

Multidimensional Reliability of Instrument for Measuring Students' Attitudes Toward Statistics by Using Semantic Differential Scale

Gaguk Margono*

Engineering Department, State University of Jakarta, Indonesia, Kampus UNJ, Jl. Rawamangun Muka, Rawamangun, Jakarta
*Corresponding author: gmarginono@unj.ac.id

Received December 24, 2014; Revised January 08, 2015; Accepted January 13, 2015

Abstract The purpose of this paper is to compare multidimensional and unidimensional reliability of instrument of students' attitude towards statistics by using semantic differential scale. Few researches utilized multidimensional reliability measurement. Multidimensional reliability is calculated by using Confirmatory Factor Analysis (CFA) in Structural Equation Model (SEM) technique. The measurement and calculation described in this article deal with instrument of students' attitude towards statistics. This instrument has been tried out to 150 students. It is found that multidimensional reliability has higher accuracy compared to the unidimensional one. Perhaps various formulas applying multidimensional reliability would be used in future research.

Keywords: *multidimensional reliability, attitudes toward statistics using semantic differential scale, confirmatory factor analysis*

Cite This Article: Gaguk Margono, "Multidimensional Reliability of Instrument for Measuring Students' Attitudes Toward Statistics by Using Semantic Differential Scale." *American Journal of Educational Research*, vol. 3, no. 1 (2015): 49-53. doi: 10.12691/education-3-1-10.

1. Introduction

Education requires reliable or trustworthy measurement judgment. According to Naga (1992), educational and psychological measurements include several things. First, it measures respondents' latent characteristics. Second, to measure the latent characteristics, respondents are given stimulus through questionnaires or appropriate measuring instruments. Third, perhaps respondents' responses are reflection of the latent characters. Fourth, the responses are scored and interpreted adequately [1]. Then, some questions rise such as: How do the scores reflect the latent characteristics accurately? Does the instrument reveal the latent characteristics (traits) properly? Both those questions regard to validity. In associate to reliability, we can ask: Are respondents' answers realistic to be used for scoring psychological attributes?

Anything used to measure can be said as measurement tools. The instrument has to be validated before used. Basically, there are two kinds of instruments; they are test and non test. Test measures maximum performance and non test calculates attitude (typical performance). Test will have true or false answer, while non test has positive or negative answer. According to Suryabrata (2000) non test measurement needs sentiment expression response that is response which cannot be judged as true or false answer. Responses here often regarded as true answers relate to each response reason [2]. Non test does not judge what someone can do; rather it values what someone tends to do.

In scientific research, a good instrument is achieved through data and it can be better interpreted through a reliable, valuable and objective process.

According to Wireman (1986) reliability is consistency of an instrument to calculate something measured [3]. Reliability indicates how far results of measurement of an instrument can be trusted. Therefore, reliability is an index for indicating if an instrument is valuable and believable. An instrument can be stated as a reliable instrument when it measures same symptoms repeatedly and results obtained are relatively stable or consistent.

Generally, there are three major categories of reliability measurement, they are: (1) stability type (e.g. retest, parallel items, and alternative forms), (2) homogeneity or internal consistency type (e.g. split half, Kuder-Richardson, Cronbach's alpha, theta and omega), and (3) equivalent type (e.g. parallel items in alternative forms and inter-rater reliability). In doing this, an instrument is given to one group of subjects once and the reliability estimation is calculated in certain way. This once type application measurement approach generates information about statement consistency of same aspects or reflect statements homogeneity.

The higher reliability coefficient of an instrument, the closer observed scores to the real scores. So the observed scores can be used as substitution of real scores. Coefficient is not the only component in deciding level of reliability. Level of reliability is acquired through calculation. It also influenced by standard of discipline involves in the measurement. Errors decrease when using a high reliability coefficient instrument.

Commonly, affective characteristics measurement provides lower reliability coefficient than cognitive measurement. It is caused by less stable score of affective characteristics. According to Gable (1986) reliability coefficient of cognitive instrument is usually about 0.90 or more [4], whereas affective instrument reliability coefficient is less than 0.70. Reliability coefficient of level 0.70 or more usually can be accepted as a good reliability (Litwin, 1995) [5]. However, Naga (1992) says that an adequate reliability coefficient should be above 0.75 [1].

Psychological research measurement always applies validity and reliability test. But in psychometrics, experts still argue about reliability coefficient also inter-rater reliability formula. It is caused by some reasons: first, many competence researchers give less correct reliability of their measurements result (Thompson, 1994) [6]. Second, some researchers use reliability coefficients monotonously without considering assumptions underlies the coefficient. The researchers do not acknowledge well the use of alpha coefficients which require hard completion assumptions. For this case, if assumptions do not suit the requirement so alpha coefficient estimates the lowest limit point. Many researchers use alpha coefficient to estimate reliability. The great range of Cronbach's alpha coefficient use caused by some factors: 1) computational technique used in processing data to get reliability coefficient is relatively easy. It only requires total score variance, and 2) sampling distribution is already provided so it is possible to decide true intervals of population (Feld et al., 1987) [7].

Third, the problem deals with assumptions in estimating reliability. Empirical measurement requires parallelism. However, tau-equivalent aspect is more complex requirement for measurement. This statement supported by Kamata et al (2003) who found that assumption of equality, test components discrimination, and unidimensionality measurement are relatively difficult to be achieved [8]. If tau-equivalent assumption cannot be obtained essentially so alpha coefficient produces very small reliability point (value). The coefficient value lies under estimation.

Fourth, the main discourse issue in measurement is problem when applying unidimensional measurement. Unidimensionality is an important aspect in estimating reliability. Unidimension psychological measurement result is very difficult to be reached, particularly in personality domain context. This domain contains broad traits variances area. Socan (2000) writes that multidimensional factor analysis studies are conducted more than the unidimensional one [9].

Assumption problem is not a major issue in setting internal consistency models. But it becomes the most chosen topic in reliability study. Research done by Vehkahlati (2000) found that unrealistic assumption in pure classical theory is genuine unidimensional score. Practically, this condition is hard to prove [10]. So, study of multidimensional measurement comes as a solution for this. In addition, many cases discovered that there is inter-item correlation in the dimension. Sometimes the correlation is greater than item correlation in test.

Education researchers use unidimensional assumption measurement. This measurement has concept that there is only one factor of ability, personality, affective and attitude which measured by one measurement instrument. But many research showed that unidimensional assumption

is difficult to be gained. Because some new factors also discovered when doing the measurement. In other words, psychological instrument which often used by researchers tends to be multidimensional.

Multidimensional reliability measurement is important for on some reasons. According to Widhiarso (2009) the reasons are: first, generally, characteristics of psychological construct are multidimensional. Second, psychological instrument involves some aspects which usually started with items generated from some theoretical aspects. The items tend to be multidimensional. The third reason is the amount of items in the instrument. Too many items can add errors variants potential in items. It may create new dimensions. Amounts of items and scale form influence respondent attitude towards items. This influence will persuade their response to the instrument [11]. Fourth reason is item writing techniques. Spector and colleagues (1997) found that item writing technique of two way direction response; positive (favorable) and negative (unfavorable) response may create new measurement dimension [12]. In fact, psychological scale uses item writing technique which has different direction in collecting data. Fifth reason relates to different measurement units. Psychological measurement is likely to have different measuring units between one item with other items. It has different capability measured as of measurement construct. This condition will cause multidimensional result.

Based on the descriptions above it can be concluded that psychological measurement tends to be multidimensional measurement rather than one-dimensional in both measure cognitive or non cognitive construct. It is suggested that psychometrics measurement involves multidimensional model analysis technique.

McDonald (1981) formulates are liability coefficient which namely McDonald composite score reliability coefficients or omega [13]. The reliability coefficient based on confirmatory factor analysis which is part of SEM modeling menu. This McDonald composite score reliability explains size of indicators proportion in explaining measured construct. Formula for obtaining construct reliability coefficients is as follows:

$$\omega = \frac{\left(\sum_{i=1}^i \lambda_i \right)^2}{\left(\sum_{i=1}^i \lambda_i \right)^2 + \left(\sum_{i=1}^i 1 - \lambda_i^2 \right)} \quad [13]$$

Descriptions:

ω = McDonald composite reliability (omega)

λ_i = Factor loading of standardized indicators to-i

According to the Latan (2012) Structural Equation Modeling (SEM) is a second-generation multivariate analysis technique that combines factor analysis and path analysis. This technique allows researchers to simultaneously test and estimate relationship between exogenous and endogenous multiple variables with many indicators [14]. In 1970s Joreskog' research discovered statistical theory of linear structural analysis which is better known as structural equation modeling or SEM. This modeling uses analysis of covariance structure. So this approach sometimes called as covariant structure model (CSM). The model includes immeasurable variables called latent

constructs which created by a set of measured variables, namely measured construct. Measurement error which reflects scores reliability of measurement is seen as a unique construct. It becomes an important part of SEM analysis. The *error* measurement becomes the advantage of SEM analysis compared to other analysis techniques (Capraro et al., 2001). SEM can estimate error variance of measurement outcome scores that actually estimate reliability [15].

According to Geffen and colleagues (2001), SEM is a multivariate statistical technique that combines multiple regressions which identify relationships between constructs and factor analysis [16]. SEM recognizes immeasurable concept through some manifest indicators which both work simultaneously. SEM has some advantages compared to other analysis techniques. In studying relationship among variables, SEM automatically reduces measurement error effect. Capraro et al., (2001) says that independent variable influence towards dependent variable persuaded by attenuation effect [15]. The value of this effect is not over the range of reliability coefficient of test score. First approach for this situation is attenuation correlation correction which caused by measurement error. Second approach is structural equation modeling in confirmatory factor analysis context. Lee and Song (2001) say that SEM is one approach to confirm measurement model [17]. SEM measurement model links latent constructs to empirical construct. Empirical constructs are expressed by combination of latent constructs. SEM may be used in generalizability theory analysis and item response theory. SEM is also able to compare measurement models and facilitates investigation of model accuracy.

Sub model of SEM is factor analysis. Factor analysis is useful for detecting measurement instrument dimension. This technique introduced by Spearman relates to intelligence factors exploring. SEM also identifies construct reliability that appears through loading items point produced. Construct reliability is counted through SEM uses this formula:

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

Descriptions:

CR = Construct reliability

λ_i = Factor loading of standardized indicators to- i

δ = Standard error of measurement

This constructs reliability gives same result with McDonald composite score reliability (ω) because $\delta = 1 - \lambda^2$.

The following rule of multidimensional reliability coefficient is construct reliability coefficients developed by Hancock and Mueller (2000) [18]. It shows how well instrument indicators can reflect construct which is being measured. This coefficient is a modification of McDonald construct reliability coefficient which cannot accommodate different weights of interdimensions. The modification result is called weighted construct reliability coefficients as follows:

$$\Omega_w = \frac{\sum_{i=1}^p \frac{l_i^2}{(1-l_i^2)}}{1 + \sum_{i=1}^p \frac{l_i^2}{(1-l_i^2)}} [18]$$

Descriptions:

Ω_w = weighted construct reliability (maximum)

l_i = Coefficient of the i -th standardized indicator

The reliability coefficient can be interpreted as square correlation of dimension and optimal linear composites score. Some experts call it as maximum reliability.

Research done by Widhiarso and Mardapi (2010) expressed that multidimensional model for reliability coefficient has greater measurement accuracy compared to unidimensional reliability [19]. For that reason, researcher only focuses on internal consistent coefficient likes α for unidimensional reliability and ω , CR dan Ω_w for multidimensional reliability in this research. Some questions appear dealing with the explanation above such as: What is reliability comparison of multidimensional and unidimensional of students' attitude instrument towards statistical by using semantic differential scale? Which one of both dimensions measurement accurate more for measuring reliability?

2. Method

This research uses survey method in developing instrument uses responses approach. This research was carried out at Engineering Technique Education of Engineering Technique Education Program, Engineering Faculty, State University of Jakarta. The target population was all students of UNJ and the research population was all post graduate students of State University of Jakarta. The sample of the research was Engineering Technique Education students who passed Statistics course. This research used simple random sampling. Research instrument (questionnaires) were given to 160 students, however only 150 students returned the questionnaires.

Scale is a set of grades or numbers that given to subject, object, or behavior for quantification and quality measurement purpose. Scale is used for measuring attitudes, values, interest, motivation, and so forth. These elements relate to psychological attributes (usually affective area). For example, we can use scale for measuring someone's attitude towards statistics.

Semantic differential scale is an instrument which is used in evaluating a stimulant concept on a set of seventh steps bipolar scale from a start point to the end point in a unity compilation (Sevilla *et al.*, 1993). Pairs of adjective words are usually separated by seventh response categories which are same units along the antonym of the words continuum. The continuum direction is usually changed randomly [20]. Semantic differential scale here is a set of adjective words which refer to stimulant characteristic provided to respondents. If the adjective words have great factor weighted so it needs a complex analysis called factor analysis.

Semantic differential scale develops a way in measuring meaning of words which called Semantic differential technique. "Meaning" is a concept in semantic

which is multidimensional. This technique can be used as psychological measurement in many aspects such as personality, attitude, or communication. Besides that, this technique has special and unique characteristics if compared to other methods. One of the uniqueness is in the way respondents respond to items. Respondents are not asked to give 'agree' or 'disagree' responses. But they are asked to grade weight of stimulant through adjective words on each continuum in the scale.

Semantic differential scale can be classified into three dimensions that are evaluation (E), potential (P) and activity (A). Evaluation dimension is like good or bad, useful or useless, honest or dishonest, clean or dirty, advantage or disadvantage and so forth. Potential dimension is such as big or small, strong or weak, heavy or light. Activity dimension can be active or passive, quick or slow, and hot or cold. These three dimensions can measure three attitude dimensions that are: (a) respondents' evaluation about measured object or concept, (b) respondents' perception about object or concept potential, and (c) respondents' perception about object activity. According to Heise (1999) evaluation dimensions include nice or awful, good or bad, sweet or sour, dan helpful or unhelpful; potential dimension is like big or little, powerful or powerless, strong or weak, and deep or shallow; also activity dimension is such as fast or slow, alive or dead, noisy or quiet, and young or old [21].

Isaac dan Michael (1985) describe statistics measurement concept into: (1) evaluation (E) has 5 items, (2) potential (P) has 5 items, and (3) activity (A) has 5 items [22]. The target variable in this research is students' attitude towards statistics. It means someone's tendency towards statistics with all his/her evaluation, potential, and activity. The response of this research is typical performance responses. It is expected that respondents respond about habit or what they think a person usually does or feels when experiencing something. This way of respond is also called expression of sentiment. The expression is a response which cannot be judged as a true or false response. All responses are true based on its reason. Dealing with this characteristics response, the instrument has certain answer option range. Each item has 7 answer choices with range grade 1 to 7. Respondents have 5 to 10 minutes to answer. The answer direction tendency is positive to negative attitude towards Statistics.

3. Results and Discussion

First Try out

The instrument of attitude towards statistics has 15 items, consists of 5 items for evaluation dimension, 5 items for potential dimension, and 5 items for activity dimension. Alpha Cronbach internal consistency reliability is reached from SPSS program that is 0.925.

McDonald composite score reliability uses structural equation modeling (SEM) gained this result:

$$\sum_{i=1}^i \lambda_i = 9.080 \text{ dan } \sum_{i=1}^i 1 - \lambda_i^2 = 8.967 ; \text{ so}$$

$$\omega = \frac{(9.080)^2}{(9.080)^2 + (8.967)} = 0.902.$$

Construct reliability obtained the same result as follows:

$$\sum_{i=1}^i \lambda_i = 9.080 \text{ and } \sum_{i=1}^i \delta = 8.980 ; \text{ so}$$

$$CR = \frac{(9.080)^2}{(9.080)^2 + (8.980)} = 0.902.$$

Weighted construct reliability used SEM and produced

this result: $\sum_{i=1}^p \frac{l_i^2}{1-l_i^2} = 15.103$, so it can be counted as:

$$\Omega_w = \frac{15.103}{1+15.103} = 0.938.$$

Second Try Out

The instrument has 15 items consists of 5 items for evaluation dimension, 5 items for potential dimension, and 5 items for activity dimension. Alpha Cronbach internal consistency reliability processed through SPSS program that is 0.912.

McDonald composite score reliability uses structural equation modelling (SEM) got this result: $\sum_{i=1}^i \lambda_i = 10.190$

and $\sum_{i=1}^i 1 - \lambda_i^2 = 7.471$; so

$$\omega = \frac{(10.190)^2}{(10.190)^2 + (7.471)} = 0.932.$$

Construct reliability obtained same result as follows: $\sum_{i=1}^i \lambda_i = 10.190$ and $\sum_{i=1}^i \delta = 7.620$ so

$$CR = \frac{(10.190)^2}{(10.190)^2 + (7.620)} = 0.932.$$

Weighted reliability used SEM and produced this result:

$$\sum_{i=1}^p \frac{l_i^2}{1-l_i^2} = 34.978 , \text{ so this can be calculated as}$$

$$\Omega_w = \frac{34.978}{1+34.978} = 0.972.$$

It can be summarized as the table below:

Table 1. Summary of Reliability Coefficient

Reliability Coefficient	Unidimensional	Multidimensional		
	α	ω	CR	Ω_w
Try Out 1	0.925	0.902	0.902	0.938
Try Out 2	0.912	0.932	0.932	0.972

The value (grade) of alpha Cronbach coefficient which was achieved is smaller if compared to construct reliability, McDonald composite score reliability, and maximum reliability. The difference is 0.013 to 0,060. However, does the difference express accuracy? There is no agreement among inter psychometrics experts about this. Because the importance role of using accurate reliability measurement instrument, Indonesia researchers should try to use this tool correctly and adequately.

Most researchers from lecturers or post graduate students do not know the formula to count reliability coefficient using SEM. This is the time to introduce and use the formula. The reasons for that are we already knew the rules and most of psychological, personality, education, and social construct is multidimensional dimension. So all the researchers should try to develop and study more about reliability coefficients.

Interpretation of reliability coefficient is evaluation of test score cautiousness. It is not only about reliability itself. Two things that should be considered when interpreting level of reliability coefficient, that is: (1) coefficient reliability of certain group subjects and situation will not be the same as other group, and (2) reliability coefficient only indicates score inconsistency of measurement result. It is not for stating the causes of the inconsistency.

Education measurement is complicated. Many journals articles discuss about measurement which should give valid, reliable, and accurate result. It is not easy since it involves mathematics knowledge. We cannot understand various education measurement journals if we do not master high level and complicated mathematics applications. We left behind so far in education measurement. Not many education experts understand content of education measurement journals with great level of mathematics application. Because of that, improving education measurement program is crucial.

The first effort for such program is change our perception about mathematics. Some educators still think that education knowledge does not need mathematics much. Mathematics is only part of science and technical disciplines, not for education main discipline. However, today the perception should be changed. Educators need to realize that not all study or disciplines use mathematics application, but some are really need it, like the example above which applies multivariate statistics that requires great level mathematics application.

4. Conclusion

It can be concluded that multidimensional reliability coefficient work more correctly to estimate reliability than unidimensional reliability. There some suggestions for that: first, this instrument estimation need to be carried out in an advance study. Second, this research used scale of 7, other scales can also be applied, such as Likert scale, dichotomy scale, Thurstone scale, and so forth. Third, this instrument needs to be examined in larger population and wider setting which involve some provinces, various levels of schools and colleges. Fourth, wide use of multidimensional reliability analysis around students or researchers study may be valuable for improving accurate result of research.

References

- [1] Naga, Dali S., *Teori sektor*, Gundarma Press, Jakarta, 1992.
- [2] Suryabrata, S., *Pengembangan alat ukur psikologis*, Andi Offset, Yogyakarta, 2000.
- [3] Wiersma, W., *Research methods in education: An introduction*, Allyn and Bacon, Inc., Boston, 1986.
- [4] Gable, R. K., *Instrument Development in the Affective Domain*, Kluwer Nijhoff Publishing, Amsterdam, 1986.
- [5] Litwin, M. S., *How to Measure survey reliability and validity*, Sage Publications, London, 1995.
- [6] Thompson, B., "Guidelines for author," *Educational and psychological measurement*, 54, 837-847. 1994.
- [7] Feld, I. S., Woodruff, D. J., and Salih, F. A., "Statistical Inference for coefficient alpha," *applied psychological measurement*, II, 93-103. 1987.
- [8] Kamata, Turhan, A., A., and Darandari, E., "Estimating reliability for multidimensional composite scales scores," *Paper presented in annual meeting of American Educational Research Association di Chicago*. April 2003.
- [9] Socan, G., "Assessment of reliability when test items are not essentially t-equivalent," In *Development in survey methodology*, Anuska Feligoj and Andrej Mrvar (Eds.), FDV, Ljubljana, 2000.
- [10] Vehkalaiti, K., "Reliability of Measurement scales tarkkonnen's general method supersedes cronbach's alpha," *Academic Dissertation*, University of Helsinki, Finland. 2000.
- [11] Widhiarso, W., Koefisien Reliabilitas pada pengukuran kepribadian yang bersifat multidimensi," *Psikobuana*, 1(1), 39-48. 2009.
- [12] Spector, P., Brannick, P., and Chen, P. "When two factors don't reflect two constructs: how item characteristics can produce artifical factors," *Journal of management*, 23 (5), 659-668. 1997.
- [13] McDonald, R. P., "The Dimensionality of test and items," *British journal of mathematical and statistical psychology*, 34: 100-117. 1981.
- [14] Latan, H., *Structural equation modeling konsep dan aplikasi menggunakan program lisrel 8.80*, Alfabeta, Bandung, 2012.
- [15] Capraro, M. M., Capraro, R. M., and Herson, R. K., "Measurement error of score on the mathematics anxiety rating scale across studies," *Educational and psychological measurement*, 61: 373-386. 2001.
- [16] Geffen, D., Straub, D. W., and Boudreau. M. D., "Structural equation modeling and regression: Guidelines for research practice," *Communications of AIS*, 4, Article 7. 2000.
- [17] Lee, S. Y., and Song, X. Y., "Hypothesis testing and model comparison in two-level structural equation model," *Multivariate behavioral research*. Volume 36, Issue 4: 639-655. 2001.
- [18] Hancock, G. R., and Mueller, R. O., "Rethinking construct reliability within latent variable systems," In *Structural equation modeling: Present and future*, R. Cudek, S. H. C. duToit, dan D. F. Sorbom (Eds.), Scientific Software International, Chicago, 2000.
- [19] Widhiarso, W., and Mardapi, D., "Komparasi ketepatan estimasi koefisien reliabilitas teori skor murni klasik," *Jurnal penelitian dan evaluasi pendidikan*, 14 (1): 1-19. 2010.
- [20] Heise, D. R., "The semantic differential and attitude research," 1999. <http://www.indiana.edu/~socpsy/papers/AttMeasure/attitude.htm>. (accessed 10 November 2013).
- [21] Sevilla, C. G., Ochoa, J. A., Punsalan, T. G., Regala, B. P., dan Uriarte, G. G., *Pengantar metode penelitian*, translated by Alimuddin Tuwu, Penerbit Universitas Indonesia, Jakarta: 1993.
- [22] Isaac, S., and Michael. W. B., *Handbook in Rresearch and evaluation: For education and the behavioralssciences*, Edits Publishers, California, 1985.