International Journal of Mathematics and Computer Research
ISSN: 2320-7167
Volume 13 Issue 06 June 2025, Page no. -5299-5304
Index Copernicus ICV: 57.55, Impact Factor: 8.316
DOI: 10.47191/ijmcr/v13i06.04

Structural Equation Modeling with Generalized Structured Component Analysis (SEM-GSCA) Estimation on The Human Development Index (HDI) of Central Java Province Equipped

Andrena Dhiffa Salsabila [1*], Sugito[2], Agus Rusgiyono[3]
[1,2,3] Department of Statistics, Universitas Diponegoro, Indonesia

ARTICLE INFO

Published Online:
04 June 2025

Corresponding Author:
Sugito

ABSTRACT

The Human Development Index (HDI) serves as a primary indicator in assessing the success of human development within a country. This study aims to analyze the latent variables influencing the HDI of Central Java in 2023 using the Generalized Structured Component Analysis (GSCA) method. GSCA is a variance-based Structural Equation Modeling (SEM) approach that does not require the assumption of multivariate normality and incorporates overall goodness-of-fit measures to evaluate the model's alignment with the data This research utilizes population data sourced from the Central Bureau of Statistics (BPS) of Central Java Province and examines the key indicators constituting the HDI. The analytical findings indicate that all indicators significantly contribute to their respective latent variables. Moreover, the education and health variables exert a significant influence on the economic variable. Furthermore, the HDI is substantially affected by economic, health, and education factors. The model evaluation results demonstrate a Fit Index (FIT) value of 62.3%, reflecting the model's explanatory capacity, and a Goodness of Fit Index (GFI) value of 96.9%, signifying a strong model fit. These findings provide a comprehensive understanding of the determinants of the HDI in Central Java and are expected to serve as a reference for policymakers in formulating strategies to enhance public welfare and overall quality of life.

KEYWORDS: Human Development Index; GSCA; SEM; and Overall Goodness of Fit.

I. INTRODUCTION

The Human Development Index (HDI) is the main indicator that measures the progress of human development in Indonesia. HDI is used as a tool to evaluate the quality of life and government performance [1]. HDI includes three main dimensions, namely the health dimension, the education dimension, and the decent standard of living dimension [2]. The Central Statistics Agency (BPS) made changes to the HDI calculation in 2014 following the revised UNDP calculation changes in 2010 [3]. Along with the development of the methodology, the Central Statistics Agency (BPS) adjusted the HDI calculation to improve accuracy, such as replacing the literacy rate indicator with the Expected Length of School (HLS) and Average School Length (RLS) indicators, as well as replacing the Gross Domestic Product (GDP) per capita with the Gross National Income (PNB) per capita [4]. The purpose of this methodological change is to present a more comprehensive picture of the conditions of human development in various regions.

In HDI data processing, the Generalized Structured Component Analysis (GSCA) method is increasingly relevant because it is able to evaluate the relationship between latent variables without requiring the assumption of normality [5]. The main advantage of the GSCA is its ability to analyze complex relationships between variables with smaller sample sizes, as well as providing a more accurate goodness-of-fit measure compared to PLS. In addition, the GSCA is able to accommodate both reflective and formative indicators simultaneously, making it more flexible in the analysis of social and economic data [6]. Previous research has applied GSCA in various fields, including customer satisfaction analysis [5] and public health [7], demonstrating the potential of this method in evaluating various social and economic phenomena. Research on the application of the GSCA method to analyze HDI in Central Java is limited, so a more comprehensive study is needed to identify each factor that affects HDI at the district/city level. Thus, the purpose of this study is to evaluate indicators that have an influence on HDI in Central Java Province using the GSCA method, with the hope of providing more accurate policy recommendations to improve human development in the area.

II. THEORETICAL FRAMEWORK

A. *Generalized Structured Component Analysis (GSCA)*

Generalized Structured Component Analysis (GSCA) is a component-based approach used to analyze the relationships between latent variables and indicator variables in a model. The GSCA method was developed [6] as an alternative to the Partial Least Squares (PLS) to overcome its limitations in model testing and global optimization of estimation parameters. The GSCA uses a matrix-based approach as well as a path diagram to describe the cause-and-effect relationships between variables, thus allowing for better model evaluation.

The GSCA includes three types of equations in the submodel specification, namely the measurement model (outer model), the structural model (inner model), and the weight relation. The measurement model describes the relationship between indicator variables and latent variables, while structural models analyze the relationships between latent variables. Weight relation explicitly represents the relationship between indicator variables and latent variables. In mathematical form, the measurement model can be explained through the following equations [8]:

$$\boldsymbol{z} = \boldsymbol{C}'\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1}$$

where, $\boldsymbol{C}'$ is a latent variable loading matrix with an ordered indicator $j \times p$, $\boldsymbol{z}$ is the vector of the indicator variable $j \times 1$, $\boldsymbol{\gamma}$ is an ordered latent variable vector $p \times 1$, and $\boldsymbol{\varepsilon}$ is a residue vector of an ordered indicator variable $j \times 1$. Whereas, mathematically, the structural model is stated as follows [8]:

$$\boldsymbol{\gamma} = \boldsymbol{B}'\boldsymbol{\gamma} + \boldsymbol{\zeta} \tag{2}$$

where is a matrix of path coefficients between ordered latent variables $p \times p$ and $\boldsymbol{\zeta}$ is the residual vector of all ordered latent variables $p \times 1$.GSCA is part of component-based SEM, where latent variables are defined as components of indicator variables. The relationship between indicator variables and latent variables is explicitly described using weighted relations, as follows [8]:

$$\boldsymbol{\gamma} = \boldsymbol{W}'\boldsymbol{z} \tag{3}$$

where, $\boldsymbol{W}'$ is the component weight matrix of an ordered indi cator $p \times j$. Based on the equations (1), (2), and (3), can be formed into a single GSCA model equation, denoted in the matrix, as follows:

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

$$\begin{bmatrix}\boldsymbol{z}\\ \boldsymbol{W}'\boldsymbol{z}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}'\\ \boldsymbol{B}'\end{bmatrix}\boldsymbol{W}'\boldsymbol{z} + \begin{bmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\zeta}\end{bmatrix}$$

$$\begin{bmatrix}\boldsymbol{z}\\ \boldsymbol{W}'\boldsymbol{z}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}'\\ \boldsymbol{B}'\end{bmatrix}\boldsymbol{W}'\boldsymbol{z} + \begin{bmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\zeta}\end{bmatrix}$$

$$\boldsymbol{V}'\boldsymbol{z} = \boldsymbol{A}'\boldsymbol{W}'\boldsymbol{z} + \boldsymbol{e} \tag{4}$$

Where,

$$\boldsymbol{V}' = \begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{W}'\end{bmatrix}; \boldsymbol{A}' = \begin{bmatrix}\boldsymbol{C}'\\ \boldsymbol{B}'\end{bmatrix}; \boldsymbol{e} = \begin{bmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\zeta}\end{bmatrix}\ ;\ \boldsymbol{V}'\boldsymbol{z}$$

is a matrix of all latent variables and ordered indicators $t \times 1$, $\boldsymbol{A}'$ is the matrix of all the values of loadings and coefficients of the ordered path $t \times p$, and $\boldsymbol{e}$ is the vector of all ordered residues $t \times 1$.

Based on equation (4), there are several parameters, namely $\boldsymbol{V}, \boldsymbol{W}$ and $\boldsymbol{A}$ that cannot be resolved directly. This is due to the presence of zero or fixed elements in these parameters, so it requires a specific approach to determine the appropriate value in further analysis [6].

B. *Parameter Estimation*

The Alternating Least Squares (ALS) method in the GSCA is used to estimate parameters by minimizing the number of residual squares. Parameter estimation $\boldsymbol{A}$ is performed using the formula :

$$\hat{a} = (\boldsymbol{\Omega}'\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}'\boldsymbol{vec}(\boldsymbol{\Psi}) \tag{5}$$

(1)

with an inverse alternative of Moore-Penrose if the matrix $\boldsymbol{\Omega}'\boldsymbol{\Omega}$ is singular. Next, the $\boldsymbol{S}$ parameters are updated using the equation:

$$\widehat{\boldsymbol{\eta}_k} = (\boldsymbol{\Pi}'\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'vec(\boldsymbol{Z\Delta}) \tag{6}$$

where the matrix $\boldsymbol{S}$ is updated based on $\boldsymbol{\eta}_k$ This process is repeated until it converges or reaches tolerance. 10^{-4} . ALS works by updating one parameter while keeping the other parameters fixed until it reaches an optimal solution.

C. *GSCA Model Evaluation*

The evaluation of the GSCA model is carried out in several systematic stages, namely:

Convergent validity is evaluated by considering the estimated value of the loading factor on each latent variable. If the estimated value of the loading factor for each latent variable is greater than 0.5, then the latent variable is considered valid or good [9]. Discriminant validity is evaluated by comparing the square root value of the Average Variance

Extracted (AVE) for each latent variable with its correlation to other latent variables in the model. When the AVE value is greater than 0.5 and the value of $\sqrt{AVE}$ greater than 0.70 than the correlation value between variables, the discriminant validity test is fulfilled [10]. Here is the formula used to calculate AVE :

$$AVE = \frac{\sum_{i=1}^{n} c_i^{\,2}}{\sum_{i=1}^{n} c_i^{\,2} + \sum_{i=1}^{n}(1 - c_i^{\,2})} \tag{7}$$

Where c_i is the loading factor of the indicator i for the related latent variable, and n is the number of indicator variables in a

latent variable. Composite reliability in the measurement of latent variables can be evaluated by internal consistency (ρ_c) and Cronbach's alpha. Based on these two measurements, if the composite reliability value is greater than 0.7, it indicates that the indicator in a construct has a good level of reliability.
The following formula for calculating composite reliability with the value of ρ_c :

$$\rho_c = \frac{(\sum_{i=1}^{n} c_i)^2}{(\sum_{i=1}^{n} c_i)^2 + \sum_{i=1}^{n}(1-c_i^2)} \quad (8)$$

In addition to the calculation of the value of ρ_c, composite reliability can be calculated by hypothesis testing. If the Critical Ratio (CR) value is significant, then the indicator is declared reliable [11]. Significance testing for each indicator loading value can be carried out by a hypothesis test, as follows:
Hypothesis:
$H_0 : c_i = 0$, (loading indicator is nonsignificant)
$H_1 : c_i \neq 0$, (loading indicator is significant)
Test statistics:

$$t_0 = \frac{\hat{c_i}}{SE(\hat{c_i})}$$

Test criteria: reject H_0 when the value $|t_0| > t_{n-1;\frac{\alpha}{2}}$. With the level of significance α , the loading value of each indicator is reliable and significant.

Significance testing for each weighted value of the indicator with a formative relationship can be done by hypothesis testing, as follows:
Hypothesis:
$H_0 : w_i = 0$, (indicator weight is nonsignificant)
$H_1 : w_i \neq 0$, (indicator weight is significant)
Test statistics:

$$t_0 = \frac{\widehat{w_i}}{SE(\widehat{w_i})}$$

Test criteria: reject H_0 when the value $|t_0| > t_{n-1;\frac{\alpha}{2}}$. With the level of significance α , the weighted value of the indicator is reliable and significant. Significance testing for each indicator loading value can be done by a hypothesis test, as follows :

Hypothesis:
$H_0 : b_i = 0$, (path coefficient γ_i is nonsignificant)
$H_1 : b_i \neq 0$, (path coefficient γ_i is significant)
Test statistics:

$$t_0 = \frac{\hat{b_i}}{SE(\hat{b_i})}$$

Test criteria: reject H_0 when the value $|t_0| > t_{n-1;\frac{\alpha}{2}}$, with the level of significance α, the path coefficient is reliable and significant.

The R-squared value in both the GSCA method and the regression analysis indicates the extent to which exogenous latent variables contribute to explaining the variations that occur in endogenous latent variables in the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{n} \quad (9)$$

Where, e_i is the residual i *th* n is sample size.

GFI is a tool used to measure the goodness of the model in the GSCA evaluation. The GFI value is considered good when it is between 0.9 and 1.0 [8]. FIT is a measure used to assess the extent to which the GSCA model is able to explain variations in data. The FIT value is in the range of 0 to 1, where the greater the FIT value, the better the model produced.

$$FIT = \frac{1}{T}\sum_{t=1}^{T} R_t^2 \quad (10)$$

Where R_t^2 is the R-Squared value of each indicator variable or endogenous latent variable and T is the number of indicator variables or endogenous latent variables.

III. RESEARCH METHOD

The type of data used in this study is secondary data obtained from the official publication of the Central Statistics Agency (BPS) of Central Java Province in 2023. The variables used were one exogenous variable, namely the economic dimension with 6 indicators (X), and one endogenous latent variable, namely the health dimension, the education dimension, and HDI with 16 indicators (Y).
The software used to process the data uses GSCA Pro Windows 1.2.1 and the R Studio 4.3.2 program. The stages of analysis carried out are:
1. Identify the background of the problem to be studied as a basis for designing measurement models and structural models.
2. Determine problems based on research background. 3. Determine the purpose of the research based on the problems that have been identified.
4. Conduct observations and literature studies to understand the topic being researched.
5. Collect information and variables related to research. 6. Identify the model under study by creating a path chart to illustrate the relationship between latent variables and their indicators.
7. Guess the value of the model parameter used.
8. Evaluate the measurement model, structural model, and overall value of the fit model
9. Interpret the final model based on the analysis to illustrate the research.

IV. RESULT AND DISCUSSION

Modeling using the Generalized Structured Component Analysis (GSCA) method, equipped involves latent variables and indicator variables. Latent variables are abstract concepts that cannot be measured directly, so they require indicator variables as representations. In contrast, indicator variables can be observed and measured directly to represent latent variables in an analysis. In this study, there is one endogenous latent variable, namely Health, Education, and Human Development Index (HDI), which is the main focus of the analysis. In addition, there are three exogenous latent variables, namely Economics, which play a role as a factor that affects endogenous latent variables. Furthermore, an estimate was made of five path coefficients that illustrate the relationship between latent variables, namely: (1) Economics to Health, (2) Economics to Education, (3) Economics to HDI, (4) Health to HDI, and (5) Education to HDI. The evaluation of the measurement model in GSCA aims to test the statistical significance and estimation of parameters that link each indicator with its latent variables. Based on the output of GSCA Pro Windows 1.2.1 in Table I, the estimated loading value shows that each indicator has a value greater than 0.5 so that it meets the convergent validity criteria

Table I. Estimated Loading Value

Latent Variable	Loading Estimate	Standard Error
EKO3	0,773	0,095
EKO4	0,855	0,050
EKO6	0,756	0,113
KES1	0,729	0,058
KES2	0,557	0,169
KES5	0,869	0,049
KES6	0,843	0,053
PEND1	0,551	0,128
PEND2	0,694	0,136
PEND3	0,720	0,081
PEND4	0,864	0,031
PEND6	0,912	0,035
IPM1	0,798	0,042
IPM2	0,971	0,008
IPM3	0,942	0,026
IPM4	0,876	0,054

Based on the output of GSCA in Table II, each variable obtained an AVE value greater than 0.5 and. AVE. greater than the correlation between other variables, so that the discriminant validity test is met.

Table II. Measurement Model Correlation Value

Laten Variable	EKO	KES	$PEND$	IPM
EKO	1			
KES	0,812	1		
$PEND$	0,754	0,808	1	
IPM	0,887	0,887	0,879	1
$\sqrt{AVE}$	0,796	0,760	0,759	0,899

Based on the output of GSCA in Table III, the values ρc and Cronbach's alpha are greater than or equal to 0.7, so that the composite reliability test is met.

Table III. Output ρ_c and Cronbach's Alpha Measurement Model

Latent Variable	ρ_c	Cronbach's alpha	Results
EKO	0,838	0,7100	Fulfilled
KES	0,842	0,7503	Fulfilled
$PEND$	0,869	0,8117	Fulfilled
IPM	0,944	0,9199	Fulfilled

The assessment of the significance of the parameter coefficient was carried out by looking at the statistical value of t_0, which was obtained from the comparison between the large parameter coefficient and the standard error using the bootstrapping method. If the value $|t_0| > t_{n-1;\frac{\alpha}{2}}$, then the parameter coefficient is considered significant, indicating that the variable has an influence on the model. Based on the output of GSCA in Table 4, the results of the significance test of the estimated load factor value with a significance level of α=5%, namely all indicators show a value of $|t_0|$ greater than the critical limit of 1.96. So it can be concluded that the estimated value of the loading factor in each indicator used is reliable and significant.

Table IV. Value t_0 of Latent Variable Indicators EKO, KES, and PEND

Latent Variable	Loading Estimation	SE	t_0	Results
$EKO3$	0,773	0,095	8,137	Reject H_0
$EKO4$	0,855	0,050	17,100	Reject H_0
$EKO6$	0,756	0,113	6,690	Reject H_0
$KES1$	0,729	0,058	12,569	Reject H_0
$KES2$	0,557	0,169	3,296	Reject H_0
$KES5$	0,869	0,049	17,735	Reject H_0
$KES6$	0,843	0,053	15,906	Reject H_0
$PEND1$	0,551	0,128	4,305	Reject H_0
$PEND2$	0,694	0,136	5,103	Reject H_0
$PEND3$	0,720	0,081	8,889	Reject H_0
$PEND4$	0,864	0,031	27,871	Reject H_0
$PEND6$	0,912	0,035	26,057	Reject H_0

Based on the output of GSCA in Table V, the results of the significance test of the weighted value of the indicator in the formative relationship with the significance level of α=5% are observed, i.e., all indicators show a value of t_0 greater than the critical limit of 1.96. So it can be concluded that the weighted value of the indicator in the formative relationship of each indicator used is reliable and significant.

Table V. t_0 Value of IPM Indicator of Latent Variable

Latent Variable	*Weighted* Estimation	SE	t_0	Results
$IPM1$	0,204	0,023	8,870	Reject H_0
$IPM2$	0,324	0,04	8,100	Reject H_0
$IPM3$	0,310	0,033	9,394	Reject H_0
$IPM4$	0,263	0,027	9,741	Reject H_0

Based on the output of GSCA in Table VI, the results of the significance test of the estimated load factor value of each path coefficient with a significance level of α=5% are observed, namely all indicators show a value of $|t_0|$ greater than the critical limit of 1.96. So it can be concluded that the estimated value of the loading factor in each line coefficient used is reliable and significant.

Table VI. Value of Structural Model Parameter Coefficient

Indicator	Estimasion	SE	t_0	Results
EKO → KES	0,812	0,043	18,884	Reject H_0
EKO → PEND	0,754	0,088	8,568	Reject H_0
EKO → IPM	0,388	0,098	3,959	Reject H_0
KES → IPM	0,284	0,142	2,000	Reject H_0
PEND → IPM	0,356	0,1	3,560	Reject H_0

Based on the output of GSCA Table VII , the results of the analysis show that the HDI variables are influenced by economic, health, and educational factors by 90.9%, while 9.1% are influenced by other variables. In addition, the education variable of 56.9% and the health variable of 65.9% were each influenced by economic factors while the rest were influenced by other variables outside the model.

Table VII. R-Square Value of Structural Model

Indicator	*R-Square*
KES	0,659
$PEND$	0,569
IPM	0,909

Based on the output of GSCA in Table VIII, the results of the analysis show that a FIT value of 0.623 indicates that the model is able to explain 62.3% of the data variations. In addition, a GFI value of 0.969 or 96.9% indicates an overall excellent model fit rate, so that the model can be relied upon in describing the relationship between latent variables and their indicator variables.

Table VIII. Goodness of Fit Model Value

Goodness of Fit Model	Results
FIT	0,623
GFI	0,969

V. CONCLUSION

The results of the evaluation of the measurement, structural, and model fit models show that all indicators in the economic, health, education, and Human Development Index (HDI) variables have met the assumptions of validity and reliability, although there is still a need for improvement in some indicators. The resulting structural model forms three equation models, namely HDI is influenced by the economy, health, and education with the model $\gamma_4 = 0{,}388\gamma_1 + 0{,}284\gamma_2 + 0{,}356\gamma_3 + \zeta_4$., Education is influenced by the economy with a model $\gamma_3 = 0{,}754\gamma_1 + \zeta_3$ and health is influenced by the economy with a model $\gamma_2 = 0{,}812\gamma_1 + \zeta_2$. The results of the analysis showed that the variables of economics, health, and education together explained the variability of HDI by 90.9%, while education and health were explained by economics by 56.9% and 65.9%, respectively. The evaluation of the overall fit of the model showed that the model had a good fit rate with a FIT value of 0.623 and a GFI of 0.969, and all latent variables were proven to be significant with the validity and reliability that had been met.

REFERENCES

[1] Badan Pusat Statistik. 2024a. *Analisis Kualitas Pembangunan Manusia Provinsi Jawa Tengah 2023* (Badan Pusat Statistik (ed.)). Badan Pusat Statistik.

[2] UNDP. 2022.. *Human Development Report 2021/2022*.

[3] Jasasila. 2020. Analisis Dimensi Pembentuk Indeks Pembangunan Manusia (IPM) Provinsi Jambi 2010-2019. *Ekonomis: Journal of Economics and Business*, *4*(2), 389–396.

[4] Badan Pusat Statistik. (2024b). *Indeks Pembangunan Manusia 2023*. Badan Pusat Statistik.

[5] Fitriani, F., Rusgiyono, A., & Windiharih, T. 2020. Penerapan Metode Generalized Structured Component Analysis pada Kepuasan Konsumen (Studi Kasus: Pasien Klinik Q). *Jurnal Gaussian*, *9*(4), 454–463.

[6] Brigo, D., Dalessandro, A., Neugebauer, M., & Triki, F. 2008. A Stochastic Processes Toolkit for Risk Management. *SSRN Electronic Journal*.

[7] Dmouj, & Abdelmoula. 2006. *Stock price modelling: Theory and Practice.* Amsterdam: BMI Paper.

[8] Hwang, H., & Takane, Y. 2004.. Generalized Structured Component Analysis. *Psychometrika*.

[9] Ghozali, I. 2008. *Generalized Structured Component Analysis (GSCA) Model Persamaan Struktural Berbasis Komponen*. Badan Penerbit Universitas Diponegoro.

[10] Kusumadewi, K. A., & Ghozali, I. 2013. Generalized Structured Component Analysis (GeSCA) Model Persamaan Struktural Berbasis Komponen. Badan Penerbit Universitas Diponegoro.

[11] Ngatno. 2019. Monograf Analisis Data Penelitian dengan Program GeSCA (2nd ed.). Penerbit Undip Press.

[12] Rizkia, A. D., Ispriyanti, D., & Sugito, S. 2022. Pengaruh Kualitas Layanan Dan Citra Merek Terhadap Kepuasan Pengguna Youtube Premium Menggunakan Partial Least Square. *Jurnal Gaussian*, *11*(3), 323–331.

[13] Hwang, H. 2009. Regularized Generalized Structured Component Analysis. *Psychometrika*, *74*(3), 517–530.

[14] Hair, Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2019. *Multivariate Data Analysis.* Eighth Edition.

[15] Thompson, P. A., Bishop, D. V. M., Eising, E., Fisher, S. E., & Newbury, D. F. 2020. Generalized Structured Component Analysis in candidate gene association studies:

Applications and limitations. *Wellcome Open Research*, *4*, 1–15.

[16] Todaro, M. P., & Smith, S. C. 2020. Economic Development. Thirteenth Edition. In *Pearson* (13th ed., Issue 13th Edition). Pearson Education Limited.