IJER
Indonesian Journal of Educational Review

MULTIDIMENSIONAL RELIABILITY OF INSTRUMENT STUDENTS' SATISFACTION USING CONFIRMATORY FACTOR ANALYSIS

Gaguk Margono
Graduate School, State University of Jakarta
g_margono@yahoo.com

ABSTRACT

The purpose of this paper is to compare unidimensional reliability and multidimensional reliability of instrument students' satisfaction as an internal costumer. Multidimensional reliability measurement is rarely used in the field of research. Multidimensional reliability is estimated by using Confirmatory Factor Analysis (CFA) on the Structural Equation Model (SEM). Measurements and calculations are described in this article using instrument students' satisfaction as an internal costumer. Survey method used in this study and sampling used simple random sampling. This instrument has been tried out to 173 students. The result is concluded that the measuringinstrument of students' satisfaction as an internal costumer by using multidimensional reliability coefficient has higher accuracy when compared with a unidimensional reliability coefficient. Expected in advanced research used another formula multidimensional reliability, including when using SEM.

***Keywords**: multidimensional reliability, instrumentof students' satisfaction as an internal costumer, confirmatory factor analysis.*

According to the Latan (2012) Structural Equation Modeling (SEM) is a second-generation multivariate analysis technique that combines factor analysis and path analysis that allows researchers to simultaneously test and estimate the relationship between exogenous and endogenous multiple variables with many indicators. Joreskog research results in the 1970s brought on the statistical theory of linear structural analysis that is better known as structural equation modeling or SEM. Important source is used in analyzing the covariance structure so that this approach is sometimes called the covariant structure model (CSM).

The model includes variables drawn immeasurable called the latent constructs were constructed by a set of measurable variables, called the construct measured. Measurement error of measurement reflecting the reliability scores are seen as a unique construct and become an important part of SEM analysis, measurement errors were included in the SEM analysis is then be compared with the advantages of SEM other analytical techniques (Capraro et al., 2001). SEM can estimate the error variance in the actual measurement outcome scores estimating the reliability.

According to Geffen and colleagues (2001), SEM as multivariate statistical technique that combines multiple regressions to identify relationships between constructs and factor analyzes that identify the concept was measured by several indicators manifest that both are used simultaneously.The first approach is the correlation correction attenuation caused by measurement error and the second approach is a structural equation model in the context of confirmatory factor analysis. Lee and Song (2001) said that SEM is one approach to confirm the measurement model. In the SEM measurement model linking the latent constructs with empirical construct. Empirical constructs expressed by a combination of latent constructs. In addition be able to and capable of handling generalizability theory and

item response theory, SEM is able to compare the measurement model and the accuracy of the model facilitates investigation.

SEM has two basic components. First, the measurement model is defined as the relationship between latent variables and a group of explanatory variables that can be measured directly. Second, the structural model is defined as the relationship between latent variables that cannot be measured directly. These variables also distinguished as independent variables and the dependent variable. Geffen and colleagues (2001) said that the measurement model is sub-models in the SEM with latent constructs that identify the indicators that can be used to determine reliability each construct were included in the model. SEM can also identify reliability constructs are visible through the resulting value of the grain loading. Based on the perspective SEM construct reliability can be calculated through the following equation:

$$CR = \frac{\left(\sum_{i=1}^{i} \lambda_i\right)^2}{\left(\sum_{i=1}^{i} \lambda_i\right)^2 + \left(\sum_{i=1}^{i} \delta\right)}$$

Description:

CR =Construct reliability

λ_i =Factor loadingof standardizedindicatorsto-i

δ =Standard error of measurement

McDonald (1981) formulate are liability coefficient which later was named a McDonald composite score reliability coefficients are also called omega (ω). Reliability coefficient is based on confirmatory factor analysis that is part of the menu SEM modeling. This composite score reliability McDonald explains the large proportion of indicators in measuring the construct explained. The formula to obtain the constructreliability coefficients are as follows:

$$\omega = \frac{\left(\sum_{i=1}^{i} \lambda_i\right)^2}{\left(\sum_{i=1}^{i} \lambda_i\right)^2 + \left(\sum_{i=1}^{i} 1 - \lambda_i^2\right)}$$

Description:

λ_i = Factor loading of standardized indicators to-i

When compared between the reliability of the constructs with a composite score reliability McDonald will give the same result as $\delta = 1 - \lambda^2$.

The following is a reliability coefficient multidimensional construct reliability coefficients developed by Hancock and Mueller (2000) that shows how well the indicator could reflect the construct to be measured. This coefficient is a

modification of the McDonald construct reliability coefficient was not able to accommodate different weights interdimensions. Modified construct called reliability coefficients weighted as follows:

$$\Omega_w = \frac{\sum_{i=1}^{p} \frac{l_i^2}{(1-l_i^2)}}{1+\sum_{i=1}^{p} \frac{l_i^2}{(1-l_i^2)}}$$

Description:

l_i = Coefficient of the i-th standardized dimensions

The reliability coefficient can be interpreted as the square of the correlation between the dimensions of the optimal linear composites, so some experts call it the maximum reliability.

In measuring student satisfaction used a measure of subjective or soft measures as quality indicators or quality. This size is called soft, because these measures focus on the perceptions and attitudes rather than concrete things called objective criteria. Therefore, focusing on the perceptions and attitudes of the gauges used to be a student satisfaction questionnaire which can be measured by the quality of service or the quality of higher education institutions.

The quality is a term that is constantly moving dynamic; moving forward if the quality is said to be better, otherwise if it moves backward to say quality deteriorated. Quality means can mean superiority or excellence that exceeds the general standard. Something said to be qualified if there is a match between the requirements that are owned by the desired object or service with the intent of the wish. According to Idrus et al. (2000) "... the purpose of fitness as perceived by the customer." For example, the quality of the learning process matches what is expected by the student; growing far beyond what one would expect more quality, if the opposite occurs, the more not qualified.

The first step is to identify the service quality measure service quality characteristics. List these characteristics can be generalized in many different ways using a variety of resources. One way is to look for literature like journals that may contain the dimensions of service quality. Researchers such as Parasuraman, Zeithaml, and Berry (1985) have concluded that the quality of services can be described on the basis of 10 dimensions. Trying to measure ten dimensions, it turns out the customer can only distinguish 5 service quality (servqual) dimensions, Parasuraman (1988) suggest that the dimensions of the original 10 overlapping one another. Five dimensions of service quality is something that materialized (tangible), reliability, responsiveness, assurance, and empathy. More on this dimension can be read from the publication of the quality of service by Zeithaml, Parasuraman and Berry (1990).

The first dimension of quality of service according to the concept of tangible servqual this is because a service cannot be kissed and intangible, tangible it becomes important as the size of the service. Tangible is the ability to provide the campus physical facilities and equipment adequate lecture concerning the appearance of employee/faculty and officials as well as public facilities. For example: the

availability of space concerning the completeness and availability of equipment, comfort and sophistication of the campus, computer and internet facilities, library, lecture halls, seminar rooms, faculty rooms, media lectures, laboratories, units of production, canteen, career guidance centers, health services, places of worship, rest areas and parking lots, as well as means of transport. Students will use the sense of sight to assess the quality of care of all the facilities and existing facilities.

Second, the reliability dimension is a dimension that measures the reliability of higher education in providing services to students. There are two aspects of this dimension are: (1) the ability of universities to provide services as promised with proper and reliable, and (2) the extent to which colleges provide a service that is not accurate or error. In other words, reliability is the ability of its officers, employees/faculty in providing services in accordance with the promised (on time), with immediate, relevant and accurate so as to satisfy the students. Example: development administration, curriculum and offers courses as demanded skill, profession and the world of work, the lectures take place smoothly on schedule, assessment study objective, fair and timely.

Third, responsiveness is a dynamic dimension of service quality, a willingness to help students and provide a service or services quickly. Expectations of students to the speed of service will almost certainly change the upward trend over time. Responsiveness is the willingness of officials, faculty/staff to assist and provide services according to the needs of the students. Example: officials easily found for the requested relief , the lecturer easily found for the purposes of consultation, ongoing learning process interactive and varied, and allows the students to develop the capacity and creativity, managers provide adequate facilities according to the needs of students and the world of work.

The fourth dimension of service quality dimensions that determine customer satisfaction is the assurance that the dimensions of quality assurance related to the company's ability and behavior of front- line staff in instilling a sense of trust and confidence to the students. Assurance includes competence, knowledge, skills, politeness, respect for every person and trust properties owned by the staff. Example: the lecturers delivered lectures in areas of expertise/experience, professors are always trying to add insight by reading, attending seminars, training , further study , do research , have a good attitude and behavior, as well as all levels of the organization reflects the professionalism and in accordance with the set in the standard.

Fifth, empathy is the ability of its officers, employees/faculty so as to give wholehearted service, among others, ease of communication, personal attention, awareness and understanding in student’s specific needs of individual students. Example: lecturers get to know the name of the students, faculty academic advisor truly act as a counselor and as a supervisor is not just a language editor, and officials can easily be reached either in the office, via phone, email and so on. This empathy relates to the development of human needs theory of Maslow. At the higher level, human needs are no longer the primary things such as physical, and social security are met, then two more needs to be pursued by humans, namely the need for ego and self-actualization. Lattertwo requirements which are associated with the dimensions of empathy.

Furthermore, Kotler (1994) stated that customer satisfaction is " ... the level of a person 's felt state resulting from comparing a product 's perceived performance

(outcomes) in relations to the person 's expectation." Student satisfaction is a condition that one feels that the result of comparison between the results expected of a product / service with the reality that is accepted. Thus, the level of student satisfaction depends on the fit between the achievement levels of quality of service or service purchased with student expectations.

In education, psychology, economics, business, and management, good judgment requires reliable measurement or trustworthy. Similarly, in the field of education and psychology. According to the Naga (1992) for educational and psychological measurement includes several things. First, measure the latent trait that is invisible to the respondent. Secondly, to measure the characteristics of the latent form of the questionnaire respondents were given stimulus or appropriate measuring instruments. Third, the stimulus responded by respondents with expectations correctly reflects the response latent trait to measure. Fourth, the response can be scored and interpreted adequately. Then, without question the extent to which scores obtained can accurately reflect the latent trait to be measured? Is the instrument was used as a stimulus to reveal the latent traits properly unseen? The second question regarding the validity. Being associated with reliability, whether the responses given by the participants are to be believed to be used as material for scoring psychological attributes that?

According to Wiersma (1986), reliability is the consistency of an instrument to measure something to be measured. Reliability indicates the extent to which the results of measurements with the device can be trusted. Therefore, reliability is an index that indicates the extent to which a measure can be reliable or unreliable. When an instrument is used repeatedly to measure the same symptoms and the results obtained are relatively stable or consistent, then the reliable instrument. In other words, the measurement results are expected to be the same if repeated measurements.

Broadly speaking there are three major categories of measurement reliability: (1) the type of stability (e.g., retest, parallel forms, and alternate forms), (2) the type of homogeneity or internal consistency (eg split half, Kuder-Richardson, Cronbach's alpha, theta and omega), and (3) type equivalent (e.g., parallel to the grain of alternate forms and inter-rater reliability. The instrument was given to one group of subjects and in a certain way calculated estimates of reliability. This one-time measurement approach generates information about the internal consistency of the instrument. Internal consistency is a measure such statements or to reflect the same aspect of a grain homogeneity statement.

The higher the reliability coefficient, the closer the value of observation scores with actual scores, so a score of observation can be used as a substitute for the real component of the score. Size of high or low reliability coefficient is not only determined by the value of the coefficient. The interpretation of high and low coefficient value obtained throughcomputation is also determined by the standard disciplines involved in the measurement. The higher the coefficient of reliability of an instrument, the possibility of errors that occur will be smaller if people make decisions based on the score obtained in the instrument.

In general, the measurement of affective characteristics provide reliability coefficient lower than the measurement of cognitive, because the cognitive characteristics tend to be more stable than affective characteristics. According to Gable (1986) cognitive reliability coefficient of the instrument usually about 0.90 or

more, whereas affective instrument reliability coefficient of less than 0.70. Reliability coefficient of 0.70 or more at the level generally accepted as a good reliability (Litwin, 1995). Whereasaccording to the Naga (1992) should be adequate reliability coefficient is above 0.75.

In each study using psychological measurement always apply the test validity and reliability. But along the way in the field of psychometrics, among the experts there were no agreement about the reliability coefficient or formula where for reliability among researchers. First, many researchers are considered quite competent still much less precise in reporting the reliability of the results of their measurements (Thompson, 1994).

Second, the problem that arises is the use of reliability coefficients by researchers consider a monotonically without the assumptions underlying coefficient. The researchers unknowingly using alpha coefficients were also without realizing that for this coefficient requires assumptions that are difficult to fulfill. If the assumptions are not met then the resulting alpha coefficient is the estimated value at the lowest limit. Many researchers only focused on the use of coefficient alpha to estimate reliability. Cronbach's alpha coefficient ppopularityies born because of factors: 1) computational technique is relatively easy, as it only requires information such as the total score variance, and 2) the sampling distribution is already known that the determination of confidence intervals on the population is very possible (Feld et al., 1987).

Third, the problems associated with the assumption that the condition for estimating the reliability. In the empirical realm other than the requirement of the nature of parallel, tau-equivalent terms is a tough challenge for researchers in developing measurement instruments. This is supported by Kamata et al (2003) who found that the assumption of equality, the power of discrimination between test components, and unidimensionality measurement is relatively difficult to achieve. If the assumption of tau-equivalenteessentially cannot be met then the coefficient alpha reliability values which produce very small, so it is below the estimatedcoefficients.

Fourth, discourse unidimensiononality measurement is the measurement problem. Unidimensionality is an important aspect in estimating reliability. Psychological measurement results are unidimension very difficult to achieve, especially in the context of the personality domain that contains most of the area variances broad traits. Socan (2000) writes that the factor analysis of several studies conducted many cases multidimensional compared with unidimensional.

Assuming the problem is not a major issue in developing a model of internal consistency, but the problem is the subject of study of many researchers in the assessment of reliability. As research Vehkahlati (2000) who concluded that the assumptions are not realistic enough to score a purely classical theory is the assumption that pure scores unidimensional practically difficult to prove. So study multidimensional surface measurements because many cases also found that the correlation between the dimensions of items in the sometimes higher than the correlation between the items in the test.

On the development of measurement instruments in the field of education assumes the use of many measurements that are conceptually formulated unidimension that there is only one kind of capability factors, personality, traits, and attitudes as measured by the measuring instrument. However, many studies have shown that the assumption unidimension difficult for the discovery of several new

factors involved are measured in one instrument. In other words, the instrument that is often used psychological research tends to be multidimensional.

Some important reasons multidimensional measurement reliability as suggested by Widhiarso (2009) with the following description: First, the general characteristics of the psychological construct are multidimensional.Second, any involvement in the preparation of the aspects of psychological instruments is usually preceded by a decrease in grain of some theoretical aspects and the tendency is multidimensional.

Third, the number of items in the instrument. The numbers of items that can add too much additional potential error variance in item giving rise to new dimensions of the original defined dimensions. Total grain and also forms the scale affect respondents' attitudes toward the item then affect their response to the instrument.

Fourth, item writing techniques. Spector and colleagues (1997) found that the technique of writing item that have reversed direction between positive (favorable) and negative (unfavorable) to form a new dimension when measuring the data capture many psychological scales using different writing techniques grain direction.

Fifth, different measurement units.Measurement of psychological fieldstends to have different measuring units between items one with the other item have different cavailities measure as an indicator construct. This condition will cause the measurement results tend to be multidimensional.

In Widhiarso research and Mardapi (2010) multidimensional model for measuring the reliability coefficient has high accuracy when compared with the reliability unidimension. Therefore, in this study, researchers focused only on multidimesional reliability and unidimensional reliability.This study aims to test the accuracy or reliability coefficient multidimensional accuracy when compared with a reliability coefficient unidimensionon. Based on the above it is raised a range of questions such as: What is the internal consistency reliability of a multidimensional instrument measuring student satisfaction as an internal customer? How does a comparison between the multidimesionalreliability and unidimensional reliability? Which is more accurate as a measure of reliability?

METHOD

The method used in this study was a survey method. The survey used in the data collection and made no treatment (treatment) or the conditioning of the variables studied, but only reveals the fact that there are symptoms based on the student or other respondents.In this trial obtained a sample of 173 respondents students from ManadoState University (Unima) in October 2013.

Scale research instrument made of two columns with details, for the first column is a reality or the fact that there was and is perceived by students to satisfy service quality with five alternative answers ranging from very dissatisfied (vd) value of 1, is not satisfied nd) value of 2, neutral (ne) value of 3, satisfied (sa) value of 4, and very satisfied (vs) value of 5. The scale was for the performance instrument.For the second column, the expectations of students to institutions with an alternative five-point scale based on the level of student interest with answers ranging from very unimportant (vu) value of 1, somewhat important (si) 2 values,

neutral (ne) 3 value, importance (im) value of 4, and very important (vi) value of 5. The scale was for expectation instrument.

RESULTS AND DISCUSION

1. Performance Instrument

The performance of the instrument consists of a 30 items questionnaire statement of student satisfaction as an internal customer. Thirtieth instrument of this item is the result of research that has been validated by researchers using factor analysis. The instrument consists of 30 items can be broken down as follows: 6 statements for tangible dimension, 7 statementsfor reliability dimension, 5 statementsfor responsiveness dimension, 7 statements for assurance dimension, and 5 statementsfor empathy dimension. First to unidimension Cronbach alpha reliability obtained directly using SPSS for 0.917.

Secondly, for which the reliability of composite reliability multidimensional omega McDonald, using the program LISREL 8.8 and Excel programs obtained: $\sum_{i=1}^{i} \lambda_i = 17.330$ and $\sum_{i=1}^{i} 1-\lambda_i^2 = 19.524$, so $\omega = \frac{(17.330)^2}{(17.330)^2 + (19.524)} = 0.939.$

Third, for the reliability of the multidimensional construct reliability obtained the same results as follows: $\sum_{i=1}^{i} \lambda_i = 17.330$ and $\sum_{i=1}^{i} \delta = 19.550$,so

$$CR = \frac{(17.330)^2}{(17.330)^2 + (19.550)} = 0.939.$$

Fourth, for the reliability of the multidimensional maximum reliability, by using by using LISREL 8.8 and Excel programs obtained: $\sum_{i=1}^{p} \frac{l_i^2}{(1-l_i^2)} = 12.744$,

so it can be calculated as follows: $\Omega_w = \frac{12.744}{1+12.744} = 0.927.$

2. Expectation Instrument

Same with the performance instrument, the expectation instrument consists of 30 items, hopes the statement level of student satisfaction questionnaire as an internal customer. First, for reliability alpha Cronbah unidimension obtained directly using SPSS for 0.932.

Secondly, for which the reliability of composite reliability multidimensional omega McDonald, using the LISREL 8.8 program using the Excel program and obtained: $\sum_{i=1}^{i} \lambda_i = 20.280$ and $\sum_{i=1}^{i} 1-\lambda_i^2 = 15.700$, so obtained

$$\omega = \frac{(20.280)^2}{(20.280)^2 + (15.700)} = 0.964.$$

Third, for the reliability of the multidimensional construct reliability obtained the same results as follows: $\sum_{i=1}^{i} \lambda_i = 20.280$ and $\sum_{i=1}^{i} \delta = 15.450$, so

$$CR = \frac{(20.280)^2}{(20.280)^2 + (15.450)} = 0.964.$$

Fourth, for maximum reliability, using LISREL 8.8 program using the Excel program and obtained: $\sum_{i=1}^{p} \frac{l_i^2}{(1-l_i^2)} = 23.260$, and so can be calculated as follows:

$$\Omega_w = \frac{23.260}{1+23.260} = 0.959.$$

From the results above description can be summarized in the form of table as follows:

Table 1. Summary of Research Findings

Reliability	α	$CR = \omega$	Ω_w
Performance	0.917	0.939	0.927
Expectation	0.932	0.964	0.959

The calculations for the above two instruments multidimensional reliability coefficient obtained relatively higher or more precise than the reliability coefficient unidimensional. It is no agreement among experts psychometrics. But among researchers in Indonesia should be after knowing the most appropriate tool should start using the tool correctly and adequately.

Indeed, most researchers among the faculty and students of both S2 and S3 do not know the formula for calculating the reliability coefficient construct, omega or the maximum reliability. So this time it's time to introduce and use the formula. With grounds already know the formula and most psychological constructs, personality, education, and social is multidimensional, so that all students and faculty researchers developing and growing to dig deeper about the reliability coefficient others.

Measurement of education is something that is quite complicated. Various writings in journals ranging from educational measurement at the measurement method are expected to provide results that are valid, reliable, and accurate. Of business experts is not easy because the longer the experts bring educational measurement that far into the area of mathematics. Without a wellmastered high mathematics and complicated, we cannot understand the various education journals measurements. So far, we are so far behind in the field of educational measurement. Very few science education experts who are able to understand the content of educational measurement journal studded high level mathematics. Therefore, it needs to be increased science education experts in the field of educational measurement.

Businesses can start by changing our perception has been that long ago, the educators among us have the notion that science education and psychology do not require math. Now, dealing with educational measurement among educators we need

to change their perception of mathematics. Educators need to be aware that there is a part of science education which hardly uses mathematics, but there is also a part of science education is in need of mathematics, such as the example above multivariate statistics requires high mathematical skills.

CONCLUSIONS

Based on the test results of this study concluded multidimensional reliability coefficient is more precise or accurate when compared with a reliability coefficient unidimensional.

Suggestions can be submitted are as follows: first, estimate the instrument needs to be tested further by using another formula that is not based SEM. Second, because this study used a five-point scale when it is necessary to continue to use a variety of different scales, such as the semantic differential scale, dichotomous scale, Thurstone scale, and so on. Third, these instruments need to be tested using a larger sample population and the wider setting and involves several provinces at the same time, also the school.

REFERENCES

Capraro, M. M., R. M. Capraro, dan R. K. Herson. (2001). "Measurement Error of Score on the Mathematics Anxiety Rating Scale across Sudies." *Educational and Psychological Measurement*, 61, 373–386.

Feld, I. S., D. J. Woodruff, dan F. A. Salih. (1987). Statistical Inference for Coefficient Alpha." *Applied Psychological Measurement*, II,: 93 – 103.

Gable, R. K. (1986.) *Instrument Development in the Affective Domain*. Amsterdam: Kluwer Nijhoff Publishing.

Geffen, D., D. W. Straub, dan M. D. Boudreau. (2001). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of AIS*. Volume 4, Article 7.

Hancock, G. R., dan R. O. Mueller. (2000). "Rethinking Construct Reliability within Latent Variable Systems. *Stuctural Equation Modeling: Present and Future*, R. Cudek, S. H. C. duToit, dan D. F. Sorbom (Eds.), Chicagp: Scientific Software International.

Idrus, N., *et al*. (2000). *Quality Assurance*. Jakarta: Directorate General of Higher Education.

Kamata, A., A. Turhan, dan E. Darandari. (2003). *Estimating Reliability for Multidimensional Composite Scales Scores*. *Paper* presented in Annual Meeting of American Educational Research Association at Chicago, April.

Kotler, P. (1994). *Manajemen Pemasaran*, terjemahan Jaka Wasana. Jakarta: Erlangga.

Latan, Hengky. (2012). *Structural Equation Modeling Konsep dan Aplikasi Menggunakan Program Lisrel 8.80*.Bandung: Alfabeta.

Lee, S. Y., dan X. Y. Song. (2001). Hyphotesis Testing and Model Comparison in Two-level Structural Equation Model. *Multivariate Behavioral Research,* Volume 36 (4): 639–655.

Litwin, M. S. (1995). *How to Measure Survey Reliabity and Validity*. London: Sage Publications.

McDonald, R. P. (1981). The Dimensionality of Test and Items. *British Journal of Mathematical and Statistical Psychology,* 34, 100 – 117.

Naga, D. S. (1992). *Teori Sekor*. Jakarta: Gunadarma Press.

Parasuraman, A. (1988). Servqual: A Multi-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing,* 64(1), 12 – 37.

Parasuraman, A., V. A. Zeinthaml, dan L. L. Berry. (1985). Conceptual Model of Service Quality and Its Iimplications for Future Research. *Journal of Marketing,* 49, 41 – 50.

Socan, G. (2000). Assessment of Reliability when Test Items are not essentially t-Equivalent. *Development in Survey Methodology*, Anuska Feligoj and Andrej Mrvar (Eds.), Ljubljana: FDV.

Spector, P., P. Brannick, dan P. Chen. (1997). When Two Factors Don't Reflect Two Constructs: How Item Characteristics Can Produce Artifictual Factors. *Journal of Management,* 23 (5), 659 – 668.

Thompson, B. (1994). Guidelines for Author. *Educational and Psychological Measurement,* 54,: 837 – 847.

Vehkalahti, K. (2000). Reliability of Measurement Scales Tarkkonnen's General Method Supersedes Cronbach's Alpha. *Academic Dissertation*, University of Helsinki, Finland.

Widhiarso, W., dan Djemari Mardapi. (2010). Komparasi Ketepatan Estimasi Koefisien Reliabilitas Teori Skor Murni Klasik. *Jurnal Penelitian dan Evaluasi Pendidikan,* 14 (1),: 1 – 19.

Widhiarso, Wahyu. (2009). Koefisien Reliabilitas pada Pengukuran Kepribadian yang Bersifat Multidimensi. *Psikobuana*, 1 (1),: 39 – 48.

Wiersma, W. (1986). *Research Methods in Education: An Introduction.* Boston: Allyn and Bacon, Inc.,.

Zeinthaml, V. A., A. Parasuraman, dan L. L. Berry. (1990). *Delivering Quality Service: Balancing Customer Perceptions and Expectations.*London: The Free Press.