanto1988).

Factor analysis carried out with the Principal Analysis Component (PCA) as an exploratory approach was extracted from the correlation matrix. the criteria used to analyze it are as follows: (1) Communities as the main variant, (2) eigenvalues value with the characteristic equation | Mvv-ƛI | = 0; with ƛ > 1 factor in according to Kaiser's rules - Guttman, (3) validity level seen from the size of the factor on each factor and Communality, the following: 0.7 special, 0.63 very good, 0.55 good, 0.45 enough and 0.32 moderate (Kaisar1974 in Norusis 2003). The acceptance limit used to test item validity is 0.32.

Research Results

All panellists agreed that dimensions and indicators and the variables are appropriate and the instrument is ready to be tested for further constructs (empirically). Of the 20 indicators, each indicator developed 5 statement items so the number of statements totalled 100 items. The Rater Test results between assessors are intended with the following anava calculation results as depicted in Table 2 below:

Table 2: Results of Anava

	ANAVA		
Sum. Var.	JK	db	Varian (F)
Item	43,03	19	2,26
Expert	2,58	9	
Rest	50,27	171	0,29
Total	95,88		

$$rkk = \frac{Vb - Ve}{Vb} = \frac{2{,}26 - 0{,}29}{2{,}26} = 0{,}87$$

Rater value (inter-rater reliability) is 0.87 so it can be concluded that overall the panellists stated that the statement of the items for the indicators was appropriate.

Test Results

Factor Analysis Approaches used in this factor analysis is the Principal Component Analysis (CPA) method obtained by the KMO MSA value of the results of practical training at 0.839 so that it is in a good category (0, 90 > MSA ≥ 0.80). This value is greater than 0.50 means factor analysis is appropriate for analyzing data in the form of correlation metrics. Besides that Bartlett's value Test of Sphericity X2 = 1856,929, on the degree of freedom (df) = 190 with a significance level of (p) 0,000 < 0,05, then the hypothesis that variables are not correlated with each other (that population correlation metrics are identity metrics with a value of 1 on the diagonal and zero outside the diagonal) is rejected, and means that these variables are indeed correlated, so that factor analysis can be continued.

Table 3: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.839
Bartlett's Test of Sphericity	Approx. Chi-Square	1856.929
	Df	190
	Sig.	.000

Extraction using PCA has communality prices for four enough factor, i.e. the lowest is 0.461 contained in item x.2.5 and the highest variant of 0.823 in item x.5.3.

Table 4: Distribution of Grains before Rotating

Factor	**Item Distribution**	**Name of Factor**
Factor 1	x1.1, x1.2, x1.3, x1.4	work procedures
Factor 2	x2.1, x2.2, x2.3, x2.4, x2.5	skills
Factor 3	x3.1, x3.2, x3.3, x3.4, x3.5.	attitudes
Factor 4	x4.1, x4.2,x4.3	knowledge
Factor 5	x5.1,x5.2,x5.3	product quality

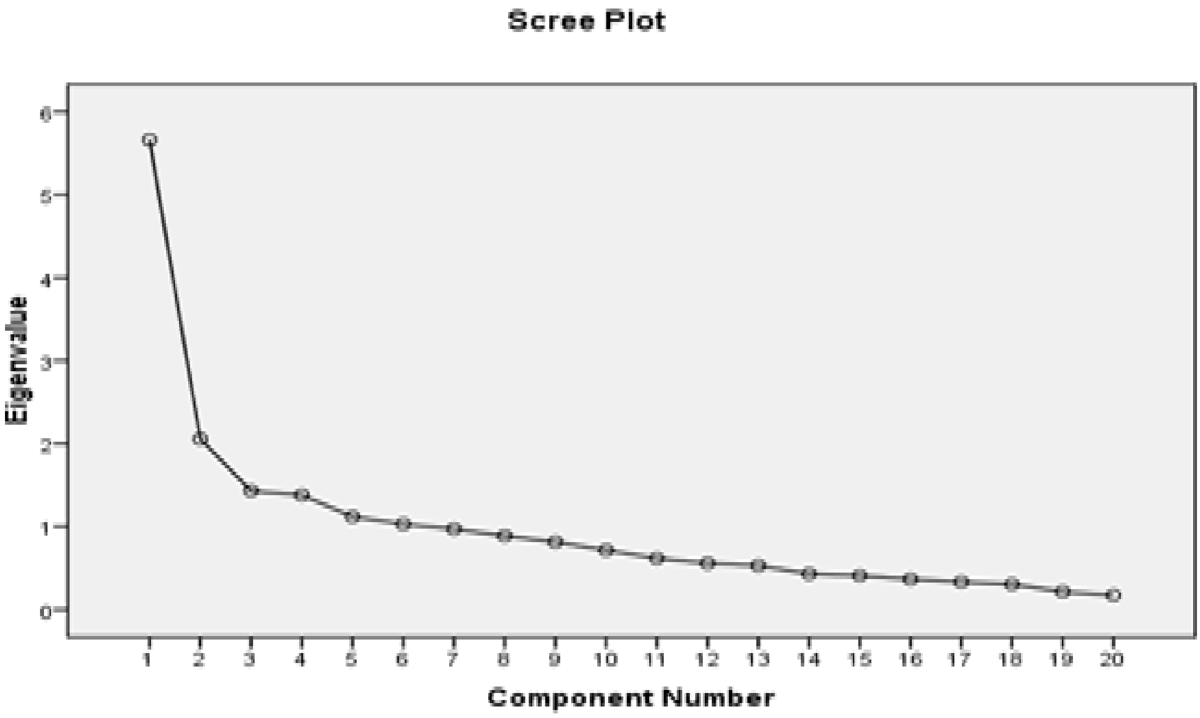
In the confirmatory approach carried out through computation with the maximum probability method (ML = Maximum Likelihood) to test whether the estimated 5 factors that have been revealed are normally distributed. Total variance explained, from 20 items included in the factor analysis and obtained the characteristic root value (eigenvalues) above 1 (≥ 1) as many as 5 factors, the same as those estimated. Shifting factors that occur as follows.

Table 5: Distribution of items after rotated

Factor	**Item Distribution**	**Name of Factor**
Factor 1	x4.2, x4.3, x5.1, x5.3	product quality,
Factor 2	x.1.3, x2.4, x3.2, x3.3, x3.4, x3.5	work procedures
Factor 3	x1.1, x1.2, x1.4, x2.5, x3.1	attitudes
Factor 4	x2.1, x2.2	skills
Factor 5	x4.1, x2.3	knowledge

In this case to test the suitability of the goodness of fit test produces an index of 208,823 with 100 degrees of freedom and 0,000 significance. The results of the scree plot diagram are shown in Figure 1 below.

Figure 1. Scree Plot



Instrument Reliability Calculations

Reliability analysis is performed on each factor for each instrument. Reliability is calculated with Alpha (α) from Cronbach. The overall reliability of learning planning instrument factors is 0.782 while for each factor the following reliability levels are obtained (Table 6). Table 6 below shows that all factors have a high reliability (α) coefficient so that it can be concluded that the instrument has been reliable.

Table 6: Reliability Calculation Results

Factor	**Name Factor**	**Coef. Reliability (α)**	**Total (α)**
Factor 1	Product quality	0,808	0,782
Factor 2	Work procedures	0,736	
Factor 3	Work attitudes	0,737	
Factor 4	skill	0,708	
Factor 5	knowledge	-0,320	

For these results, the items in factor 5 namely x2.3 and x5.2 are declared unreliable.

Discussion

After two stages of testing the validity of the instrument, we got a picture of quality the instrument being tested. The first trial through the panel of experts ((Expert Judgment) resulted in a grid of questions that had a match between indicators and descriptors. The Rater Test results rkk = 0.878 pointed out that the ten panellists had strong consequences. The study was conducted by (Ahmad, Mukharil, Informatika, Teknik, & Indonesia, 2016) the inter-related reliability coefficient (Kappa) is 0.605., The Kappa coefficient can be interpreted in the good category because the Kappa $\kappa > 0.60$ coefficients and conclude that there is an expert agreement, and can be said to be valid and reliable. Testing by using the Inter-rater Reliability Coefficient by (Ahmad et al., 2016) with the results of average measure 0.824 concluded that the overall instrument has high stability. Validity test through the panel can be continued because the items are developed from the indicators that are, where each indicator is made into 5-scale assessment points, criteria are needed to fill it. Each indicator has 5 descriptors that are used as statement items or items as criteria. Statements of items that have been declared valid according to the Panellists (the expert) are used as checklist items, using descriptive writing.

The results of the instrument testing were empirical through the Principal Component Analysis (CPA) method obtained by the KMO MSA value of the results of practical training at 0.839. According to (Kayisoglu, 2015). Leech et al. (2005); Sencan (2005) and Tavsancil (2001) state that the KMO criticism value is 0.50, and factor analysis cannot be done below this value. Compared to the critical value, the Kaiser-Meyer-Olkin value obtained was 0.839 so that it was in a good category ($0, 90 > MSA \geq 0.80$). These values indicate that data obtained from convenient pre-implementation for factor analysis. And the sample quantity is sufficient for factor analysis.

The results of the instrument testing were empirical through Factor analysis which shows that each item has a significant charge factor so that there are no factors that fall. There are two items which have a large enough charge on two or more factors. For items such as this, improvements are made to the item statement and not aborted.

Rotation results show no change in a number of factors, but there is a shift in grouping items so that a new naming is needed for the factors formed. Naming is given based on clustered items. For example Factor 1 is given a new name "the results of the practice" because clustered items include: (1) understanding concepts that are relevant to practical material, (2) understanding the benefits of practice objects, (3) the appearance of practical results, and (4) the strength and durability of the results of the practice. Items (1) and (2) are the results of the practice in the form of knowledge while items (3) and (4) are the results of the practice or practical products in the form of objects, with the right name for the new factor are "results of practice". In the same way, the name of another new actor is given.

The instrument reliability calculation is done by calculating the Alpha coefficient indicating that each factor formed has a value greater than 0.5 ($\alpha > 0.5$), except for factor 5 where the reliability value ($\alpha = - 0.320$). Thus a factor of 1 to 4 is feasible to use while factor 5 is aborted or repaired first if it will be used. Factor 5 is formed from two items, namely: (1) the skills of students choosing and using practical materials, and (2) functioning whether or not the results are practiced, have $\alpha \leq 0.5$ (means unreliable). It is recommended not to be aborted because overall the instruments are developed reliable with $\alpha = 0{,}782 \geq 0.5$.

Conclusion

Based on the results of research and discussion about practical learning outcomes assessment instruments can be summarized as follows:

1. The test results through a panel of experts (expert judgment), states that there is a match between variables with dimensions, dimensions with indicators, and indicators with descriptors (statement of items).
2. Based on Factor analysis shows that the results of the development have a contract the valid to measure performance students Practice in Engineering because of (1). All variables have KMO MSA ≥ values; 0.50 so it's right factor analysis, (2). All items have a factor charge above 0.32 so that there are no items that fall; (3). Varimax rotation results show a reduction in the number of factors
3. Reliability Instruments have index $\alpha \geq 0.50$, the instrument is the same general factor and is a reliable instrument.
4. Validity and reliability test proves that indicators and descriptors have strong correlations or constructs, so descriptor can be an item in the checklist item.

The standardization of the instrument needs to be effected by continuously refining the statement of the item in accordance with the existing developments and this must be continued by testing the validity and reliability of the instrument.

REFERENCES

Ahmad, N., Mukharil, A., Informatika, T., Teknik, F., & Indonesia, U. K. (2016). *Jurnal Sistem Informasi (Journal of Information Systems*). 2 / 12 (2016), 82-89 DOI : http://dx.doi.org/10.21609/jsi.v12i2.481, *12*(3), 82–89.

Azwar Saifudin,(1986) *Reliabilitas dan Validitas Interprestasi dan Komputasi*, Yogyakarta: Penerbit Liberty.

Biggs, J (1996) Assessing learning quality: Reconciling institutional, staff and educational demands, *Assessment & Evaluation in Higher Education*; Mar1996, Vol. 21 Issue 1, p5, 11p

Bonney, K. M. (2015). Case Study Teaching Method Improves Student Performance and Perceptions of Learning Gains †. *Journal of Microbiology & Biology Education*, *16*(1), 21–28. https://doi.org/10.1128/ jmbe.v16i1.846

Carole.L. Camberline and Almut G. Wensterstain (2008): Validity and reliability of measurement instruments used in research , *Article, American Journal of Health-System Pharmacy* Vol. 65, Issue 23 Des, 2008

Ebel R.L., & Frisbie A. David.(1991): *Essentials of Educational Measurement,* New Jersy : Printice Hall Englewood Cliffs.

Gronlund, Norman E, (1985): *Measurement and Evaluation in Teaching, 5Th Editions*.Macmilan Publishing Company : New York.

Hodgman, M. R. (2014). Using Authentic Assessments to Better Facilitate Teaching and Learning: The Case for Student Portfolios. *Journal of Studies in Education*, *4*(3), 59. https://doi.org/10.5296/jse.v4i3.6149

Jones, C. (2005). Assessment for Learning. *Learning and Skills Development Agency*, 16–19.

Kayisoglu, N. B. (2015). Validity and reliability studies for scale of evaluating physical education teachers based on student ratings (SEPETBSR). *Journal of Physical Education and Sport Management* , *6*(8), 60–69. https://doi.org/10.5897/

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*(23), 2276–2284. https://doi.org/10.2146/ajhp070364

Leighbody, Gerald B. & Kidd Donald M., (1968): *Methods of Teaching Shop and Technical Subyects*. New York: Delmark Publisher.

Liesbeth K.J. et.al (2006), Teachers' opinions on quality criteria for Competency Assessment Programs: Teaching and Teacher Education

Learning and Teaching Centre. (2010). Developing Checklists. *British Columbia Institute of Technology*.

Milanowski, A. T., Prince, C. D., & Koppich, J. (2007). *Observations of Teachers ' Classroom Performance Guide to Implementation : Resources for Applied Practice Observations of Teachers ' Classroom Performance*. Retrieved from https://www.wested.org/wp-content/uploads/classroom-practice. pdf%5Cnhttp://files.eric.ed.gov/ fulltext/ED533045.pdf

Mohajan, H., & Mohajan, H. K. (2017). M P RA Two Criteria for Good Measurements in Research: Validity and Reliability Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University*, *17*(3), 58–82. https://doi.org/10.26458/1746

Munawar, W., Berman, E. T., Pendidikan, D., Mesin, T., & Indonesia, U. P. (2014). *Ketercapaian hasil belajar siswa smk pada praktik pemeliharaan ac split*, *1*(2), 307–312.

Nor Hasnida, C. M. G. (2016). A Reliability and Validity of an Instrument to Evaluate the School-Based Assessment System : A Pilot Study. *International Journal of Evaluation and Research in Education (IJERE)*, *5*(2), 148–157. https://doi.org/ http://dx.doi.org/10.11591/ijere.v5i2.4533

Nurosis, J. Marija,(2003) *Spss.12.0 Statistics Procedures Compnion*,Published by Printece Hill. Inc. New Jersey

Presiden Republik Indonesia. (2012). Peraturan Presiden Republik Indonesia Nomor 8 Tahun 2012 Tentang Kerangka Kualifikasi Nasional Indonesia.

Shirley Lesch. (2012). *Learning Outcomes and samples*.

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, *38*(3), 323–342. https://doi.org/ 10.1080/ 03054985.2012.689693

UNESCO. (2012). *Proposed Indicators for Assessing Technical and Vocational Education and Training. UNESCO Working Document*, (April), 1–50. Retrieved from http://www.etf.europ/Report on indicators April 2012.pdf

Vleuten, C. Van Der, & Sluijsmans, D. (2017). Competence-based Vocational and Professional Education. *Mulder, M. (Ed.). (2017). Competence-Based Vocational and Professional Education (Vol. 23). Cham: Springer International Publishing. Https://Doi.Org/ 10.1007/978-3-319-41713-4*, *23*, 607–630. https:// doi.org/10.1007/978-3-319-41713-4

Wahba, E. M. (2010). Competence Standards for Technical and Vocational Education and Training TVET. *Www.Unevoc.Unesco.Org/e-Forum/Competence StandardsforTVET.Pdf*, 1–23.

Wood, B. (2010). Defining *"Needs Assessment", "Learning Gaps" and the "Outcomes Measurement" Process*. *PediaLink*, 4–25. Retrieved from https://pedialink.aap.org/File Library/About AAP CME/Gaps-and-Needs-Assessment-Information.pdf