

Application of the fuzzy clusterwise generalized structured component method to evaluate implementation of national education standard in Indonesia

Susetyo Budi^{a*} and Wahyuni Rezi^a

^aIPB University, Indonesia

CHRONICLE

ABSTRACT

Article history:

Received: July 10, 2020

Received in revised format:

October 18 2020

Accepted: November 4, 2020

Available online:

November 4, 2020

Keywords:

FCGSCA

Heterogeneity Group

National Education Standards

National Examination

Results of school accreditation and national examinations are two indicators that are often used to describe the achievement of quality in education in Indonesia. ‘Accreditation’ reflects the fulfillment of 8 national education standards (NES), while the national exam (NE, or UN in Bahasa) for students describes academic performance. Eight NES and academic performance are latent variables. The relationship between the two variables and the validity of its indicators can be evaluated by several methods. Path analysis with latent variables can be obtained through general structured component analysis (GSCA) with the assumption of homogeneity of variance. Since the data are not homogeneous, this study aims to apply the fuzzy clusterwise generalized structured component analysis (FCGSCA) to evaluate the relationship between the NES and the UN, and the validity of the indicators. The results showed that there were two school clusters in Indonesia. The evaluation of the measurement model indicated that some indicators of the accreditation instrument were not valid, i.e., 6 indicators in cluster 1 and 15 indicators in cluster 2. The structural model evaluation of the two clusters indicated that standard of process to the UN was not significant. Based on the overall goodness of the fit model, the total diversities of all variables that could be explained were 61.60% in cluster 1 and 59.90% in cluster 2.

© 2021 by the authors; licensee Growing Science, Canada

1. Introduction

Law No. 20 of the year 2003 of the Republic of Indonesia regulates the national education system. The government, in charge of the educational system, established eight national education standards (NES) for schools, such as the standard of content (SI), the standard of process (SPR), the standard of competency (SKL), the standard of education and staff (SPT), the standard of infrastructure (SSP), the standard of management (SPL), the standard of cost (SB), and the standard of assessment (SPN). The NES, as the minimum criterion of the education system in Indonesia, is used as a reference in planning, implementing, and evaluation by all education stakeholders to guarantee the quality of the education. Accreditation is a process to measure compliance with the 8 national standards. As a subsystem of the NES, SKL is a level of competence be achieved by students at a given level of education. Several statistical methods have been implemented to evaluate relationships among NES. In a study on accreditation data, the SPR and SPN were shown to have the greatest influence on the SKL using the generalized structured component analysis (GSCA) method (Vita et al., 2013). Furthermore, SB had a significant influence on SKL in the assessment of vocational school accreditation using a partial least square path modeling (PLSPM) method (Hijrah et al., 2018). One of the assessments in SKL is done based on the result of the national examination (UN). Some studies indicated the relationship between the NES and the UN. For instance, SKL, SPN, and SPR had significant effects on academic achievement in the secondary schools (Setiawan et al. 2018).

* Corresponding author.

E-mail address: buset008@yahoo.com (S. Budi)

NES, as a latent variable, is measured through indicators consisting of several elements in the accreditation instrument. The structural equation modeling (SEM) measures the relationship between latent variables and indicator variables. There are two types of SEM, i.e., covariance-based structural equation modeling (CBSEM) and variance-based structural equation modeling (VBSEM). The CBSEM introduced by Joreskog (1978), requires several assumptions (Joreskog, 1970). Unlike CBSEM, the VBSEM, namely the PLSPM, was introduced by Wold (1982) and it is free of assumptions (Wold, 1982). Unfortunately, it cannot measure the overall goodness of a fit model. Then, Hwang and Takane (2004) proposed the GSCA, as another VBSEM method to overcome the weaknesses of the two previous methods (Hwang & Takane, 2004). The GSCA is an assumption-free method with optimal global criteria to measure the overall goodness of a fit model. The parameters of GSCA are estimated under the assumption that the objects are sampled from a homogeneous population. When the data are heterogeneous, a technique for dealing with the heterogeneity of data in latent variables is the combination of clustering and GSCA in a single framework (Bezdek, 1981). The fuzzy clusterwise generalized structured component analysis (FCGSCA) is a method that combines the GSCA with fuzzy clustering (Hwang et al., 2007). One of the reasons for choosing the fuzzy clustering as a clustering technique is that it is compatible with the assumption-free concept of the GSCA.

This paper aims to reveal the characteristics of school clustering in Indonesia and evaluate the relationship between accreditation and the UN at the secondary school level using the FCGSCA.

2. Materials and methods

2.1 Data

The data consisted of 6434 secondary schools in Indonesia based on accreditation and computer-based national examination (UNBK) data in 2017 and 2018. The accreditation data obtained from the National Accreditation Board (in Bahasa BAN-S/M) consisted of 124 items of instrument statements with a Likert scale of 0 to 4. Meanwhile, the UNBK data were obtained from the Research and Development Board of the Ministry of Education and Culture, consisting of 4 subject scores, namely Indonesian Language (BIN), English (ING), Mathematics (MAT), and Natural Sciences (IPA). The eight NES are defined as the formative measurement model and the UNBK as a reflective measurement model (Table 1).

Table 1

Latent, indicator and measurement model

Latent	Indicator	Measurement model
SI	item 1 – item 9	formative
SPR	item 10 – item 30	formative
SKL	item 31 – item 37	formative
SPT	item 38 – item 56	formative
SSP	item 57 – item 80	formative
SPL	item 81 – item 95	formative
SB	item 96 – item 111	formative
SPN	item 112 – item 124	formative
UNBK	4 subjects	reflective

2.2 Data analysis

The stage of data analysis used in this study:

Step 1. Checking the homogeneity of the variances.

A homogeneous variability test is conducted on inter-provincial school data with hypotheses as follows:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k \quad (k = 1, 2, \dots, 34)$$

$$H_1 : \text{there is least one } \Sigma_i \neq \Sigma_j, \text{ for } i \neq j \quad (i, j = 1, 2, \dots, 34).$$

The statistic test used Box's M – test (Rencher, 2002)

Step 2. Fuzzy clustering using fuzzy c-means (Bezdek, 1981).

a. Generating a random number u_{ki} which is the value of membership for i th object in k th cluster, where i is the number of observations and k is the number of clusters, and u_{ki} is the initial membership value (U^0).

b. Calculating the center of the cluster using the following formula:

$$V_{kj} = \frac{\sum_{i=1}^N (u_{ki})^m X_{ij}}{\sum_{i=1}^N (u_{ki})^m}$$

where m is the weight of the component and X_{ij} is i th object in j th variable.

c. Calculating the objective function for t th iteration (U^t) by:

$$J_{FCM} = \sum_{k=1}^K \sum_{i=1}^N (u_{ki})^m d_{ki}^2$$

where d_{ki} is the distance between i th object in k th cluster.

d. Updating membership values by:

$$u_{ki} = \frac{(d_{ki}^2)^{-1}}{(\sum_{k=1}^K d_{ki}^2)^{-1}}$$

e. Checking the convergence value.

If $|U^t - U^{t-1}| < \varepsilon$, then it is convergent and stops the iteration. Otherwise, it goes back to step 2(b).

Step 3. Choosing the optimum number of clusters. Fuzziness performance index (FPI) and normalized classification entropy (NCE) are used to measure the validity of the number of clusters. The FPI and NCE are given by (Roubens, 1982):

$$FPI = 1 - \frac{K \left(\frac{1}{N} \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \right) - 1}{K - 1}$$

$$NCE = \frac{-\frac{1}{N} \sum_{k=1}^K \sum_{i=1}^N u_{ki} \log u_{ki}}{\log K}$$

The FIT and AFIT values indicated the extent to which the model can explain the diversity of data. The criteria for the optimum cluster are given by the smallest values of FPI and NCE and the largest values of FIT and AFIT (Hwang et al., 2007). The FIT and AFIT are stated as follows:

$$FIT = 1 - \frac{\sum_{k=1}^K SS(\mathbf{Z}_k \mathbf{V}_k - \mathbf{Z}_k \mathbf{W}_k \mathbf{A}_k)}{\sum_{k=1}^K SS(\mathbf{Z}_k \mathbf{V}_k)}$$

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

where $Z = [z_1, \dots, z_N]'$ is a matrix of indicator variables, \mathbf{V}_k , \mathbf{W}_k , \mathbf{A}_k are parameters of GSCA for k th cluster, $d_0 = NJ$ is the degree of freedom for the null model ($\mathbf{V}_k = \mathbf{1}$, $\mathbf{W}_k = \mathbf{0}$, $\mathbf{A}_k = \mathbf{0}$), $d_1 = NJ - G$ is the degree of freedom for the model being tested, where G is the number of free parameters.

Step 4. Estimating the parameters for each cluster, including weighted estimator, loading factor estimator, path coefficient and standard errors.

Step 5. Evaluating measurement models.

- a. Evaluate the reflective model based on convergence validity, discriminant validity, and composite reliability. The convergence validity can be evaluated through the loading factor values. It is recommended that the loading factor should be greater than 0.70 and significant for all indicators (Henseler et al., 2009). Furthermore, the discriminant validity is examined by looking at the square root of the average variance extracted (AVE) for each latent variable. The AVE value should be greater than the correlation with the other latent variables in the model (Fornell & Larcker, 1981). The AVE is measured by:

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

where λ_i is loading factor component and $1 - \lambda_i^2$ is variance.

Meanwhile, the composite reliability is measured by Cronbach's alpha. The value of Cronbach's alpha should be above 0.70 is recommended (Nunnally, 1978). Cronbach's alpha is defined by (Cronbach, 1951):

$$\text{cronbach's } \alpha = \left(\frac{J}{J-1} \right) \left(1 - \frac{\sum_{i=1}^J \sigma_i^2}{\sigma_t^2} \right)$$

σ_i^2 is the variance for i th indicator, σ_t^2 is variance all of the indicators.

- b. Evaluate the formative model based on the statistic significance of the weighted value, the critical ratio (CR), and the multi-collinearity test by using the variance inflation factor (VIF) value. The recommended weight significance is more than 1.96, while the VIF value should be less than 10 (Hair et al., 1995).

Step 6. Evaluate the structural models based on the significance of the path coefficient value on the latent variable and coefficient of determination (R^2).

Step 7. Evaluate the overall goodness of the fit model based on FIT and AFIT values of each cluster.

3. Results and discussion

The results of the homogeneity test using Box's M-test concluded that the school data come from heterogeneous populations, then used the clustering method through fuzzy c-means to resolve. The number of clusters in this study was determined with $K = 2, 3, 4, 5,$ and 6 . Based on the number of clusters, there were a different number of iterations until the convergence condition of the objective function. Choosing the number of optimum clusters was determined based on criteria such as FPI, NCE, FIT, and AFIT values. Number of cluster $K=2$ was chosen as optimum cluster because it has the smallest FPI and NCE values and the largest FIT and AFIT (Table 2). The number of schools in cluster 1 are 4653 schools and cluster 2 are 1781 schools.

Table 2

The value of FPI, NCE, FIT and AFIT

K	FPI	NCE	FIT	AFIT
2	0.5047	1.3548	0.6077	0.6076
3	0.6486	1.5335	0.5550	0.5549
4	0.7414	1.6644	0.5105	0.5104
5	0.7951	1.7419	0.4939	0.4938
6	0.8362	1.8009	0.4808	0.4807

Table 3 represents the average scores of 8 NES and UNBK in each cluster. Schools in cluster-1 have lower scores than in cluster 2.

Table 3

The average scores for 8 SNPs and UNBK

Criteria	Cluster 1	Cluster 2
SI	87.36	92.85
SPR	84.93	90.85
SKL	83.91	90.73
SPT	75.83	83.27
SSP	78.72	89.30
SPL	85.37	91.56
SB	87.65	92.21
SPN	86.39	92.04
UNBK	43.33	62.07

3.2 Evaluation of the Measurement Model

The measurement model specifies the relationship between indicators and latent variables. The UNBK was defined as a reflective measurement model. From the results of the analysis in clusters (Table 4), it was observed that the loading factor values were above 0.70 and significant at the 5% level for each indicator.

Table 4

Evaluation of a reflective measurement model

Indicator	Cluster 1		Cluster 2	
	Loading	CR	Loading	CR
BIN	0.850	126.74	0.829	66.47
ING	0.824	104.52	0.813	88.53
MAT	0.804	98.66	0.926	172.04
IPA	0.889	166.10	0.904	143.96
AVE	0.7100		0.7560	
Cronbach	0.8440		0.8810	

By examining the composite reliability based on Cronbach's alpha, the value was greater than 0.7 for cluster 1 and 2 with around 0.844 and 0.881, respectively. Therefore, it can be concluded that the evaluation of the reflective measurement model based on the convergent validity, discriminant validity, and composite reliability test, was fulfilled in the clusters. The eight NESs were defined as the formative measurement model. The evaluation of the formative measurement model was based on the weight significance and VIF value of each indicator variable. The indicator was valid if the significance of the weight test was greater than 1.96. Based on the results of the formative model, there were invalid indicators in each cluster. Out of 124 indicators, 6 invalid indicators were observed in cluster 1, i.e., items 51, 57, 59, 65, 76 and 80. Meanwhile, cluster 2 had 15 invalid indicator variables, i.e. items 39, 46, 51, 54, 57, 59, 64, 65, 73, 75, 76, 80, 103, 105 and 107. The invalid indicators

were used as input in the education evaluation, especially 6 invalid items in both clusters, i.e., items 51, 57, 59, 65, 76, and 80. Furthermore, the multi-collinearity test of each indicator in each latent variable had the VIF value less than 10, so it can be concluded that there was no multi-collinearity among variables.

3.3 Evaluation of Structural Model

The structural model showed the relationship between latent variables. Fig. 1 shows the path diagram of the structural model in cluster 1. The results of parameter estimation obtained in the path coefficient value from SPL to SPT had the highest value (0.794). Also, SPL to SB showed the second-highest value (0.773). It was concluded that the relationship between SPL to SPT and SPL to SB had significant effects on them. Furthermore, the dashed line in the SPR to UNBK had the smallest path coefficient value (0.047), and it was not statistically significant, with a significance value that was less than 1.96. Thus, the SPR just had a small effect on UNBK.

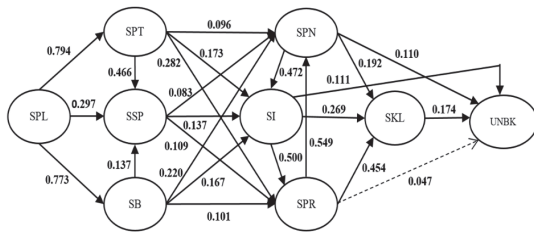


Fig. 1. The path diagram of cluster 1

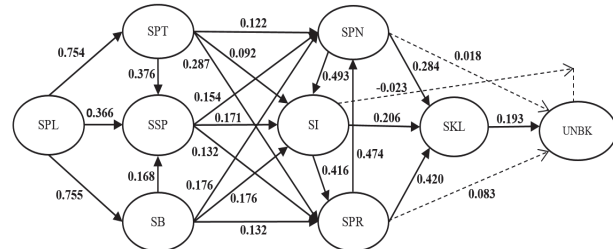


Fig. 2. The path diagram of cluster 2

Fig. 2 shows the path diagram of the structural model in cluster 2. The path coefficient from the SPL to SB and SPL to SPT had the highest value, with around 0.755 and 0.754, respectively. Like cluster 1, SPL to SB and SPL to SPT have had significant effects among them. However, 3 dashed lines were showing that some coefficient paths were not significant in cluster 2. There were path coefficient values from the SI to UNBK, SPN to UNBK and SPR to UNBK which were not statistically significant, i.e., around 0.32, 0.54, and 1.78, respectively. These three standards did not show a significant effect on the UNBK value in cluster 2 with the value of each path coefficient below 1.96. Further, the structural model was also evaluated by the coefficient of determination. The latent variable of the UNBK had the smallest in cluster 1 (0.171) and cluster 2 (0.068). The total diversity of the UNBK can be explained by the model equal to 1.71% in cluster 1 and 0.68% in Cluster 2, and other unavailable variables in the model explain the remainder. The form of the structural model equations obtained in cluster 1 and cluster 2 were as follows in Fig. 4 and Fig. 5.

$$\begin{bmatrix} \text{SPT} \\ \text{SB} \\ \text{SSP} \\ \text{SI} \\ \text{SPR} \\ \text{SPN} \\ \text{SKL} \\ \text{UNBK} \end{bmatrix} = \begin{bmatrix} 0.794 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.773 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.297 & 0.466 & 0 & 0.137 & 0 & 0 & 0 & 0 \\ 0 & 0.173 & 0.137 & 0.167 & 0.472 & 0 & 0 & 0 \\ 0 & 0.282 & 0.109 & 0.101 & 0 & 0.500 & 0 & 0 \\ 0 & 0.096 & 0.083 & 0.220 & 0 & 0 & 0.549 & 0 \\ 0 & 0 & 0 & 0 & 0.192 & 0.269 & 0.454 & 0 \\ 0 & 0 & 0 & 0 & 0.110 & 0.111 & 0.047 & 0.174 \end{bmatrix} \begin{bmatrix} \text{SPL} \\ \text{SPT} \\ \text{SSP} \\ \text{SB} \\ \text{SPN} \\ \text{SI} \\ \text{SPR} \\ \text{SKL} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Fig. 3. The structural model equations of cluster 1

$$\begin{bmatrix} \text{SPT} \\ \text{SB} \\ \text{SSP} \\ \text{SI} \\ \text{SPR} \\ \text{SPN} \\ \text{SKL} \\ \text{UNBK} \end{bmatrix} = \begin{bmatrix} 0.754 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.755 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.366 & 0.376 & 0 & 0.168 & 0 & 0 & 0 & 0 \\ 0 & 0.092 & 0.171 & 0.176 & 0.493 & 0 & 0 & 0 \\ 0 & 0.287 & 0.132 & 0.132 & 0 & 0.416 & 0 & 0 \\ 0 & 0.122 & 0.154 & 0.176 & 0 & 0 & 0.474 & 0 \\ 0 & 0 & 0 & 0 & 0.284 & 0.206 & 0.420 & 0 \\ 0 & 0 & 0 & 0 & 0.018 & -0.023 & 0.083 & 0.193 \end{bmatrix} \begin{bmatrix} \text{SPL} \\ \text{SPT} \\ \text{SSP} \\ \text{SB} \\ \text{SPN} \\ \text{SI} \\ \text{SPR} \\ \text{SKL} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Fig. 4. The structural model equations of cluster 2

The overall goodness of the fit model was examined based on the FIT and AFIT models of each cluster. The FIT obtained in cluster 1 was equal to 0.616, and that of cluster 2 was equal to 0.599. Like the FIT, the AFIT had the same value as FIT. So, it indicated that models could explain the total diversity of all variables was 61.60% for cluster 1 and 59.90% for cluster 2. Then, cluster 1 was better at describing the diversity of data than cluster 2.

4. Conclusion

In this study, there were two clusters for Indonesian schools. Cluster 1 was characterized by schools that had average NES and UNBK scores lower than schools those in cluster 2. The evaluation of the measurement model in clusters obtained 15 invalid items of accreditation statements and there were 4 items (point 39, 46, 51, and 54) in education and staff standards (SPT), 8 items (point 57, 59, 64, 65, 73, 75, 76, and 80) in infrastructure standards (SSP) and 3 items (point 103, 105 and 107) in cost standards (SB). Furthermore, the path coefficient of the management standard (SPL) to education and staff standards (SPT) and cost standards (SB) had the highest values. Meanwhile, the evaluation of the structural model showed that the path coefficient of the content standard (SI), assessment standard (SPN), and process standard (SPR) did not have significant effects on UNBK. Based on the overall goodness of the fit model, the FIT and AFIT values obtained were 61.60% in cluster 1 and 59.90% in cluster 2.

References

- Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: algebra and statistics. *Journal of Marketing Research*, 18(3), 328-388.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis with Readings*. Prentice-Hall International, Inc. Englewood Cliffs, New Jersey.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20, 277-319.
- Hijrah, M., Susetyo, B., & Sartono, B. (2018). Structural equation modeling of national standard education of vocational high school using partial least square path modeling. *International Journal of Scientific Research in Science Engineering and Technology*, 4(4), 1418-1422.
- Hwang, H., DeSarbo, S. W., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, 72(2), 181-198.
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69(1), 81-99.
- Joreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57, 409-426.
- Nunnally, J. C. (1978). *Psychometric Theory Second Edition*. McGraw-Hill, New York.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis Second Edition*. John Wiley and Sons, Inc. Publication, New Jersey.
- Roubens, M. (1982). Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*, 10(3), 294-301.
- Setiawan, A. I., Susetyo, B., & Fitrianto, A. (2018). Application of generalized structural component analysis to identify relation between accreditation and national assessment. *International Journal of Scientific Research in Science Engineering and Technology*, 4(10), 93-97.
- Vita, F., Susetyo, B., & Indriyanto, B. (2015). Generalized structured component analysis for national education standards of secondary school in Indonesia. *Global Journal of Pure and Applied Mathematics*, 11(4), 2441-2449.
- Wold, H. (1982). *Soft Modeling: The Basic Design and Some Extensions*. In Joreskog KG, Wold H. North-Holland, Amsterdam.



© 2021 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).