Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2441-2449

http://www.ripublication.com

Generalized Structured Component Analysis (GSCA) For National Education Standards (NES) Of Secondary School In Indonesia

[1)]Debrina Vita, [2)]Budi Susetyo,[3)]Bambang Indriyanto

[1,2)]Departemen of Statistics, Bogor Agricultural University,
[3)]Center for Research and Policy, Ministry of education and culture Indonesia
Departemen of Statistics
Bogor Agricultural University
Jalan Pajajaran, Kampus IPB Baranangsiang, Bogor 1615, Indonesia
[1)]debrinaferezagiaa@gmail.com, [2)]buset008@yahoo.com, [3)]indri_diknas@yahoo.com

Abstract

National Education Standards (NES) is the minimum criteria of the national education system in Indonesia. Generalized Structured Component Analysis (GSCA) is Structural Equation Modeling (SEM) based on components and a powerful analytical method. The advantages SEM-GSCA is not based parametric assumptions, avoiding factor inderterminacy and inadmisable solution. GSCA gives optimum solution and can provide mechanism to evaluate overall goodness-fit of the model. The objective of this research is the modeling of eight standards NES with GSCA. The data used is secondary data, Indonesian high school accreditation. Results from this study is not all variables indicator valid and reliable. SPT latent variables path coefficient of the SPN is not significant and the coefficient is very small parameter (0.002). SKL are influenced by all the other latent variables either directly or indirectly. SPT, SI, SPR, SSR indirect effect on SKL through SPN. The most indirect influence of SKL is SPR through SPN (0.205). All variables are valid and reliable indicators to measure latent variables after modification. FIT = 46.7%, the diversity of all variables can be explained by the model.

Key Words: Structural Equation Modeling (SEM), Generalized Structured Component Analysis (GSCA), indicator variables, latent variables

1. Introduction

National Education Standard (NES) is the minimum criteria of the Indonesia education system and must be implemented by all education units. NES serve as a basis for planning, implementation and monitoring education in order to realize the quality of education in Indonesia. NES consists of 8 standards ie Competency Standards (SKL), Content Standards (SI), Processing Standards (SPR), Staff Education Standards (SPT), Infrastructure Standard (SSP), Standard Management (SPL), Standard Funding (SB) and Assessment Standards (SPN). The eight standards is called laten variables, because it can not be measured directly. Latent variables can be measured by the indicator/ manifest variables. The relations between laten variables and relationships between latent variables with indicator variables can be analyzed by Structural Equation Modelling (SEM).

SEM was first develop by Joreskog (1973), that called Covariance-Based SEM (CBSEM) using likelihood. CBSEM influenced by the parametric assumptions, is all variables have a multivariate normal distribution, the observations should be independent of each other and the sample size should be large. The minimum recommended range is 5-10 times the variables used (Ghozali, 2008). The small sample size in CBSEM can not give representative parameter and good model (Cou and Betler, 1985). It can also produce a negative variance (often called Heywood case).

SEM can be approximated by variance/ component-based, which is Partial Least Square (PLS) and Generelized Structured Component Analysis (GSCA). SEM-PLS and GSCA avoid problems in admisable solution and inderterminacy factor (Fornell and Bookstein, 1982). SEM-PLS has limitations in estimating the parameters, because it has no global optimization. As a result, there is no guarantee that the PLS will provide optimal solutions and difficult to determine the overall goodness-fit of the model (Hwang H and Tanake, 2004).

GSCA was recently proposed as a *component-based* approach to SEM. GSCA can be use for substantive researchers and practitioners in various scientific disciplines for many reasons. First, GSCA does not require the assumption of multivariate normality of observed variables, which is usually violated in practice. Second, it is shown to perform well in small samples, which appear practically inevitable in many studies. Third, GSCA does not yield inadmissible solutions such as negative variance estimates and correlations greater than one in absolute value. Inadmissible solutions are often difficult to circumvent effectively and obscure the interpretation of results. Fourth, this approach results in unique estimates of latent variable scores which can be used for various subsequent analyses. Finally, it enables the provision of overall model-fit measures for theory testing and model comparison (Hwang and Takane, 2009). Based on the background of the problem, the objective of this study is to apply GSCA for NES modelling in secondary school.

2. Materials and Mathods

2.1 Matrials

The data ued in this reseach is secondary data of secondary school accreditation in 2014. The data was obtained from the reasearch and development of Indonesia's ministry education and culture. The sample size is 1000 secondary school, that collected using random proportional sampling. The variables consist of 8 latent variables and 165 indicators variables. There are SB: 24 indicators variables, SPL: 20 indicators variables, SPT: 20 indicators variables, SSP: 30 indicators variables, SI: 18 indicators variables, SPR: 9 indicators variables, SPN: 19 indicators variables and SKL: 25 indicators variables. Based on the education theory, direct and indirect relationships between the latent variables shown in the following path diagram.

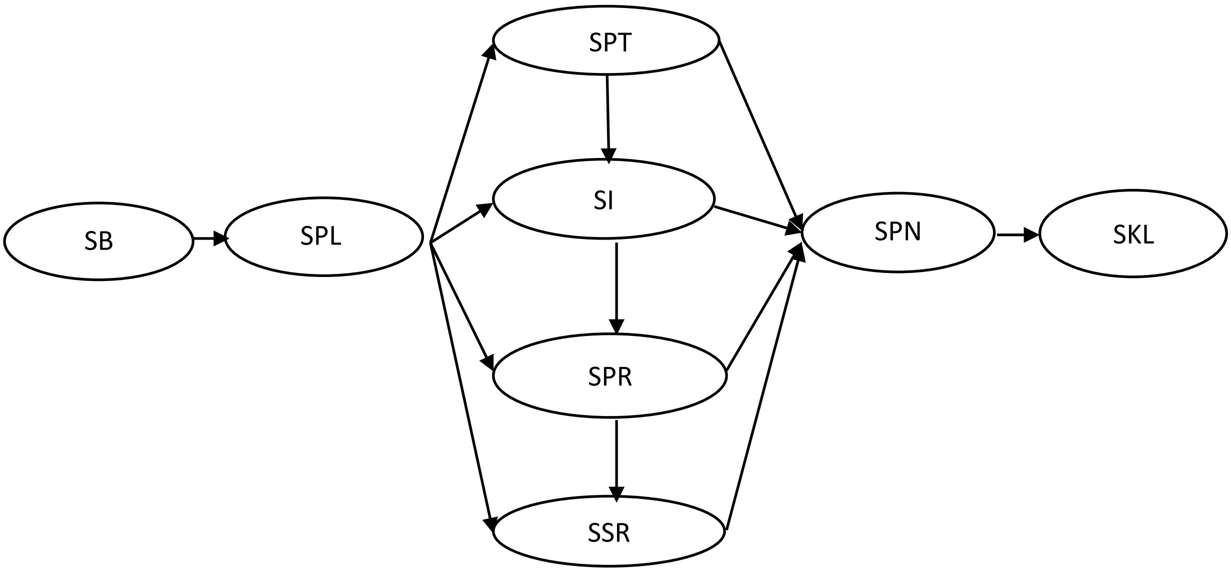


Figure 1. Path diagram of 8 Laten Variables

2.2 Method

GSCA model combine three equations into a single one. There are weighted composited of observed variables, measurement model and structural model. GSCA models as follows:

$$\begin{bmatrix} Z \\ \gamma \end{bmatrix} = \gamma \begin{bmatrix} C \\ B \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \xi \end{bmatrix}$$

$$Z \begin{bmatrix} I \\ W \end{bmatrix} = ZW \begin{bmatrix} C \\ B \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \xi \end{bmatrix}$$

if I is identity matrix, $V = [I, W]$, $A = [C, B]$, $E = [\varepsilon, \xi]$ so:

ZV=ZWA+E

Ψ=τA + E

To estimate the unknown parameters W and A, the following least-squares criterion is minimized:

f = trace((ZV-ZWA)' (ZV-ZWA))

f = trace((Ψ-τA)'(Ψ-τA))

An Alternating Least Squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) was developed to minimize that criterion.

Goodness of fit GSCA models are divided into Goodness of measurement model, Goodness of Structural model and Goodness of overall models. The evaluation of the measurement model by looking at convergent validity, discriminant validity, and the Average Variance Extracted (AVE) (Fornell and Lacker, 1981). The evaluation of structural models to see the path coefficient from endogenous to exogenous variables and see the values of significance. Structural model was evaluated by looking at the value of the coefficient parameters and statistical values T as well as the significance of the parameter coefficients. The T statistics obtained from the bootstrapping by dividing the coefficient parameters by standard error. The R^2 measure diversity of endogen construct that can be explained by diversity of exogenous constructs.

GSCA currently provides two measures of overall model fit – FIT (Hwang & Takane, 2004) and AFIT (Hwang, DeSarbo, & Takane, 2007). The FIT indicates the total variance of all endogenous variables explained by a model specification. It is given by FIT $=1-[SS(\mathbf{ZV}-\mathbf{ZWA}) / SS(\mathbf{ZV})]$. The values of FIT range from 0 to 1. The larger this value, the more variance in the variables is accounted for by the specified model. The AFIT was developed to take model complexity into account. It is given by:

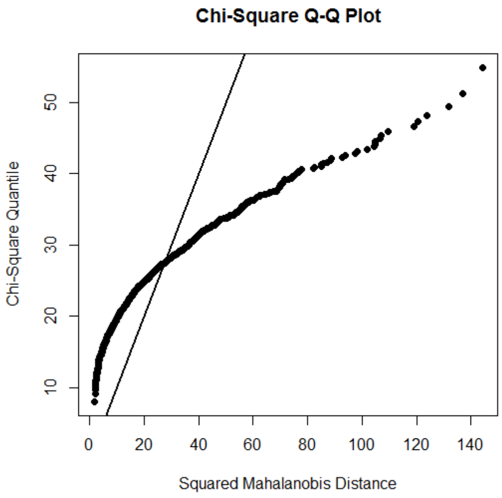
$$AFIT = 1 - (1 - FIT)\frac{d_0}{d_1}$$

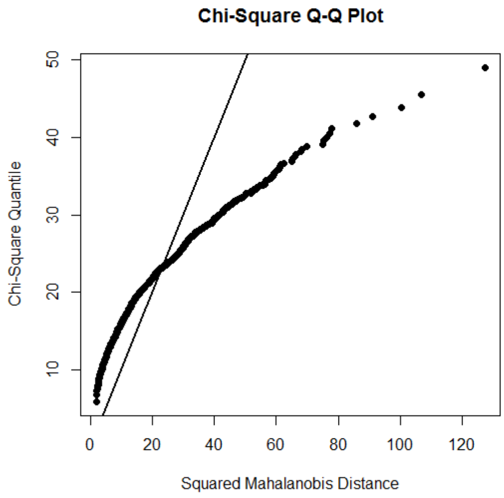
Where $d_0 = NJ$ is the degrees of freedom for the null model ($\mathbf{W} = \mathbf{0}$ and $\mathbf{A} = \mathbf{0}$) and $d_1 = NJ - G$ is the degrees of freedom for the model being tested, where G is the number of free parameters.

3. Result and Discussion

3.1 Normal multivariate

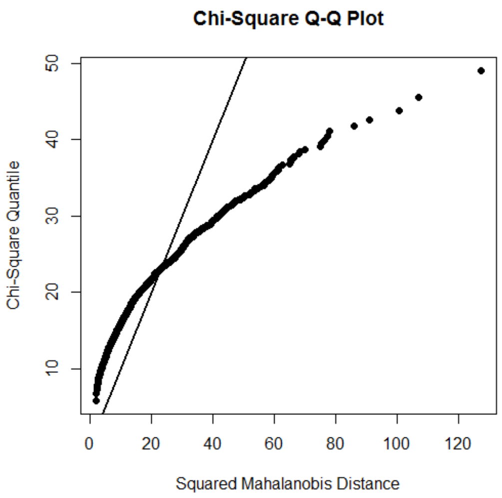
Multivariate Normal testing by exploration using plot quantile-quantile and Zirkler Henze test. Multivariate normal testing conducted in each laten variables. Based on the results, can be shown in figure 2. The p value $< \alpha$ (0.01) for the eight latent variables, H_0 is rejected, so it can be concluded that the data did not distributed multivariate normal. CBSEM-likelihood is not appropriate if the data is not distributed normal multivariate (Chou and Bentler, 1985). To see the relationship between latent variables with the indicator and the relationship between latent variables, the approach used in this study using SEM-GSCA.

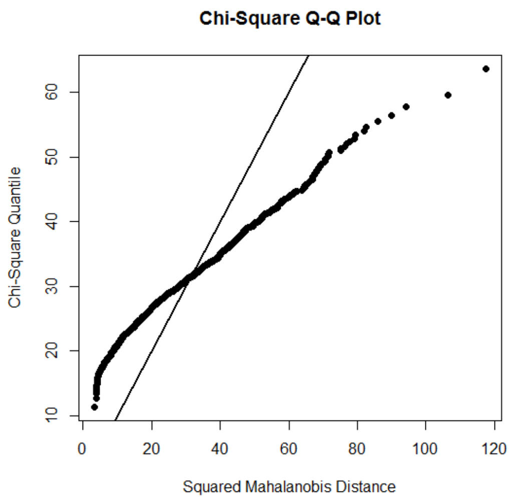




SB HZ: 3.502355 p-value: 0

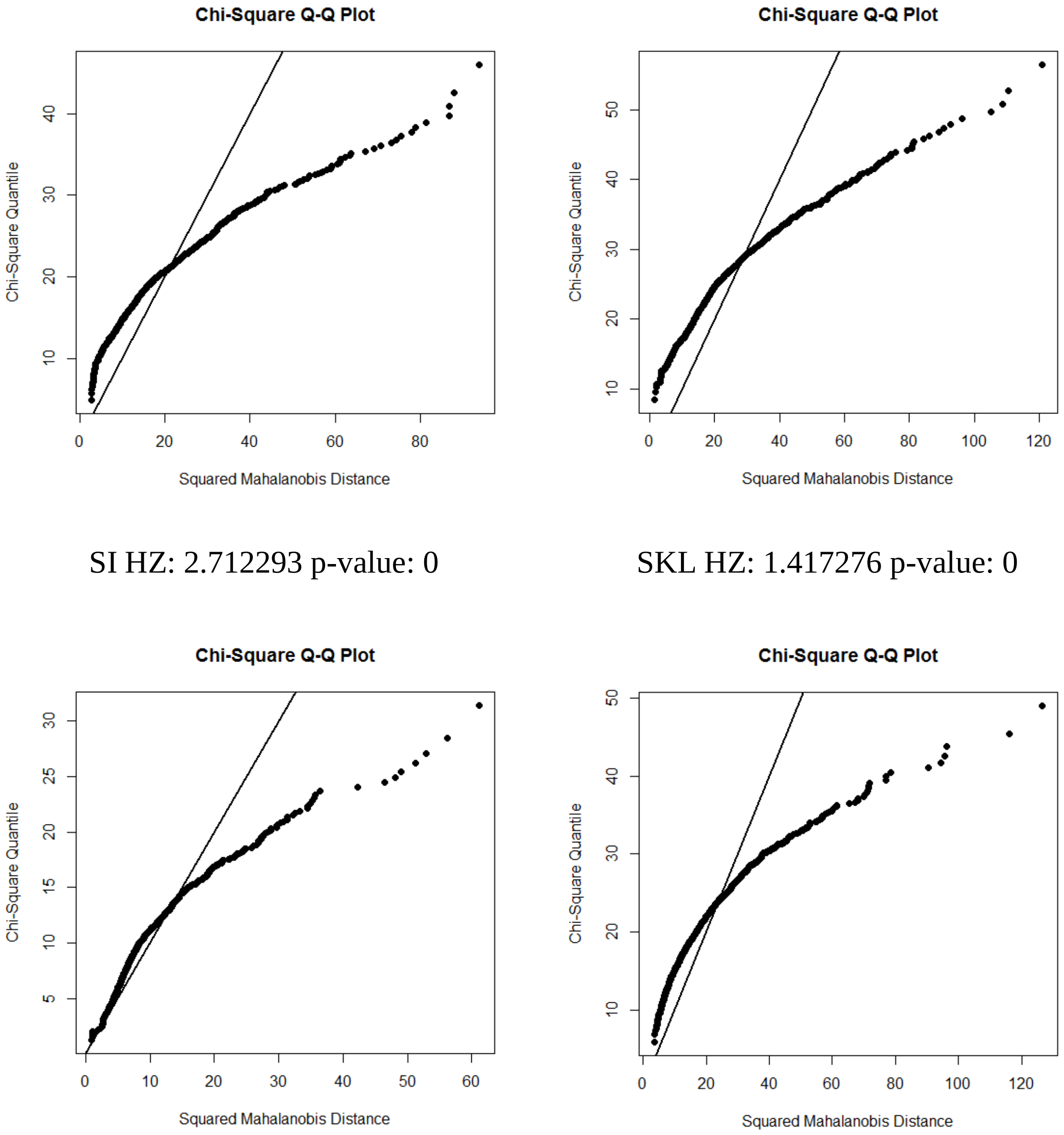
SPL HZ: 2.305078 p-value: 0





SPN HZ: 2.557918 p-value: 0

SSP HZ: 1.697766 p-value: 0



SI HZ: 2.712293 p-value: 0

SKL HZ: 1.417276 p-value: 0

SPR HZ: 2.545809 p-value: 0

SPT HZ: 1.589162 p-value: 0

Figure 2. Testing of Normal multivariate for eight laten variables

3.2 Measurement Models

Evaluation of the measurement model by looking at the loading factor of indicators to latens variables. The significant test and loading factor > 0.6 are a good indicator to measure latent variables (Chin, 1998). Indicator variables that have a loading factor < 0.6 were excluded from further analysis. Based on testing, the indicator variables with high loading for eight latent are SI: 8 indicators, SPR: 8 indicators, SKL: 21 indicators, SPT: 3 indicators, SSP: 13 indicators, SPL: 10 indicators, SB: 9 indicators and SPN: 7 indicators. Indicators that have a high loading factor to be analyzed again. Evaluation measurement model was repeated with three criteria: convergent validity, discriminant validity and reliability.

Reliability testing of latent variables using Conbrach Alpha, with a values in each laten is SI = 0836, SPR = 0826, SKL = 0861, SPT = 0747, SSP = 0904, SPL = 0861, SB 0876 and SPN = 0.8601. According to the results, Conbrach Alpha on each laten variables >0.7. This show that the reliability of each laten variables is good.

The evaluation measuremnet model can also be analyzed by discriminat validity. The value of square root of the AVE is compared with correlation laten variables. The value of the square root of the AVE is SB 0711, SPL 0671, SPT 0.82, SI 0686, SPR 0672, SSP 0692, SPN 0742 and SKL 0654. Those values are greater than the value of the correlation between latent variables concerned with other latent variables (can be seen in Table 1). As a result, having a good discriminant validity. So that, it can be concluded that all indicator variables are valid and reliable to measure latent variables.

Table 1. Correlation of laten variables

	SB	SPL	SPT	SI	SPR	SSP	SPN	SKL
SB	1	0.547 (0.023)*	0.327 (0.026)*	0.495 (0.027)*	0.499 (0.032)*	0.453 (0.031)*	0.486 (0.033)*	0.494 (0.026)*
SPL	0.547 (0.023)*	1	0.503 (0.022)*	0.721 (0.016)*	0.714 (0.020)*	0.693 (0.015)*	0.628 (0.023)*	0.743 (0.014)*
SPT	0.327 (0.026)*	0.503 (0.022)*	1	0.562 (0.019)*	0.472 (0.024)*	0.607 (0.019)*	0.401 (0.026)*	0.582 (0.019)*
SI	0.495 (0.027)*	0.721 (0.016)*	0.562 (0.019)*	1	0.763 (0.016)*	0.676 (0.017)*	0.619 (0.022)*	0.779 (0.012)*
SPR	0.499 (0.032)*	0.714 (0.020)*	0.472 (0.024)*	0.763 (0.016)*	1	0.631 (0.020)*	0.625 (0.021)*	0.760 (0.014)*
SSP	0.453 (0.031)*	0.693 (0.015)*	0.607 (0.019)*	0.676 (0.017)*	0.631 (0.020)*	1	0.546 (0.024)*	0.743 (0.015)*
SPN	0.486 (0.033)*	0.628 (0.023)*	0.401 (0.026)*	0.619 (0.022)*	0.625 (0.021)*	0.546 (0.024)*	1	0.641 (0.018)*
SKL	0.494 (0.026)*	0.743 (0.014)*	0.582 (0.019)*	0.779 (0.012)*	0.760 (0.014)*	0.743 (0.015)*	0.641 (0.018)*	1

3.3 Structural Model

Table 1 shows the correlation between laten variables. A strong correlation was shown on two variables that have a significant correlation and more than 0.5. And then a very strong correlation indicated by correlation value over 0.7. Whereas, weak correlation occours between latent variables SPT with SB, SB with SI, SB with SPR, SSP with SB, SB with SPN, SB with SKL, and SPT with SPN.

By using GSCA, the results are not all path coefficients significant. SPT latent variables path coefficient to SPN is not significant and the coefficient is very small parameter (0.002). The latent variables SPN influenced by SPT at 0.002. In other words, variable SPT has a relationship with SPN but no significant effect on α = 10%. The model is modified by removing the path of SPN and SPT. Thus, the structural models obtained are:

$$\begin{bmatrix} SPL \\ SPT \\ SI \\ SPR \\ SSR \\ SPN \\ SKL \end{bmatrix} = \begin{bmatrix} 0.547 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.503 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.586 & 0.267 & 0 & 0 & 0 & 0 \\ 0 & 0.341 & 0 & 0.517 & 0 & 0 & 0 \\ 0 & 0.495 & 0 & 0 & 0.277 & 0 & 0 \\ 0 & 0 & 0 & 0.262 & 0.319 & 0.168 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.641 \end{bmatrix} \begin{bmatrix} SB \\ SPL \\ \mathrm{SPT} \\ SI \\ SPR \\ SSR \\ SPN \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \\ \xi_7 \end{bmatrix}$$

Based on Figure 3, the largest coefficient is the relationship between SPN and SKL. The latent variables SPN give positive effect to SKL for 0.641.The smallest effect given by laten variables SSP with SPN (0.1682). All of path have postive coefficient. SKL are latent variables that are influenced by all the other latent variables either directly or indirectly. SPT, SI, SPR, SSR indirect effect on SKL through SPN. The biggest indirect influence is SKL given by SPR through SPN with coefficient 0.205.

The diversity of endogenous latent variables that can be explained from the diversity of latent exogenous measured by R^2, which can be seen in Figure 3. The largest value of R^2 is 0.639. This means that 63.9% of the diversity SPR that can be explained by SI and SPL. The 37.1% is explained by other variables that are not included in the model.

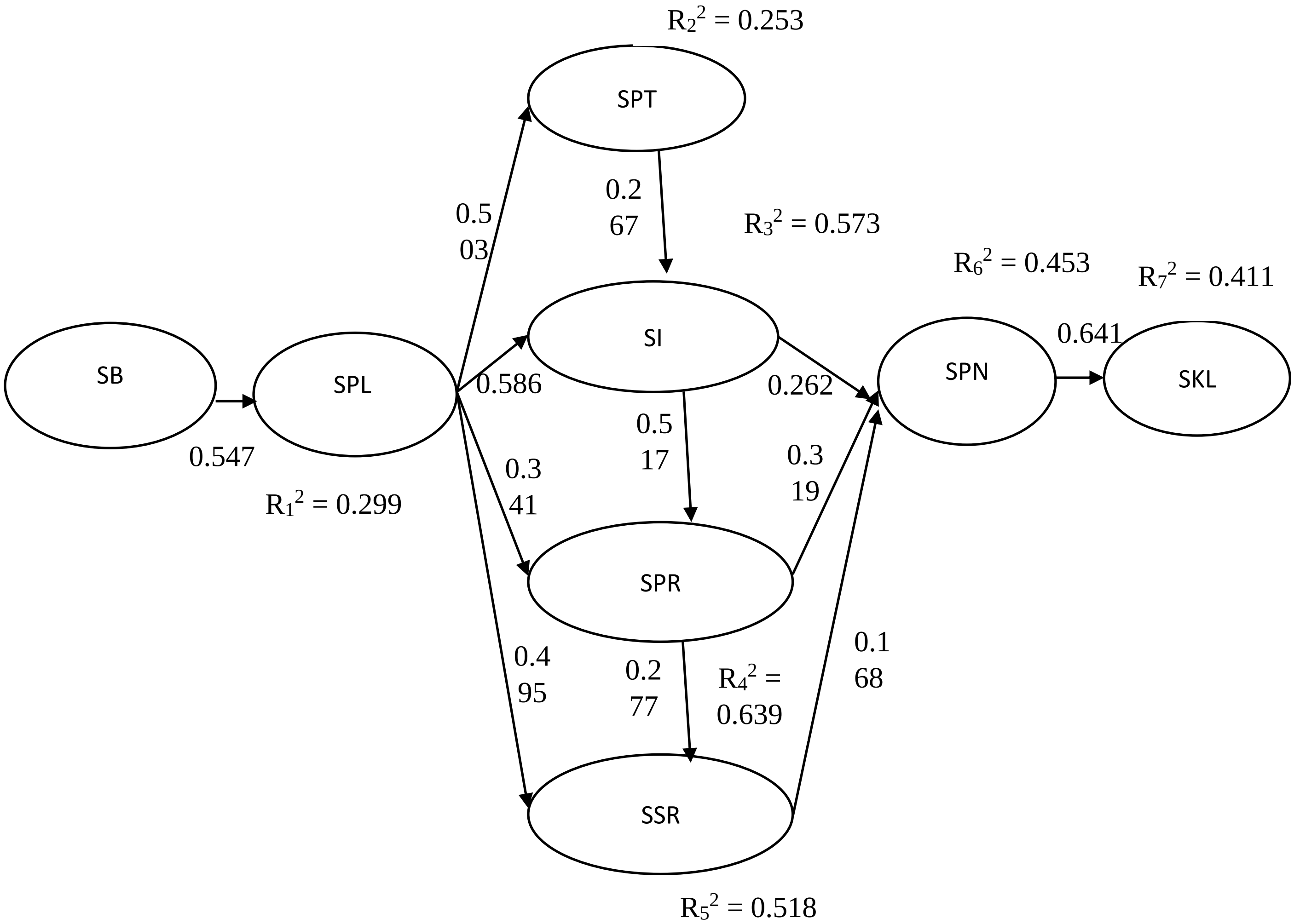


Figure 3. The result of Structural Models

3.4 Overall Evaluation Model

The evaluation of overall model is test of goodness of fit. The value of SRMR is 0.107 and GFI is 0.997. Thus, it can be said that the model is good. FIT measure the diversity of all the variables that can be explained by the model. The value of FIT is 0.467. This means that 46.7%, the diversity of all variables can be explained by the model.

4.Conclusion and Remarks

GSCA is a method of SEM without any assumption. All of indicator variables that used on this research was valid and reliable to measure latent variable after modification. The endogen laten variables (SI, SPR, SPT, SSP, SPL, SB and SPN) significant influence of the SKL. The latent variables that greatly affect SKL is SPN. The latent variables that greatly indirect affect SKL is SPR through SPN.

References

[1] Bollen K.A. 1989. “Structural Equation with Laten Variabels”. Departement of Sociology, John Wiley & Sons, New York

[2] Chin, W.W. 1989. “Structural Equation with Laten Variables”. New York. John Wiley and Son.

[3] Chou, C.P and Bentler, P.M. 1985. Estimate and Tests in Structural Equation Modelling. In R.H. Hoyle (Ed). “Structual Equation Modelling: Concepts, Issues and Application” (pp. 37-55) Newbury Park. CA. Sage

[4] De Leeuw, J., Young, F.W., &Takane, Y. 1976. “Regression With Qualitative and Quantitative Variables: An Alternating Least Squares Method With Optimal Scaling Features”. Psychometrica.Vol. 41 No.4.pp.505-529.

[5] Fornell, C and Bookstein, F. 1982. “Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory”. Journal of Marketing Research.19. 440-452.

[6] Fornell, C and Larcker, D. 1981. “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error”. Journal of Market Research. 18. Pp. 39-50

[7] Ghozali, I. 2008. “Generalized Structured Component Analysis (GSCA)”.Universitas Diponegoro, Semarang, Indonesia

[8] Joreskog, K.G. 1973. “A general Method For Estimating a Linear Structural Equation System”. In A.S Goldgerger & O. D. Duncan (Eds). Structural Equation Models in Social Sciences (pp. 85-112). New York Academic Press.

[9] Hwang, H and Takane. 2004.”Generalized Structured Component Analysis”. Psychometrica.Vol.69 No.1. pp. 81-99

[10] Hwang, H and Takane. 2009. “Nonlinear Generalized Structured Component Analysis”. Psychometrica

[11] Hwang, H., De Sarbo, S. W., &Takane, Y. (2007). “Fuzzy Cluster Wise Generalized Structured Component Analysis”. Psychometrika, 72, 181-198.