Structural Causality Between National Examination Score and the School Accreditation Categories

Anwar Fitrianto[1*], Budi Susetyo[1], Iswan Achlan Setiawan[2]

[1] Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia.
[2] Ministry of Education and Culture, Republic of Indonesia, Jakarta, Indonesia.

Abstract: This study aims to compare and determine the best model to describe the relationship between National Education Standard (NES) and CBNE scores using generalized structured component analysis. Model 1 describes the causal relationship between the NES and CBNE based on the educational theory of the Ministry of National Education and the Ministry of Religion (2010), Model 2 describes the causal relationship between the NES and CBNE based on the educational theory of the Ministry of Education and Culture (2012), and Model 3 describes the causal relationship between the NES and CBNE based on the educational theory of the Ministry of Education and Culture (2017). The results of the structural model evaluation have found that in Model 1, the SI path coefficient to Academic Achievement (PA) is not significant, in Model 2, the SI path coefficient to PA and SPT to SPN is not significant and in Model 3, the SI path coefficient to PA is also not significant. The coefficient of determination of each endogenous latent variable for each model ranges from 0.20 - 0.75. While the resulting Q-square value for all models is more than 0.9 to represent very good predictive relevance. Based on the overall goodness of fit, it is found that Model 3 produces the largest FIT and AFIT values. So it can be said that model 3 is better than other models. This model produces 11 invalid indicator variables, namely points 17, 39, 51, 55, 57, 59, 73, 75, 76, 80, and 108. The study found that National Education Standards that significantly affect academic achievement are graduate competency standards, process standards, and educational assessment standards.

Keywords: Structural equation modeling; Generalized structured component analysis; National education standards; National exams

Introduction

The National Education Standard (NES or SNP) is a minimum requirement for the quality of the education system in Indonesia. It is used as a reference in planning, implementing, and supervising education and being the criteria for various elements of education implementers and providers, especially in monitoring and guaranteeing the quality of education. Quality assessment for accreditation of academic units must refer to the NES. Based on Government Regulation (GR) Number 19 of 2005, the NES consists of 8 standards,

namely process standards (SPR), graduate competency standards (SKL), financing standards (SB), content standards (SI), facilities and infrastructure standards (SSP), educational assessment standards (SPN), management standards (SPL), and standards for educators and education personnel (SPT). SKL is an essential part of the NES, used as the primary standard in developing other standards. The results of the SKL can be used to evaluate the NES regularly, for example through the national exam (NE or UN). Since 2015, the Indonesian government has implemented the NE in 2

*Email: anwarstat@gmail.com

forms: the paper and pencil national exam (PPNE) and the computer-based national exam (CBNE or UNBK). The causality relationship between the NESs has always been the concern of many people who pay attention to education. Several educational theories explaining the causality of the eight NESs were published by the Ministry of National Education and the Ministry of Religion (2010), the Ministry of Education and Culture (2012), and the Ministry of Education and Culture (2017). Several researchers have also researched the causality of NESs. The NES causality relationship at the SMK level has been carried out by Hijrah (2018). Meanwhile, the causal relationship at the high school (MA or SMA) level has been carried out by Ferezagia (2015) using generalized structured component analysis (GSCA).

Structural equation modeling (SEM) can be used to see the relationship between latent variables and related indicators. Eight NESs are latent variables measured through the indicator variables in statement items contained in the accreditation instrument. One of the analyzes that can be used to measure the pattern of the relationship between latent variables and the relationship between latent variables and their indicator variables is structural equation modeling (SEM). According to (Ghozali and Kusumadewi, 2016), SEM has two types of approaches, namely variance-based structural equation modeling, namely PLSPM and GSCA, and covariance-based structural equation modeling CBSEM. Unfortunately, the use of CBSEM must meet several assumptions, so its use is limited. The assumptions needed in CBSEM include parametric assumptions, reflective indicator models, and large sample sizes (Reinartz et al., 2009). To overcome the limitations of CBSEM, Wold (1982) developed partial least square path modeling (PLSPM) as a VBSEM method that does not require parametric assumptions. PLSPM is very effective on small samples and can be used on reflective and formative indicator models (Haenlein & Kaplan, 2004). The limitation of PLSPM is in measuring the overall goodness of the model. To overcome the shortcomings of several previous methods, Hwang and Takane (2004) introduced GSCA as a method free of assumptions and has global optimum criteria that can measure the model's overall goodness.

Estimating GSCA parameters is carried out with the assumption that all objects are sampled from a homogeneous population. However, differences in behavior, attitudes, or preferences on objects result in heterogeneity. Bezdek (1981) introduced a technique to deal with data heterogeneity on latent variables by combining clustering methods and GSCA in one framework. Clustering techniques are divided into two, namely hierarchical and non-hierarchical methods. Non-hierarchical methods are divided into hard

clustering and fuzzy clustering. Fuzzy clustering is a clustering method that allows each object to be part of several clusters. Hwang et al., (2007) proposed the fuzzy cluster-wise generalized structured component analysis (FCGSCA) method to combine GSCA with fuzzy clustering. There are several reasons for choosing fuzzy clustering as the clustering technique used, namely the assumption-free concept in line with GSCA, easier to use in calculations, and the results obtained are not affected by local-optimum problems. It is essential to construct a generalized structured component analysis to see the structural relationship between the computer-based national examination and the NES.

Method

This study uses secondary data on accreditation status data and CBNE from public schools/madrasah at the SMP/MTs level in 2017–2018, including 2069 schools. Accreditation data consists of 8 NESs and a score of 124 statement items on a Likert scale of 0-4. Meanwhile, the CBNE data contains scores of 4 subject indicators, namely Indonesian Language (BIN), English (ING), Mathematics (MAT), and Science (IPA) obtained from the Research and Development Agency, Ministry of Education, and Culture. The questions used to become the variables that will be grouped into nine latent variables.

Generalized Structured Component Analysis

GSCA is viewed as a component-based SEM with latent variables defined as components or weighted composites of indicator variables. GSCA has been widely used in various psychological and biochemical studies (Hwang et al. 2012, 2013; Jung et al. 2012; Romdhani et al. 2015).

GSCA consists of 3 sub-models, namely measurement model, structural model, and weighting model. The measurement model describes the relationship between indicator variables and latent variables. Mathematically the measurement model in the GSCA is written as follows (Ryoo and Hwang 2017):

$$z = C'\gamma + \varepsilon, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

where z is the indicator variable matrix, γ is the latent variable matrix, C is the loading matrix, and ε is the residual vector for z. The structural model describes the relationship between indicator variables and latent variables. Mathematically the structural model in the GSCA is written as (Ryoo and Hwang, 2017):

$$\gamma = B'\gamma + \zeta, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(2)$$

where B is the path coefficient matrix, and ζ is the residual vector for γ. While the weighting model is used

to estimate the value of the latent variable with the following formula (Ryoo and Hwang, 2017):

$$\gamma = W'z, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3)$$

where W is the component weight matrix.

From these three equations, a single equation is obtained as follows:

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

$$\begin{bmatrix} I \\ W' \end{bmatrix} z = \begin{bmatrix} C' \\ B' \end{bmatrix} W'z + \begin{bmatrix} \varepsilon \\ \xi \end{bmatrix}.$$

If I is an idebtity matrix, then

$$V = \begin{bmatrix} I \\ W \end{bmatrix}, A = \begin{bmatrix} C \\ B \end{bmatrix}, e = \begin{bmatrix} \varepsilon \\ \xi \end{bmatrix}$$

$$V'z = A'W'z + e$$

$$ZV = ZWA + E. \dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

If $\Psi = ZV$ and $\Gamma = ZW$, then the GSCA model is obtained as follows (Hwang and Takane 2004):

$$\Psi = \Gamma A + E. \dots\dots\dots\dots\dots\dots\dots\dots\dots.. (5)$$

Parameter Estimations

Estimation of GSCA parameters was carried out using the alternating least square (ALS) method. The unknown GSCA parameters (V,W dan A) are estimated so that the least-squares value of all residuals (e_i) is as small as possible for all observations. This is equivalent to minimizing the following least-square criteria:

$$f = SS(ZV - ZWA) = SS(\Psi - \Gamma A) \dots\dots ..(6)$$

ALS is a general approach to parameter estimation that involves grouping a parameter into parts and then obtaining the least squares for one of the parameter parts, assuming that all the remaining parameters are constant. The GSCA method consists of two subsets, namely A, V, and W.
Hwang and Takane (2014) wrote that the ALS algorithm used in the GSCA consists of 2 stages. In the first stage, A is estimated at fixed V and W. In the second stage V and W are estimated at fixed A. To estimate matrix A in the first step, Equation (6) can be written in the form:

$$f = SS(vec(\Psi) - vec(\Gamma A)), \dots\dots\dots\dots.. (7)$$

where $vec(X)$ is a super vector formed by stacking all X columns into one column.

Structural Model Evaluation
Evaluation of the structural model is done by looking at the path coefficients from exogenous to endogenous variables and seeing their significance values. The significance value is obtained from the results of bootstrap resampling by dividing the coefficient value by the standard error value. In addition, the structural model is also evaluated by using the coefficient of determination for endogenous latent variables and by looking at the Q-square predictive relevance.

a. Coefficient of determination
The coefficient of determination is used to measure the predictive power of the structural model. The interpretation is the same as the coefficient of determination in the regression. The coefficient of determination represents the amount of variation that endogenous variables can explain to exogenous variables. The value of the coefficient of determination ranges from 0 to 1.

b. Q-square predictive relevance
Q-square predictive relevance is used to test the overall goodness of the structural model. Q-square can be calculated using the formula:

$$Q^2 = 1 - (1 - R_1^2)(1 - R_2^2) \dots (1 - R_p^2), \qquad (8)$$

where $R_1^2, R_2^2, \dots, R_p^2$ is the coefficient of determination of the endogenous latent variable. The value of Q^2 is between 0 and 1. The value of Q^2 is more than 0.5, indicating that the model has good predictive relevance (Chin 2010).

Measurement Measurement Model
The evaluation of the model is distinguished based on the form of the relationship between the measurement model. If the measurement model is reflective, then the evaluation of the model is done by looking at the convergent validity, discriminant validity, and composite reliability.

Discriminant validity measures the extent to which one latent is entirely different from another. It represents a good model if the value of the $\sqrt{AVE}$ of each latent is greater than the value of the correlation between the other latents in the model. The AVE is a coefficient that describes the variability in indicators that common factors can explain. It can be calculated using the following equation (Fornell and Lacker 1981):

$$AVE = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum(1 - \lambda_i^2)}, \dots\dots\dots\dots\dots\dots(9)$$

where λ_i is the loading factor, and $1 - \lambda_i^2$ is $var(\varepsilon_i)$.

Overall Goodness-of-fit
The overall goodness of fit model is carried out with FIT and AFIT tests. The FIT model is defined as follows (Ryoo and Hwang 2017):

$$FIT = 1 - \left[\frac{trace((ZV - ZWA)'(ZV - ZWA)}{trace((ZV)'(ZV))} \right], \dots\dots(10)$$

The FIT value is in the range of 0 to 1 and can be interpreted as a large variety of data that the model can explain. The larger the value, the greater the diversity of data that the model can explain. However, the FIT value is influenced by the complexity of the model, namely the more parameters, the higher the FIT value. Adjusted FIT (AFIT) was developed unaffected by the complexity of the model. The model with the largest AFIT value is considered the best model. AFIT is defined as follows (Ryoo and Hwang 2017):

$$AFIT = 1 - (1 - FIT)\frac{d_0}{d_1}, \ldots \ldots \ldots \ldots (11)$$
$$d_0 = N*J,$$
$$d_1 = N*J - k,$$

where d_0 is the degree of freedom of the model 0 (W=0 and A=0), d_1 is the degree of freedom of the model under the test and k is the number of parameters.

Result and Discussion

The Structural Model

The resulting $\sqrt{AVE}$ value for the three models is greater than the correlation value between Academic Achievement (PA) and other latent variables. The PA is good discriminant validity while the Cronbach alpha value obtained is > 0.70, which means that the model has good reliability. The formative measurement model is evaluated by looking at the Critical Ratio (CR) value. The indicator is valid if the CR value is more than 1.96 and vice versa. Indicator variables that are not excluded from further analysis.

Of the 124 indicators, model 1 produces 14 indicators with a CR value of less than 1.96, namely at points 13, 17, 38, 39, 46, 51, 55, 57, 59, 74, 75, 76, 80, and 108. Model 2 resulted in 13 invalid indicator variables, namely items 13, 38, 39, 46, 51, 55, 57, 59, 72, 74, 75, 80, and 108.

Table 1: Coefficient of determination and Q^2 of each latent variable in each model

	Model 1	Model 2	Model 3
SB	0	0	0.485
SPL	0.486	0.484	0.000
SPT	0.563	0.577	0.563
SI	0.550	0.577	0.626
SPR	0.678	0.717	0.710
SSP	0.600	0.584	0.641
SPN	0.633	0.644	0.631
SKL	0.473	0.655	0.655
PA	0.243	0.242	0.241
Q^2	0.9981	0.9989	0.9992

Table 1. describes the variability of endogenous latent variables that exogenous latent variables can

explain. The coefficient of determination of each endogenous latent variable for each model ranges from 0.20 - 0.75. The smallest value of the coefficient of determination is around 0.24 on the latent variable PA. This means that the variance of latent PA variables that can be explained in the model is around 24% and the rest is explained by other variables not included in the model. While the resulting Q-square value for all models is more than 0.9, it can be said that all models have excellent predictive relevance. However, Model 3 produces the highest Q-square predictive relevance among other models.
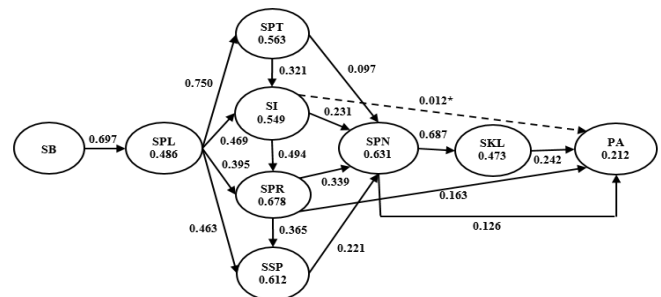

Figure 1. Structural Model 1 according to the Ministry of National Education and the Ministry of Religion (2010)

Based on the evaluation of the model, the structural model obtained can be seen in Figure 1, Figure 2 and Figure 3. The model is equipped with the coefficient of determination of each endogenous latent variable.
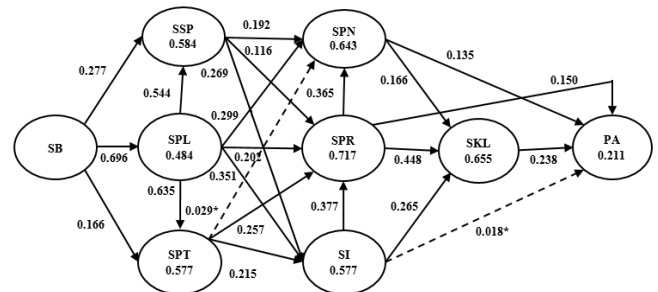

Figure 2. Structural Model 2, according to the Ministry of Education and Culture (2012)
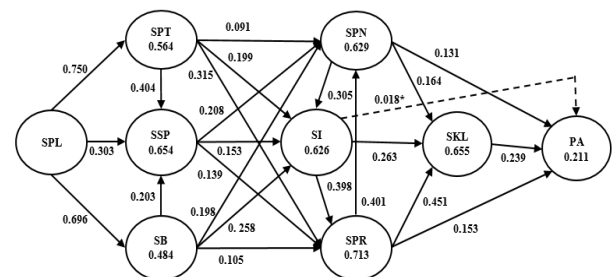

Figure 3. Structural Model 3, the Ministry of Education and Culture (2017)

Comparisons between Accreditation Instrument Weights and Model Weights

A comparison of the weight of the accreditation instrument and the model weight is carried out to see how far the difference in the level of importance of each instrument/statement item is. These results can be taken into consideration in determining the weight of the next accreditation instrument. The SMP/MTs accreditation instrument contains 124 questions, each with a different weight depending on their learning support. The lowest statement item is given a weight of 1, and the highest statement is given 4. The weight of the instrument can be seen in Appendix 7. The operational definition of item weight is as follows (Ministry of Education and Culture, 2017):

- Weight 1 is the minimum weight to support component functions in the learning process.
- Weight 2 is the weight that supports the component's function in a proper learning process.
- Weight 3 is a weight that supports the function of these components in a good learning process.
- Weight 4 is the maximum weight that supports the function of these components in a very good learning process.

The model's weight produces the lowest value of 0.043 and the highest value of 0.298. The model weight scale is converted into a scale of 1 to 4 according to the weight of the accreditation instrument. Spearman correlation value of the weight of the accreditation instrument and the model's weight is 0.422. These results indicate a correlation between the weight of the accreditation instrument and the model's weight. The direction of the relationship shows a positive value, that is, if the weight of the accreditation instrument is high, the weight of the model will also be high.

Conclusion

Based on the study, it can be concluded that the model published by the Ministry of Education and Culture (2017) is the best model when compared to the other two models, especially in SMP/MTS data in 2017. It was found that Model 3 produces the highest Q-square predictive relevance among other models. This model produces 11 statements in the accreditation instrument that are not valid, namely points 17, 39, 51, 55, 57, 59, 73, 75, 76, 80, and 108. National Education Standards that significantly affect academic achievement are graduate competency standards, process standards, and educational assessment standards.

References

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms.* New York. Plenum Press.

Chin, W.W. (2010). How to write up and report PLS analyses. In Vinzi VE, Chin WW, Henseler J, Wang H. *Handbook of Partial Least Squares: Concepts, Methods, and application*. Vinzi VE, Chin WW, Henseler J, Wang H. Jerman (DE): 645-689. Springer. Retrieved from: https://link.springer.com/chapter/10.1007/978-3-540-32827-8_29

Ferezagia D.V. (2015). *Model persamaan struktural delapan standar nasional pendidikan dengan generalized structured component analysis.* (Unpublished Master Thesis). Institut Pertanian Bogor.

Ghozali I, Kusumadewi KA. (2016). *Model persamaan struktural PLS-PM GSCA RGCCA*. Semarang: Yoga Pratama.

Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, *3*(4), 283–297. https://doi.org/10.1207/s15328031us0304_4

Hijrah, M, et al., Susetyo B, Sartono B. 2018. Structural Equation Modeling of National Standard Education of Vocational High School Using Partial Least Square Path Modeling. *IJSRSET*. 4(4), 1418-1422. https://doi.org/10.32628/IJSRSET1844465

Hwang H, et al., (2012). Functional multi-set canonical correlation analysis. *Psychometrika.* 77(1), 48–64. https://doi.org/10.1007/S11336-011-9234-4

Hwang H, et al., (2013). A unified approach to multi-set canonical correlation analysis and principal component analysis: an application to functional neuroimaging data. *Br. J. Math. Stat. Psychol*. 66(1), 308-321. https://doi.org/10.1111/j.2044-8317.2012.02052.x

Hwang H., De Sarbo S. W., Takane Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika* 72(1), 181–198. https://doi.org/10.1007/s11336-005-1314-x

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*. US: Chapman and Hall/CRC.

Jung K, et al., (2012). Dynamics generalized structured component analysis with applications to the analysis of effective connectivity in functional neuroimaging data. *Psychometrika*. 77, 827–848. https://doi.org/10.1007/s11336-012-9284-2

Ministry of Education and Culture. 2017. *Peraturan menteri pendidikan dan kebudayaan Nomor 003/H/AK/2017 tahun 2017 tentang kriteria dan*

perangkat akreditasi SMP/MTs. Jakarta. Sekretariat Negara.

Ministry of National Education and the Ministry of Religion. 2010. *Sistem penjaminan mutu pendidikan: panduan teknis evaluasi diri sekolah (EDS)*. Jakarta. Direktorat Jenderal Pendidikan Dasar dan Menengah.

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332–344. https://doi.org/10.1016/j.ijresmar.2009.08.001

Romdhani H, et al., (2015). Pathway-based association study of multiple candidate genes and multiple traits using structural equation models. *Genet. Epidemiol*. 39, 101–113. https://doi.org/10.1002/gepi.21872

Ryoo, J. H. & Hwang H. (2017). Model evaluation in generalized structured component analysis using confirmatory tetrad analysis. *Frontiers in Psychology*. 8, 1-10. https://doi.org/10.3389/fpsyg.2017.00916

Wold, H. (1982). Soft modeling: The basic design and some extensions. In *systems under indirect observation: causality, structure, prediction*. Amsterdam (NL): North Holland. https://doi.org/10.1207/s15328031us0304_4