ISSN 2250-3153

Structural Equation Modeling on Likert Scale Data With Transformation by Successive Interval Method and With No Transformation

Yani Quarta Mondiana1), Henny Pramoedyo, Eni Sumarminingsih

[*] Department, Institute Name
[**] Department, Institute Name, if any

DOI: 10.29322/IJSRP.8.5.2018.p7751
http://dx.doi.org/10.29322/IJSRP.8.5.2018.p7751

Abstract- Structural Equation Modeling (SEM) is a multivariate data analysis, and one of the requirements in using SEM is that the data has interval scale. Some researchers argue that Likert scale is an interval, yet many others assume that this type of data is ordinal, and therefore transformation is important to apply to uplift the measurement scale. This paper tries to identify whether there is a difference in the result of analysis between SEM analysis using Likert scale with no transformation and with transformation, by employing secondary data. The results revealed that RMSEA (Root Mean Square Error of Approximation) values without and with transformation were 0.000 and 0.000, respectively. Therefore, both are close fit. It can be stated that the results of both treatments, without and with transformation, come to the same conclusion.

Index Terms- Likert scale, transformation, Structural Equation Modeling (SEM)

I. Introduction

Many researches in economy, social, education and psychology, as well as medical field generally involve multi variables and multi relations. The variables in these fields of study are usually qualitative in nature, such as attitude, motivation, performance, commitment, satisfaction, behavior, strategy, loyalty, and so forth, and therefore, they are observable.
In order to reach a reliable result, a method of data analysis which is simultaneous and integrated is necessary. SEM (Structural Equation Modeling) is a multivariate analysis which is able to be applied in multivariable and multi relations data simultaneously, and able to test complicated relations between variables. The relationships can be built between one or more dependent variable and one or more independent variable. From each variable, factor can be built (a construct built from some indicator variables). In addition, SEM is also a measurement model which can be used to test validity and reliability of an instrument.

One of the requirements in using SEM is the data should have interval scale. In many fields of study, the data are often in the form of Likert scale. Some researchers assume that Likert scale data is interval data, yet many others assume that it is ordinal data. Joreskog (1994) states that ordinal data describes multi-level category like Likert scale. Meanwhile, Deny (2007) mentions that Likert scale can be analyzed parametrically. This is due to the fact that Likert scale can be taken as interval data, in which the range between points is the same, and therefore Likert scale should be arranged so that the data can be categorized as interval data. This is supported by Clason & Dormody (2004) stating that 5-point Likert scale can be categorized as interval scale. This paper will discuss further whether or not there is a difference between SEM analysis using questionnaire data which employs Likert scale without transformation and the one with transformation.

II. THEORETICAL FRAMEWORK

Measurement Scale

If the respondents are the data source, we need a scale that can measure an attitude becoming characteristics of a population. There are two types of scale measurement based on social phenomenon, namely a scale to measure attitude and personality (scale for morale, the result of character test, social participation) and the scale for measuring other cultural aspects and social condition. According to Riduwan and Kuncoro (2007), alongside the development of sociology and psychology, research instrument is now focusing on measurement.
Researches in social studies use some measurement scale such as Likert, Guttman, semantic differential, Rating, and Thurstone.

1. Likert Scale

This scale is used to measure psychological attitudes to be measured mathematically. Riduwan and Kuncoro (2007) explain that the Likert scale is used to measure the attitude, opinion and perception of a person or group of people about social phenomena. On the Likert scale, the variables to be measured are translated into variable indicators. Each answer is related to a question or attitude support expressed in words, for instance, for positive statements: Strongly Agree (5), Agree (4), Neutral (3), Disagree (2) and Strongly Disagree (1); or negative statements: Strongly disagree (1), Disagree (2), Neutral (3), Agree (4) and Strongly Agree (5).

The Likert scale is easy to create and implement, and there is freedom to include questions in the questionnaire, as long as it is still in the context of the problem.

2. **Guttman Scale**

This type of measurement scale has two possible firm answers of a question: 'yes' – 'no' so that it can produce interval or rational dichotomy data.

3. **Semantic Differential Scale**

This scale contains a set of bipolar characteristics (two poles) used to measure attitudes, only the form is neither multiple choice nor checklist, but is arranged in a continuous line whose "very positive" answer lies at the right of the line, and the "very negative" answer lies in the left side of the line, or vice versa. The resulting data is an interval scale.

4. **Rating Scale**

The raw data obtained in the form of numbers is then interpreted in a qualitative sense. In Rating Scale, respondents choose one of the quantitative answers that have been provided. Therefore, Rating Scale is more flexible, not limited to attitude measurement but to measure respondent's perception of other phenomena, such as scale to measure socioeconomic status, knowledge, ability and others. Rating Scale should be able to interpret every number given in the answer of each question. Answer number 2, by a certain person is not necessarily meaningful to another person who also chooses answer 2.

5. **Thurstone Scale**

Thurstone-scale data are obtained from respondents who are asked to select an approved answer from several different statements. In general, each question has an association of values between 1 to 10, but these values are unknown to the respondents. The scoring is based on the specified number of statements selected by the respondent regarding the question.

Method of Successive Interval (MSI)

Analysis of quantitative data that is often encountered in exact sciences is so far more well known. Social research uses a lot of qualitative data as a reflection of abstract concepts, or cannot be measured directly. Therefore, the analysis used is limited to descriptive analysis or non parametric analysis. Today, research in the field of social science has been developed and many quantitative analyses has been done, yet qualitative variables that produce Likert scale are classified as ordinal data (Waryanto and Millafati, 2006).

Deny (2007) suggests using Method of Succesive Interval (MSI) to transform ordinal data using Likert scale into interval data for regression analysis to be applied.

A list of questions answered with a Likert-scale approach will yield ordinal data that does not show a comparison of one answer with the other answer to the same question. In the interval data, the comparison between the actual answers will look sharper so it can be processed to obtain the value of the respondent's answer (Sukawati, 2007).

According to Riduwan and Kuncoro (2007), transforming ordinal data into interval data is useful to meet some requirements of parametric analysis. The steps of transforming ordinal data into intervals using MSI are as follows:

1. On each point of the answer, the calculated frequencies are scored 1, 2, 3, 4, and 5
2. Each frequency is divided by the number of respondents, the result is called proportion
3. The cumulative proportion value is determined by summing the proportions in sequence per score column
4. Normal Distribution Table is used to calculate the Z value of each cumulative proportion
5. The high density for each value of Z is then determined (using the High-Density table)
6. Scale value (NS) is determined with this formula:

$$NS = \frac{(\text{Density at Lower Limit}) - (\text{Density at Upper Limit})}{(\text{Area Below Upper Limit}) - (\text{Area Below Lower Limit})} \quad (2.1)$$

7. Transformation value (Y) is determined by this formula:

$$Y = NS + [\, 1 + |NS_{min}| \,] \quad (2.2)$$

in which NS_{min} is minimum scale value.

Path Analysis

Path analysis is used to determine the direct influence of a number of variables based on cross coefficients. Path analysis is not a method for finding causes, but only testing the theoretical truths that have been theorized. In the path analysis, it can be drawn conclusions about which exogenous variables have a strong influence on endogenous variables.

According to Riduwan and Kuncoro (2007), the assumption of path analysis is as the following:

1. The relationship between variables is linear and additive.
2. There is only a one-direction causal flow system, which means no direction of causality is reversed
3. Minimum response (endogenous) variable in the interval scale
4. The variables studied can be observed directly
5. The model analyzed is correctly specified based on relevant theories and concepts that means theoretical models are examined or tested based on a particular theoretical framework that is able to explain the causal relationship between the variables studied.

Simple path diagrams can be described as follows:

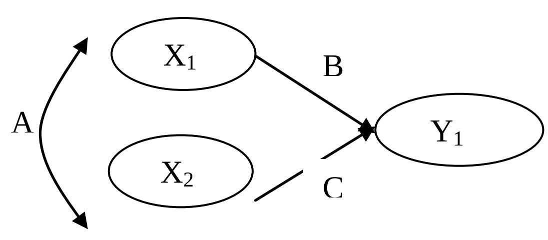

Figure 2.1 Simple Path Diagram
(Hair, Anderson, Tatham & Black, 2006)

One-way arrow ($X_1 \rightarrow Y_1$) shows the direct influence of variable X_1 to variable Y_1. While the two-way arrow between the variables X_1 and X_2 shows the existence of unanalyzed influence, which is the influence of two exogenous variables are mutually

correlated. Figure 2.1 explains that two correlated exogenous variables (X_1 and X_2) predict Y_1, and can be written in the equation:
$Y_1 = b_1 X_1 + b_2 X_2$

Path analysis allows researchers to use simple correlation coefficients between the variables involved to predict the magnitude of the causality relationships b_1 and b_2 (Hair, Anderson, Tatham & Black, 2006). The correlation coefficient between two variables X and Y (ρ) is assumed by the correlation coefficient of example *r* ,i.e.: (Walpole, 1988)

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \qquad (2.3)$$

A is the correlation between X_1 and X_2, B is the effect of X_1 in predicting Y_1 and C is the effect of X_2 in predicting Y_1.

$\rho_{x1,x2}$ = A

$\rho_{x1,x2}$ = B+AC

$\rho_{x1,x2}$ = C+AB

Confirmatory Factor Analysis

Factor analysis is one of interdependence analysis between variables. The general purpose of factor analysis is to reduce the variables, so that the information contained in the origin variable can be explained by the variable resulting from the reduction which has less number (Hair, *et.al.*, 2006).

An example of a factor analysis model is presented in Figure 2.2.

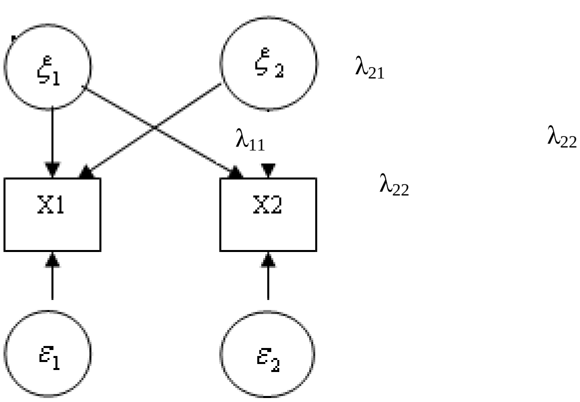

Figure 2.2 Factor Analysis Model (Sharma, 1996)

The Factor analysis model in Figure 2.2 can be written in the following equation:

$$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \varepsilon_1$$

$$x_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \varepsilon_2$$

where

ξ_p = the latent variable to -p or can also be referred to as the P-th factor
λ_{ij} = loading from -i variable to -j variable which shows the importance of j-factor in the composition of the i-th variable
ε_p = error
The Factor Analysis Model is :
$\mathbf{X_1} = c_{11}\,\mathbf{F}_1 + c_{12}\,\mathbf{F}_2 + c_{13}\,\mathbf{F}_3 + \ldots + c_{1m}\,\mathbf{F}_m + \varepsilon_1$
$\mathbf{X_2} = c_{21}\,\mathbf{F}_1 + c_{22}\,\mathbf{F}_2 + c_{23}\,\mathbf{F}_3 + \ldots + c_{2m}\,\mathbf{F}_m + \varepsilon_2$
$\mathbf{X_3} = c_{31}\,\mathbf{F}_1 + c_{32}\,\mathbf{F}_2 + c_{33}\,\mathbf{F}_3 + \ldots + c_{3m}\,\mathbf{F}_m + \varepsilon_3$
...
$\mathbf{X_p} = c_{p1}\,\mathbf{F}_1 + c_{p2}\,\mathbf{F}_2 + c_{p3}\,\mathbf{F}_3 + \ldots + c_{pm}\,\mathbf{F}_m + \varepsilon_p$

or

$$\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\\ \mathbf{X}_3\\ \ldots\\ \mathbf{X}_p\end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \ldots & c_{1m}\\ c_{21} & c_{22} & c_{23} & \ldots & c_{2m}\\ c_{31} & c_{32} & c_{33} & \ldots & c_{3m}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ c_{p1} & c_{p2} & c_{p3} & \ldots & c_{pm}\end{bmatrix} \begin{bmatrix}\mathbf{F}_1\\ \mathbf{F}_2\\ \mathbf{F}_3\\ \ldots\\ \mathbf{F}_m\end{bmatrix} \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \ldots\\ \varepsilon_p\end{bmatrix}$$

(p x 1) (p x m) (m x1)

$$\mathbf{X} = \mathbf{c}\,\mathbf{F} + \boldsymbol{\varepsilon} \qquad (2.4)$$

where,
$\mathbf{X_1}, \mathbf{X_2}, .., \mathbf{X_p}$ are origin variables
$\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_m$ are common factors
c_{ij} is the loading of the origin variable i on the j-th factor
$\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ is error

The prediction method of loading in factor analysis, among others, is the main component method. The input data can be either a correlation matrix or a covariance matrix. From the covariance matrix (S) or the correlation (R), λ_j (eigen values) and a_j (eigen vectors) are obtained. Loading factor is:

$$c_{ij} = a_{ji}\sqrt{\lambda_j} \qquad (2.5)$$

Matters relating to factor analysis (Soemarno, 2003):

1. Variety of Origin Variables (X)

Variety of X variables is divided into two components, namely communal (h_i^2) and φ_i.

$$\mathrm{Var}(X_i) = c_{i1}^2 + c_{i2}^2 + \ldots + c_{ip}^2 + \varphi_i \qquad (2.6)$$

or

$$\mathrm{Var}(X_i) = h_i^2 + \varphi_I;\ \text{where } h_i^2 = \sum_j c_{ij}^2 \qquad (2.7)$$

The h_i^2 component is called a comunality which denotes the proportion of variance X that can be explained by the common factor p. The component φ_I represents the proportion of the range of X caused by either a specific factor or an error.

The magnitude of the variety of X_i that can be explained by F_j is:

Var (X_i) that is explained by F_j = $\frac{c_{ij}^2}{\sum_j c_{ij}^2}$ x 100%
(2.8)

2. Factor of Significance

The factor that is considered significant is if the value is larger than one (λ >1) or the cumulative diversity is approximately 75%.

3. The Covariance between X and F

The covariance between Xi and Fj is as follows:

Cov (Xi, Fj) = c¬ij (2.9)

Factor loading is used to interpret every significant factor. Factor with large loading means it is the largest constituent component of a variable, while the sign (positive and negative) indicates the direction. Thus, the factor as a new variable that is unobservable can determine the origin of variable X.

4. Factor Rotation

Where significant factors are significant, it is often found that interpretation of factors as new or unobservable variables is difficult. This is due to the overlap of factors that exist as components of the compilers of variables X. To overcome this, factor rotation is applied.

5. Factor Score

Often, factor analysis is a preliminary analysis of a problem in a research, namely the effort to get a new variable or unobservable variable. Thus, the new variable must have data, which is the factor score. If the input matrix is a covariant matrix (S), the factor score is calculated by the formula:

S-Fa = c'S-1 (xj -) (2.10)

However, if the input matrix is the correlation matrix (R), the factor score is calculated by the formula:

S-Fa = c'R-1Zj (2.11)

According to Sharma (1996), in the confirmatory factor analysis, the structure of the factor model has been underlied by a theory. Therefore, the number of factors that are formed is already known in advance. This is in contrast to exploratory factor analysis, where the previous researcher did not have a theory or hypothesis that made up the factor structure. Confirmatory factor analysis is a continuation of exploratory factor analysis. In this case, once a researcher finds new variables resulting from exploration of variables owned before, the researcher need to confirm the new variables to check the validity and reliability. The factor as a new variable resulting from exploratory factor analysis process is unobservable, and it is often called latent variable. This factor cannot be observed directly by researchers because it is a collection of several sizes or observations.

Simultaneous Equation Model

The simultaneous equation model is a model containing more than one dependent variable and more than one equation. This model is useful for predictions as in regression. The typical feature of the simultaneous equation model is that the dependent variable in an equation may appear as a variable explaining in other equations in the model. In this model, a number of equations form a system of equations which describes the dependence among the variables in the equations. Before completing the simultaneous equation model, the equations contained in the model must be shown first that they have satisfied the proper identification conditions.

The identification for the function of a simultaneous model is (Imam, 2000):

1. if K – k > m – 1, the function is overidentified.
2. if K – k = m – 1, the function is just identified.
3. if K – k < m – 1, the function is underidentified.

where *m* is the number of endogenous variables in a particular single function

K is the number of exogenous variables in the simultaneous model

k is the number of exogenous variables in a particular single function.

Structural Equation Modeling (SEM)

SEM is a statistical technique used to construct and test a causal model. SEM is a technique that includes confirmatory aspects of factor analysis, path analysis and regression that can be considered a special case in SEM. From the definition. it can be said that SEM has characteristics that are as analytical techniques to be more confirm than explanation, meaning that SEM is more suitable to be used to determine the validity of a model than to use it to find a model fit, although SEM analysis also includes elements for explanation. The most critical error in model development is the existence of specification error, that is when one or more predictor variables is not involved.

There are two models in SEM that are structural model and measurement model. The structural equation is formulated as a means to express the relationship of mutuality between constructs with the following guidelines:

Endogenous Variables = Exogenous Variables + Endogenous Variables + Error. The general structural modeling equation can be written as follows (Hayduk, 1987 in Wijayanto, 2008):

$$\eta_1 = \beta_2\eta_2 + \beta_3\eta_3 + ... + \beta_m\eta_m + \gamma_1\xi_1 + \gamma_2\xi_2 + ... + \gamma_n\xi_n + \zeta_1$$

$$\eta_2 = \beta_1\eta_1 + \beta_3\eta_3 + ... + \beta_m\eta_m + \gamma_1\xi_1 + \gamma_2\xi_2 + ... + \gamma_n\xi_n + \zeta_2$$

.
.
.

$$\eta_m = \beta_1\eta_1 + \beta_3\eta_3 + ... + \beta_{m-1}\eta_{m-1} + \gamma_1\xi_1 + \gamma_2\xi_2 + ... + \gamma_n\xi_n + \zeta_m$$

Or (in matrix):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$
(2.12)

where:

$\boldsymbol{\eta}$: eta, matrix (sized m x 1) endogenous latent variable (dependent)

B: beta, matrix (sized m x m) coefficient of the influence of endogenous variable toward other endogenous variables

$\mathbf{\Gamma}$: gamma, matrix (sized m x n) coefficient of the influence of exogenous variable toward endogenous variable

ζ : zeta, matrix (sized m x 1) structural error

ξ : matrix (sized n x 1) exogenous latent variable

Measurement model for exogenous variable can generally be written as:

$$x_1 = \lambda_1 \xi_1 + \delta_1$$

$$x_2 = \lambda_2 \xi_2 + \delta_2$$

.
.
.

$$x_q = \lambda_q \xi_q + \delta_q$$

Or in matrix it is written:

$$\mathbf{x} = \mathbf{\Lambda_x \xi} + \boldsymbol{\delta}$$

(2.13)

where:

$\mathbf{x}$: matrix (sized q x 1) indicator of exogenous variable

$\mathbf{\Lambda_x}$: matrix (sized q x n) loading factor of exogenous variable

$\boldsymbol{\delta}$: matrix (sized q x 1) measurement error

Measurement model for endogenous variable can generally be written as: (Hayduk, 1987 in Wijayanto, 2008):

$$y_1 = \lambda_1 \eta_1 + \varepsilon_1$$

$$y_2 = \lambda_2 \xi_2 + \varepsilon_2$$

.
.
.

$$y_p = \lambda_p \xi_p + \varepsilon_p$$

Or in matrix it is written:

$$\mathbf{y} = \mathbf{\Lambda_y \eta} + \boldsymbol{\varepsilon}$$

(2.14)

where:

$\mathbf{y}$: matrix (sized p x 1) indicator of endogenous variable

$\mathbf{\Lambda_y}$: matrix (sized p x m) loading factor of endogenous variable

$\boldsymbol{\varepsilon}$: matrix (sized p x 1) measurement error for endogenous variable

The three stages of SEM are: a) examining of the validity and reliability of the instrument (confirmatory factor analysis), b) testing the relationship model among the latent variables to determine the determinant factor (path analysis), c) the acquisition of a useful model for prediction equivalent to the structural model or regression (Sarwono & Narimawati, 2007).

The terms of SEM are as the following:

1. Large sample size

Sample size plays an important role in the estimation and interpretation of SEM analysis results. There is no single criterion explaining how many sample sizes are required in SEM. However, for estimation using maximum likelihood, the recommended sample size ranges from 100-200. If the sample size exceeds 400, the possibility of goodness of fit will indicate a model mismatch.

2. Continuous measurement scale (interval)

The scale of measurement of variables in SEM analysis is the most controversial and much debated. This controversy arises because the treatment of ordinal variables is considered a continuous variable. Generally, the measurement of indicators of a latent variable uses 5-point Likert scale, namely strongly disagree, disagree, neutral, and strongly agree, which actually is the ordinal scale (rank). Many researchers change this ordinal Likert scale into an interval scale with successive interval (MSI) methods.

The assumptions in SEM are as follows (Hair, *et. al.*, 2006):

1. All relationships are linear

Examining the linearity of relationship can be done with the Curve Fit approach and applying the principle of parsimony, i.e. when all models are significant or non-significant, the model chosen is the simplest model that is linear (Ljung, 2003).

2. Normality

Basically the normality assumption for using SEM analysis is not very critical when the observation data reaches 100 or more because based on the Central Limit Theorm of a large sample size can be generated average samples close to normal distribution. (Mendenhall, *et.al.*, 1981).

3. Data does not contain outliers

Univariate outliers and multivariate outliers must be examined. For univariate outliers, observations with z-score $\geq$ 3.0 will be categorized as outliers, and for large samples above 80 observations, evaluation guidelines are the threshold values of z-scores ranging from 3 to 4 (Hair, *et.al.*, 2006). As for multivariate outliers, it can be detected by the distance of Mahalanobis. The Mahalanobis distance between the i and j-individuals is expressed by equations (Senior, 2000):

$$d_{ij}^2 = (\mathbf{\bar{x}_i} - \mathbf{\bar{x}_j})'\mathbf{S}^{-1}(\mathbf{\bar{x}_i} - \mathbf{\bar{x}_j})$$

(2.15)

d_{ij}^2 : Mahalanobis distance between the i- and j- individuals

$\mathbf{\bar{x}}_i$: average vectors of i- individual observation

$\mathbf{\bar{x}_j}$: average vectors of j- individual observation

$\mathbf{S}^{-1}$: inverse of covariance matrix

The test of outliers is done by looking at the value from Mahalanobis distance, with the hypothesis:

H_o: there is no outliers

H_1: there are outliers

If $d_{ij}^2 < \chi^2_{p(\alpha)}$, where p is the number of indicators, then it is concluded that there is no outliers.

According to Bollen and Long (in Wijayanto, 2008), SEM modeling is made through several stages:

1. Model specification
2. Identification
3. Assumption
4. Evaluation
5. Re-specification

One of the evaluation criteria model is RMSEA (Root Mean Square Error Approximation), which can be calculated by the formula 2.16 for interval data and 2.17 for ordinal data.

RMSEA for ordinal data

$$\text{RMSEA} = \sqrt{\frac{F(S,\Sigma(\theta))}{db} - \frac{1}{N-1}} \tag{2.17}$$

RMSEA for interval data

$$\text{RMSEA} = \sqrt{\frac{\chi^2 - db}{(N-1)db}} \tag{2.16}$$

where $\chi^2 = (N-1)F(S,\Sigma(\theta))$

III. RESEARCH METHODOLOGY

The data used in this study were secondary data from the research result of a student of Faculty of Agricultural Technology Universitas Brawijaya entitled "Analysis of the Relationship between Productivity and Quality to Internal Company Business Process with Structural Equation Modeling (SEM) Method" (Pramudiya, 2006). The variables in this study are:

Latent Variable	Indicator
1. Productivity	X11. Workers' productivity X12. Extra value X13. Operation level X14. Machine error
2. Quality	X21. Process error X22. Product error X23. Customers' claim
3. The Process of Company's Internal Business	Y1. Innovation process Y2. Operation process

Method

The stages in this research are:

1. Transforming data by MSI method
2. Applying SEM (Structural Equation Modeling) on data without transformation and with transformation with the following steps:
   a. Estimating model parameter with maximum likelihood method using Equation 2.16
   b. Evaluating the whole model through the Root Mean Square Error of Approximation (RMSEA) criterion, using Equation 2.16 for interval data and equation 2.17 for ordinal data
   c. Comparing the model of SEM analysis results from data without transformation and data with transformation using the RMSEA goodness of fit criteria

IV. RESULTS AND DISCUSSION

The comparison of two or more estimation models can be performed if the quantities used as comparators are the same. In order to compare the two models, the RMSEA criterion is used in this study. Therefore, in the first step it is necessary to prove that the RMSEA of the model generated from ordinal data is the same or equivalent to RMSEA for the model generated from interval data.

Analysis of RMSEA

According to Joreskog (1994), RMSEA for ordinal data is

$$\sqrt{\frac{\chi^2 - db}{(N-1)db}}$$

Meanwhile, RMSEA for interval data is (Scermelleh & Muller, 2003):

$$\sqrt{\frac{F(S,\Sigma(\theta))}{db} - \frac{1}{N-1}}$$

Substitution $\chi^2 = (N-1)F(S,\Sigma(\theta))$ which is one of the statistical analysis of the goodness of fit into RMSEA as a model of ordinal data will result in:

$$\text{RMSEA} = \sqrt{\frac{(N-1)F(S,\Sigma(\theta)) - db}{(N-1)db}} = \sqrt{\frac{F(S,\Sigma(\theta))}{db} - \frac{1}{N-1}}$$

which is also the formula for RMSEA for interval data. Therefore, RMSEA can be used as a comparison factor for the obtained model from ordinal and interval data.

The Results of Parameter Estimates and RMSEA on Data With and Without Transformation

The summary of parameter estimation results in data without transformation is presented in the following table.

Table 4.5 The Results of Parameter Estimates on Data 2 Without Transformation with Standardized Loading Factor

Loading Factor	Estimates	p

	value	value
Business process_internal <- Productivity	0.453	0.213
Business process_internal <--- Quality	0.376	0.233
x14 <----------------- Productivity	0.180	
x13 <----------------- Productivity	0.681	0.239
x12 <----------------- Productivity	0.520	0.194
x11 <----------------- Productivity	0.242	0.324
x23 <---------------------- Quality	0.280	
x22 <---------------------- Quality	0.189	0.332
x21 <---------------------- Quality	0.572	0.415
y1 <--------- Business process_internal	0.696	
y2 <--------- Business process_internal	0.548	0.030

The result of conversion of line diagram into structural model is

$$\eta_1 = 0.453\xi_1 + 0.213\xi_2 + 0.127$$

in which η_1 : internal business process

ξ_1 : productivity

ξ_2 : quality

The biggest contribution to internal business processes is given by productivity variables. This can be seen from the value of the standardized coefficients for productivity variables is the largest, which is 0.45 (Table 4.5). Both productivity and quality variables contribute positively to internal business processes. The model analyzed is a recursive model (there is no reciprocal relationship). From the analysis, it was obtained $\chi^2_{hitung} = 21.257$ and p-value = 0.678, meaning that at the error rate of 5%, the hypothesis stating that the model is in accordance with empirical data, is received. RMSEA value obtained for data without transformation is 0.000 (Figure 4.5), which is included in the close fit category.

The result of parameter estimation for data with transformation is presented in Table 4.7

Table 4.7 The Results of Parameter Estimates on Data 2 With Transformation with Standardized Loading Factor

Loading Factor	Estimates value	p value
Business process_internal <- Productivity	0.450	0.352
Business process_internal <--- Quality	0.374	0.242
x14 <----------------- Productivity	0.134	
x13 <----------------- Productivity	0.659	0.388
x12 <----------------- Productivity	0.446	0.350
x11 <----------------- Productivity	0.311	0.419
x23 <--------------------- - Quality	0.281	
x22 <--------------------- - Quality	0.186	0.348
x21 <--------------------- - Quality	0.568	0.431
y1 <--------- Business process_internal	0.684	
y2 <--------- Business process_internal	0.546	0.038

The result of conversion of line diagram into structural model is

$$\eta_1 = 0.450\xi_1 + 0.374\xi_2 + 0.239$$

From the overall model evaluation, it was obtained RMSEA for Data 2 With Transformation of 0.000; and because the RMSEA is less than 0.05, it is included into the close fit category.

Comparison of RMSEA Value

The statistical value of RMSEA for each data, with and without transformation can be seen in table 4.9.

Tabel 4.9 RMSEA for each data

Without transformation	With transformation
0.000	0.000

Table 4.9 shows that RMSEA value for the data without and with transformation is included into close fit data, in which both shows RMSEA<0.05.

V. CONCLUSION

Based on the results of this study, it can be concluded that there is no difference in the SEM analysis result in the questionnaire data in Likert scale with and without transformation. This is shown by the RMSEA value in both data which come into the same conclusion in testing model suitability.

REFERENCES

[1] Clason, D.L. & T.J. Dormody . 2004. Analyzing Data Measured by Individual Likert-Type items. http://pubs.aged.tamu.edu/jae/pdf/Vol35/35-04-31.pdf. Tanggal akses : 20 Januari 2009.

[2] Deny, K. 2007. Transformasi Data Ordinal ke Interval dengan Method of Successive Interval (MSI). http://jurnal-sdm.blogspot.com/2007/12/transpormasi-data-ordinal-ke-interval.html. Tanggal akses : 1 Oktober 2008

[3] Hair, J.F., R.F. Anderson, R.L. Tatham dan W.C. Black. 2006. Multivariate Data Analysis. 5th edition. Prentice Hall Inc., New Jersey.

[4] Imam, K. 2000. Simultaneous Equation Models.

[5] http://elearning.unej.ac.id/courses/MAA01/document/SIMULTANEOUS_EQUATION_MODELS.pdf?cidReq=MAA01. Tanggal akses : 15 Januari 2009.

[6] Joreskog, K.G. 1994. Structural Equation Modeling with Ordinal Variables. http://www.samsi.info/200405/socsci/opening/karl.joreskog.pdf. Tanggal akses : 17 Oktober 2008.

[7] Ljung, L. 2003. Linear System Identification as Curve Fitting. http://www.springerlink.com/content/t40uqtmh2a6v62u5/. Tanggal akses : 20 Desember 2008.

[8] Mendenhall, W., R.L. Scheaffer, D.D Wackerly. 1981. Mathematical Statistics With Applications. 2nd edition. Wadsworth Inc, California.

[9] Pramudiya, A. 2006. Analisis Hubungan Produktivitas dan Kualitas Terhadap Proses Bisnis Internal Perusahaan dengan Metode Structural Equation Modeling (SEM) "Studi Kasus di Perusahaan Rokok CV. Cempaka Tulungagung". Skripsi. Universitas Brawijaya, Malang.

[10] Riduwan & E.A. Kuncoro . 2007. Cara Menggunakan dan Memaknai Analisis Jalur (Path Analysis). Alfabeta, Bandung.

[11] Sarwono, J & U, Narimawati. 2007. Structural Equation Model (SEM) Dalam Riset Ekonomi " Menggunakan LISREL". Gava Media, Yogyakarta.

[12] Scermelleh, K & H. Muller. 2003. Evaluating the Fit of Structural Equation Models : Test of Significance and Descriptive Goodness of Fit Measures. http://user.uni-frankfurt.de/~moosbrug/schermelleh/mpr_Schermelleh.pdf. Tanggal akses : 17 Oktober 2008.

[13] Senior, G. 2000. Analysing RAVLT Learning and Serial Position Curves Using Mahalanobis Distance. http://www.usq.edu.au/users/senior/Assessment/RAVLT-Mahalanobis-Analysis.htm. Tanggal akses : 27 Oktober 2008.

[14] Sharma, S. 1996. Applied Multivariate Techniques. John Wiley & Sons Inc, Canada.

[15] Soemarno. 2003. Analisis Faktor. http://images.soemarno.multiply.com/attachment/0/RftLWAoKCpkAAA7KtoI1/Analisis%20Faktor.ppt?nmid=22306021. Tanggal akses : 19 Januari 2009

[16] Sukawati, 2007. Pengaruh Strategi Diferensiasi Terhadap Kepuasan Pelanggan Pita Maha A Tjampuhan Resort & Spa di Ubud. Http://ejournal.unud.ac.id/abstrak/cok%20sukawati.pdf. Tanggal akses: 8 Oktober 2008

[17] Walpole, R. Alih bahasa Bambang Sumantri. 1988. Pengantar Statistika. Edisi ketiga.PT. Gramedia, Jakarta.

[18] Waryanto, B. & Y.A. Millafati. 2006. Transformasi Data Skala Ordinal ke Interval dengan Menggunakan Makro Minitab. http://www.litbang.deptan.go.id/warta-ip/pdf file/4.budiwaryantoipvol-15.pdf. Akses : 1 Oktober 2008

[19] Wijayanto, S.H. 2008. Structural Equation Modeling dengan Lisrel 8.8. Graha Ilmu, Yogyakarta.

AUTHORS

First Author – Yani Quarta Mondiana, yqmondiana@gmail.com
Second Author – Henny Pramoedyo
Third Author – Eni Sumarminingsih