ME - 27

MULTIDIMENSIONAL RELIABILITY ESTIMATION IN INSTRUMENT OF STUDENTS' SATISFACTION AS AN INTERNAL COSTUMER

Gaguk Margono
Universitas Negeri Jakarta, Kampus UNJ, Jl. Rawamangun Muka, Rawamangun, Jakarta 13220,
Email: g_margono@yahoo.com

Abstract

The purpose of this paper is to compare multidimensional and unidimensional reliability toward 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 measuring 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 99 students. The result of the calculation is concluded that the measuring instrument 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, instrument of students' satisfaction as an internal costumer, confirmatory factor analysis

INTRODUCTION

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 immeasurable variables called latent constructs. This construct created by a set of measurable variables, namely measureable construct. Error of measurement which reflects score reliability is seen as a unique construct and become an important part of SEM analysis. The error include of measurement in the SEM becomes a benefit for using SEM compared to other analysis techniques (Capraro and colleagues, 2001). SEM can estimate error variance which actually estimates reliability.

According to Geffen and colleagues (2001), SEM is a multivariate statistical technique that combines multiple regressions. This technique identifies relationships between constructs and factor analysis also recognize immeasurable concept through some manifest indicators. Both multiple regressions and factor analysis are used simultaneously.

The first approach is attenuation correlation correction which caused by measurement error. The second approach is a structural equation model in context of confirmatory factor analysis. Lee and Song (2001) said that SEM is one approach to confirm the measurement model. SEM measurement model links latent constructs with empirical construct. Empirical

constructs expressed by a combination of latent constructs. SEM can process generalizability and item response theory. It is also able to compare measurement models and facilitates the accuracy of accurate model 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 dependent variable. Geffen and colleagues (2001) said that the measurement model is sub-models in SEM which identifies latent construct with its indicators. This identification can be used to determine which shown through loading items produced. Based on the SEM perspective 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 loading of standardized indicators to-i

δ = Standard error of measurement

McDonald (1981) formulates a reliability coefficient which later named as McDonald composite score reliability coefficient which also called omega (ω). Reliability coefficient is based on confirmatory factor analysis that is part of the menu SEM modeling. This composite score explains the amount of indicators proportion of indicators which explains measurement construct. The formula for construct reliability coefficients is:

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

Description:

λ_i = Factor loading of standardized indicators to-i

Construct reliability or McDonald composite score reliability will give the same result as $\delta = 1 - \lambda^2$.

Hancock and Mueller (2000) develop construct reliability coefficients which show how well instrument indicators could reflect the construct which is going to be measured. This coefficient is a modification of McDonald construct reliability coefficient. It cannot accommodate different weights of interdimensions. The result of the modification called weighted coefficients reliability, it can be shown as:

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

Description:

l_i = Coefficient of the i-th standardized dimensions

The reliability coefficient can be interpreted as the square of the correlation between dimensions with optimal linear composites score. Because of this some experts call the coefficient as maximum reliability.

Students' satisfaction is measured by using subjective or soft measurement applied as quality indicators. It is called soft because this measurement focuses on attitudes and perceptions, not concrete things (objective criteria). Therefore, the instrument utilized for this could be student's satisfaction questionnaire which discusses about service quality of the institution.

Quality is a term that is constantly moves dynamically; if it moves forward, the quality is said to be better, otherwise it moves backward it is defined deteriorate. Quality can be understood as superiority or excellence that exceeds the general standard. Something said to be qualified if there is a match between the requirements of the desired object or service with the purpose of the person who wants it. According to Idrus and colleagues (2000) "... the purpose of fitness as perceived by the customer." For example, the quality of the learning process matches with what is expected by the students; the better it is served, the higher quality it is stated.

The first step to measure service quality is identifying characteristics of service quality. List of these characteristics can be generalized in many different ways using various resources, for instance, by studying literature such as journal that may contain dimensions of service quality. Researchers such as Parasuraman, Zeithaml, and Berry (1985) stated that the quality of services can be described on into 10 dimensions. When it is studied, respondents can only distinguish 5 dimensions of service quality (servqual). Parasuraman (1988) comments that the 10 original dimensions are overlapped. The five dimensions of service quality are materialized (tangible), reliability, responsiveness, assurance, and empathy. More information about this can be read from the publication about service quality, written by Zeithaml, Parasuraman, and Berry (1990).

Service is untouchable and invisible, so the tangible dimension becomes more important. First, tangible is the ability to provide physical facilities of campus and equipments adequate lectures 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 equipments, comfortable and sophisticated 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 see and judge quality of all facilities.

Second, reliability dimension is a dimension that measures the reliability of higher education in providing services to students. There are two aspects of this dimension, they are: (1) the ability of universities to provide services as informed to students, and (2) how far college provides correct and accurate service. In other words, reliability is the ability of its officers, employees/faculty in providing services on time, relevant and accurate to satisfy students. These aims are done by developing good administration, curriculum and courses which produce skills, profession and meet job requirement. The lectures run smoothly as scheduled and the assessment is arranged well also encourage students to increase their ability.

Third, responsiveness is a willingness to help students and provide service quickly. Students' expectations to the speed (fast) service will almost certainly change over time. Responsiveness is the willingness of officials, faculty/staff to assist and provide services according to students' needs. For instance, officials easily found for help, lecturers easily met for consultation, there is interactive class in learning which allows students to develop their capacity and creativity for getting better future.

Fourth, assurance dimension which relates to the company's ability and front-line staff professionalism to provide a good service. Assurance includes competence, knowledge, skills, behavior and characteristics, professionalism in work, owned by lecturers, administrators, and staff. For example, the lecturers deliver lectures in areas of their expertise/experience, add their insight by reading, attending seminars, training, further study, do research, have a good attitude and behavior, and all the persons of the institution have professionalism and regulated in the rule or standard.

Fifth, empathy is the ability of officers, employees/faculty, and academic advisors to give good service, good communication, personal attention, awareness and understand students' specific needs. It can be illustrated as lecturers know the name of the students, faculty academic advisor truly act as a counselor and as a supervisor rather than just as 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 as primary things such as physical and social needs, but also needs of ego and self-actualization. Later, the last two needs become the dimension 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." Thus, the level of students' satisfaction depends on match of between achievement level of service quality or service purchased with students' expectation.

In education, psychology, economics, business, and management, good judgment requires reliable measurement or trustworthy. According to Naga (1992) educational and psychological measurement includes several things, they are: first, measure the latent trait which is invisible to the respondent. Secondly, to measure the characteristics of the latent form, respondents given stimulus or appropriate measuring instruments. Third, the stimulus responded by respondents should correctly reflects the latent trait. Fourth, the response can be scored and interpreted adequately. One thing which is important is how far the score achieved shows the latent trait? Does the stimulus express the characteristic of the latent trait? Are the responses given by participants reliable for scoring psychological attributes that? Those questions regard to validity and reliability of the measurement.

According to Wiersma (1986), reliability is the consistency of an instrument to calculate something to be measured. Reliability indicates how far the instrument and its results can be trusted. Therefore, reliability is an index that indicates a measurement is reliable or not. If the use of an instrument is repeated for couple of times for the measuring the same symptoms, then the instrument could be reliable. In other words, the measurement results are expected to be the same if the measurements repeated.

Generally, there are three major categories of measurement reliability, they are: (1) type of stability (e.g. retest, parallel forms, ad alternative forms), (2) type of homogeneity or consistency internal (e.g., split half, Kuder-Richardson, Cronbach's alpha, theta and omega), and (3) type of equivalent (e.g. parallel to alternative forms and inter-rater reliability). The instrument was given to one group of subjects once, and then the reliability is estimated. This kind of measurement approach generates information about the internal consistency of the instrument. Internal consistency reflects the same aspect of homogeneity statement.

The higher the reliability coefficient, the closer the value of observation scores with actual scores. So the observation score can be used as substitute for the real component of the score. High or low reliability coefficient is not only determined by the value of the coefficient. The interpretation of high and low coefficient value obtained through computation and also standards of disciplines involved in the measurement. The higher the coefficient of reliability of an instrument, the smaller errors occur in decisions.

Commonly, the measurement of affective characteristics provide lower reliability coefficient than the cognitive measurement, 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 is less than 0.70. Reliability coefficient of 0.70 or generally accepted as a good reliability (Litwin, 1995). In addition, Naga (1992) says that the adequate reliability coefficient should be above 0.75.

Psychological measurement always applies validity and reliability test. But in psychometrics, there were no agreement among the experts about the reliability coefficient or formula for reliability among researchers. The problems are: first, many researchers who considered competent but do not report the reliability of their measurement (Thompson, 1994).

Second, reliability coefficients are used monotonically by researchers without considering assumptions that underlie the coefficient. The researchers use alpha coefficients were also without realizing that for this coefficient requires assumptions that are difficult to fulfill. If the assumptions are not achieved, so resulting alpha coefficient is the lowest value of estimation. Many researchers only focused on the use of coefficient alpha to estimate reliability. The popularity of Cronbach's alpha coefficient caused by some factors such as: 1) its computational technique is relatively easy, as it only requires information such as the total score variance, and 2) sampling distribution is already known so the determination of confidence intervals on population is very possible (Feld and colleagues, 1987).

Third, the problems associated with assumption which becomes requirement for estimating reliability. In the empirical realism, instead of parallel characteristic requirement, *tau-equivalent* becomes a challenge for researchers in developing measurement instruments. This is supported by Kamata and colleagues (2003) who found that the assumption of equality, the power of discrimination between test components, and unidimensionality measurement is relatively difficult to achieve. If *tau-equvalent* assumption cannot be gained then the coefficient alpha reliability values which produced is very small. It is below the estimated coefficients.

Fourth, the problem is dealing with unidimensiononality measurement. Unidimensionality is an important aspect in estimating reliability. Result of psychological unidimension measurement is very difficult to achieve, especially in the context of the personality domain that contains broad variances area traits. Socan (2000) writes that most of factor analysis in several studies conducted are multidimensional case, not unidimensional.

Assumption is not major issue in developing a model of internal consistency, but it becomes subject of study of many researchers in reliability assessment. Like research done by Vehkahlati (2000) that concluded that unrealistic assumptions in classical theory is assumption of purity scores of unidimensional which practically is difficult to prove. It makes study of multidimensional measurements grow because many cases found that correlation between dimensions of items higher than the correlation between items in test.

Most of education measurement use unidimension. This measurement conceptually process one capability factor, personality, trait, and attitude which measured by one measuring instrument. However, many studies shown that the unidimension assumption is difficult since there is another new factor measured in one instrument. In other words, the instrument that is often used in psychological research tends to be multidimensional.

Some important reasons why multidimensional measurement reliability is important suggested by Widhiarso (2009). He describes some reasons, they are: First, the general characteristics of the psychological construct is multidimensional. Second, any involvement in preparation of psychological aspect instruments are usually preceded by having item from theoretical aspects and its tendency is multidimensional.

The third reason is the number of items in the instrument. Too much items may add more error variance potential in item and this will encourage new dimensions. The total of the

items and forms the scale affect respondents' attitudes toward items which then affect their response to the instrument.

The fourth reason is items writing techniques. Spector and colleagues (1997) found that item writing technique which has reversed direction between positive (favorable) and negative (unfavorable) direction could form a new dimension. In fact many psychological scales using different item writing techniques with reversed direction in collecting data.

Different measurement units become the fifth reason. Psychological measurement tends to have different measuring units between one item with another item and it has different capability as measurement construct indicator. This condition will cause the measurement results tend to be multidimensional.

Widhiarso and Mardapi (2010) did research on multidimensional model. It has higher accuracy reliability coefficient than unidimension model. Therefore, in this study, researchers focused only on internal consistency coefficient like α untuk unidimensional reliability and ω, CR dan Ω_w.

This study aims to test the accuracy of reliability coefficient multidimensional compared to reliability coefficient of unidimension. Based on the descriptions above, some questions arise: What is the internal consistency reliability of multidimensional instrument measuring student satisfaction as an internal customer? How is the comparison between multidimesional reliability and unidimensional reliability? Which one is more accurate to measure reliability?

RESEARCH METHOD

The method used in this study was a survey method. There was no treatment in this survey. The survey reveals the fact based on symptoms found from students or respondents. In this trial obtained a sample of 103 respondents students from Yogyakarta State University (UNY) in October 2013.

The instrument has scale, it is divided into two columns: first column is a reality or the fact perceived by students about service quality. There are five alternative answers ranging from Very dissatisfied (Vd) values of 1, is Dissatisfied (D) values of 2, Neutral (N) values of 3, Satisfied (S) values of 4, and Very satisfied (Vs) values 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) values of 1, Somewhat unimportant (Su) values of 2, Neutral (N) values of 3, Importance (I) values of 4, and Very important (Vi) values of 5. The scale was for expectation instrument.

RESULTS OF RESEARCH AND DISCUSSION

1. For Performance Instrument

The performance of the instrument consists of a 30 items questionnaire statement of students' satisfaction as an internal customer. Thirtieth items of the instrument are the result of research that has been validated by researchers using factor analysis. The instrument consists of 30 items, divided into: 6 statements for tangible dimension, 7 statements for reliability dimension, 5 statements for responsiveness dimension, 7 statements for assurance dimension, and 5 statements for empathy dimension. First, for unidimension Cronbach alpha reliability obtained directly using SPSS, that is 0.917.

Secondly, for which the reliability of McDonald omega composite reliability in

multidimensional, using LISREL 8.8 and Excel programs, it is obtained: $\sum_{i=1}^{i}\lambda_i = 17.020$ and $\sum_{i=1}^{i}1-\lambda_i^2 = 19.870$, so $\omega = \frac{(17.020)^2}{(17.020)^2 + (19.870)} = 0.936.$

Third, for the reliability of the multidimensional construct reliability, it is obtained the same results as follows: $\sum_{i=1}^{i}\lambda_i = 17.020$ and $\sum_{i=1}^{i}\delta = 19.930$, so

$$CR = \frac{(17.020)^2}{(17.020)^2 + (19.930)} = 0.936.$$

Fourth, reliability of multidimensional maximum reliability, by using LISREL 8.8 and Excel programs, it is obtained: $\sum_{i=1}^{p}\frac{l_i^2}{(1-l_i^2)} = 12.584$, so it can be calculated as follows:

$$\Omega_w = \frac{12.584}{1+12.584} = 0.926.$$

2. For Expectation Instrument

The amounts of the items are the same as the performance instrument. The expectation instrument consists of 30 items. First, for reliability alpha Cronbach unidimension obtained directly using SPSS, that is 0.920.

Secondly, for which the reliability McDonald omega composite reliability multidimensional, using the LISREL 8.8 and Excel program, it is obtained: $\sum_{i=1}^{i}\lambda_i = 19.060$ and $\sum_{i=1}^{i}1-\lambda_i^2 = 17.489$, so obtained

$$\omega = \frac{(19.060)^2}{(19.060)^2 + (17.489)} = 0.954.$$

Third, for the reliability of the multidimensional construct reliability obtained the same results as follows: $\sum_{i=1}^{i}\lambda_i = 19.060$ and $\sum_{i=1}^{i}\delta = 17.750$, so

$$CR = \frac{(19.060)^2}{(19.060)^2 + (17.750)} = 0.954.$$

Fourth, for maximum reliability, using LISREL 8.8 and Excel program, it is obtained: $\sum_{i=1}^{p}\frac{l_i^2}{(1-l_i^2)} = 16.098$, and so it can be calculated as follows: $\Omega_w = \frac{16.098}{1+16.098} = 0.942.$

From the results above it can be summarized as table 1 below:

Table 1. Summary of Research Findings

Reliability	α	$CR = \omega$	Ω_w
Performance	0.917	0.936	0.926
Expectation	0.932	0.954	0.942

The calculations for the above two instruments are the multidimensional reliability coefficient obtained relatively higher reliability coefficient than unidimensional model. Still there is not agreement among experts psychometrics about this. But for Indonesia researchers there should be change in using way to find out reliability after knowing the most appropriate, correct and adequate tools.

Indeed, most researchers among the faculty and post graduate students (of both S2 and S3)) do not know the formula for calculating the coefficient of reliability construct, omega or maximum reliability. This is the time to introduce and use the formula because many psychological constructs, personality, education, and social are multidimensional. So researchers and students can develop and grow insight and information about other reliability coefficient.

Education measurement is something that is quite complicated. It is hoped that various writings in journals provide valid, reliable, and accurate result. Experts try to bring educational measurement far into the area of mathematics. Without mastering high level and complicated mathematics calculation, we cannot understand the various journals of education measurements. So far, we left behind in the field of educational measurement. Only few science education experts who understand the content of educational measurement journals. These journals consist of high level mathematics. Therefore, the amount of science education experts in educational measurement should be increased.

The first effort to do it is by changing our perception that education and psychology measurement do not require math. Educators need to change their perception of mathematics. Educators need to be aware that there is a part of education which hardly uses mathematics such as the example above uses multivariate statistics that require high mathematical skills.

CONCLUSION AND SUGGESTION

Based on the test results of this study it can be concluded that multidimensional reliability coefficient is more precise or accurate when compared to reliability coefficient in unidimensional.

There are some suggestions for this: first, estimation of the instrument need to be tested further by using another formula which is not based on SEM. Second, because this study used a five-point scale, it is necessary to continue to use 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 population sample and the wider setting and involves several provinces at the same time, also the school level and different type of university or college.

Thank statement. Thanks to the Directorate of Higher Education, Ministry of Education and Culture. Research Institute of UNJ for Fundamental Research grants for Fiscal Year 2013, and all students who have been willing to be the respondents; the data of this paper is part of Fundamental Research finding.

BIBLIOGRAPHY

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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