BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1955–1962
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1955-1962

PROVINCIAL SEGMENTATION IN INDONESIA: EXPLORING FACTORS INFLUENCING EDUCATION WITH SEM-PLS METHOD, INCORPORATING MODERATION EFFECTS AND FIMIX-PLS APPROACH

Davina Shafa Vanisa [1], Tentri Ryan Rahmanita[2], Elly Ana[3*], Ardi Kurniawan[4]

[1,2,3,4]*Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Surabaya, 60115, Indonesia

*Corresponding author's e-mail: * **elly-a@fst.unair.ac.id***

Article History:

Received: 14th March 2024
Revised: 22nd April 2024
Accepted: 15th July 2024
Published: 1st September 2024

Keywords:

Provincial Segmentation; FIMIX-PLS; Moderation; Indonesian Education.

ABSTRACT

The significance of education as a developmental metric is underscored by its designation as the 4th goal in the SDGs, which emphasizes ensuring inclusive, equitable, and high-quality education while also expanding lifelong learning opportunities for all. This research relies on two primary sources: secondary data from publications by the Indonesian Central Statistics Agency (BPS RI) in 2023 and the BPS website. The educational variables examined in this study are believed to be influenced by latent variables, including school performance, infrastructure, and poverty levels. Employing the Finite Mixture Partial Least Squares (FIMIX-PLS) approach, the research identified 13 valid and reliable indicators of educational variables. It delineated three regional groups based on the lowest BIC and CAIC values. In this structural equation research, the moderation effect is seen in the significance of the indirect relationship, especially the influence of Regional Poverty on Education with School Outcomes as a moderating construct.

How to cite this article:
D. S. Vanisa, T. R. Rahmanita, E. Ana and A. Kurniawan., "PROVINCIAL SEGMENTATION IN INDONESIA : EXPLORING FACTORS INFLUENCING EDUCATION WITH SEM-PLS METHOD, INCORPORATING MODERATION EFFECTS AND FIMIX-PLS APPROACH," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1955-1962, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Quality education certainly affects quality in all fields. It plays an important role in realizing productive, dynamic, skilled, knowledgeable, and technological human resources supported by global and industrial talent cooperation [1]. The primary focus of the SDGs revolves around fostering sustainable development by prioritizing well-being, environmental regulations, economic prosperity, and academic progress [2]. The importance of education as an indicator of development is also evidenced by the point that education is the 4th goal in the SDGs, namely, ensuring that the quality of education is inclusive and equitable and that learning opportunities throughout life are increased [3]. According to [4], the economic and social aspects of the SDGs are the most important for developing nations. With the support of educational progress, it is possible to achieve the acceleration of national development in the future; besides that, moral and moral education is remembered to increase the element of morality.

Law No. 20/2003 on the National Education System states that the education budget allocation in the APBD is at least 20%. The allocated amount indicates that the government has pursued optimization in advancing education. However, education equity has not been fully achieved. According to [5], education inequality includes two aspects: the quality of education, which provides access distribution of schools, and the ratio of schools to the population. The results of Susenas 2023 show that 6.93% of the population aged 13-15 years have never been to school, and 21.61% of the population aged 16-18 years have never been to school. In addition, based on BPS data, the illiteracy rate is also relatively high, especially in 4 provinces in Indonesia, reaching above 6%, including Papua, West Nusa Tenggara, East Java, and South Sulawesi. In terms of the quality of human resources, based on the Human Development Index (HDI) data released by UNDP, Indonesia is ranked 114th in the world, below four other ASEAN member countries, with quite far-adrift rankings. Namely, Singapore ranked 12th, Brunei Darussalam ranked 51st, Malaysia ranked 62nd, and Thailand ranked 66th.

Some previous research on Education Indicators, [6] conducted research with the title "Application of K-Medoids in Provincial Clustering in Indonesia based on High School Level Education Indicators". The study concluded that Papua and West Papua Provinces have several aspects of education that could be better. It aligns with research conducted by [7] titled "Comparison of K-Means and AG K-Means Methods in Clustering Provinces in Indonesia based on EducationIndicators". The study concluded that Papua Province must make evaluations and improvements to increase the value of education indicators. Another study conducted by [8] concluded that the results of cluster analysis could show that provinces in Indonesia have different characteristics, especially in terms of uneven development in the field of education, which is indicated by the difference in the average value of each variable in clusters that are geographically located in the western, central and eastern regions of Indonesia.

Based on previous research, no research has yet analyzed the factors influencing education in Indonesia using the Structural Equation Modeling (SEM) method with the Finite Mixture Partial Least Square (FIMIX-PLS) approach. SEM has more capabilities in solving complicated problems, namely being able to estimate the relationship between variables that are multiple relationships with output in the form of a measurement model from several indicators and a structural model composed of several constructs (latent variables) [9]. The Partial Least Square method is used because it has several advantages, including the data does not have to have a specific measurement scale, the number of samples does not have to be significant, and the data does not have to have a normal multivariate distribution [10].

Grouping objects into several groups is done based on a measure of similarity or common characteristics between objects. In grouping, objects are often found that cannot be measured directly because they have quantitative values. These objects are called latent variables. The method that can be used to group latent variables based on indicator variables is the Finite Mixture Partial Least Square method developed by [11] which produces Segments with more homogeneous members based on the relationship between latent variables in the structural equation model. Therefore, based on the introduction that has been presented, there has yet to be any research that discusses the analysis of factors that influence education in Indonesia using SEM-PLS with the FIMIX-PLS approach. This study aims to analyze the factors that affect education in Indonesia and examine the best education segmentation in each province in Indonesia using FIMIX-PLS analysis so that the results can be used to implement an education equality program in Indonesia.

2. RESEARCH METHODS

2.1 Structural Equation Modelling in Partial Least Square

This research uses Structural Equation Modelling (SEM) methods and data analysis. SEM is a statistical technique that can analyze the relationship pattern between latent construct variables and their indicators and determine the relationship pattern of latent construct variables with each other. According to [12], SEM is a type of multivariate analysis that can analyze relationships between variables in a more complex manner. The SEM method can be divided into covariance-based SEM (CB-SEM) and variance-based SEM or Partial Least Square (PLS-SEM). Partial Least Square (PLS) is a multivariate statistical technique that can handle many response variables (dependent) and explanatory variables (independent) at once [13]. PLS is an analytical method that is considered powerful because it is not based on various assumptions [14]. This method also does not require normally distributed data and can be used for small sample sizes [15].

2.2 Bootstrap

The bootstrap method is a non-parametric statistical procedure used to test significance in SEM PLS. The bootstrap method uses resampling with replacement to obtain standard errors in hypothesis testing. The bootstrap method is a tool to help reduce unreliability by forming shadow data whose characteristics are very similar to the original data. The resulting values are Path Coefficients and $\boldsymbol{R^2}$ values [16]. The higher the $\boldsymbol{R^2}$ value, the better the predictive ability of the variables in the model [17]. [18] explain that in theory there are 3 categories of $\boldsymbol{R^2}$ value limits, namely above 0.67 is categorized as substantial, meaning that the contribution of exogenous variables to endogenous variables is strong. $\boldsymbol{R^2}$ between 0.33 and 0.67 is categorized as moderate, meaning that the contribution of exogenous variables to endogenous variables is sufficient or moderate. Meanwhile, $\boldsymbol{R^2}$ between 0.19 and 0.33 is categorized as weak.

2.3 Finite Mixture in Partial Least Square

Clustering methods can be applied in PLS models to overcome heterogeneity, but these methods cannot be applied to models that contain latent variables or variables that cannot be measured directly. In FIMIX-PLS, the statistical measure used to indicate the best number of Segments is the selection of Segments based on several criteria where researchers can compare the Consistent AIC (CAIC) and Bayesian Information Criterion (BIC) values [19]. The selection of the best segment class based on CAIC and BIC values is seen from the smallest value when comparing segment class values [20]. The assumption in FIMIX-PLS is that if the observation units are separated according to their strata, there will be no cases of structural heterogeneity in the model [21].

2.4 Research Data Source

The data used in this study are secondary data obtained from the Indonesian Central Statistics Agency (BPS RI) publication in 2023 and the BPS website. The data is cross-section data, where the population and samples used are variables that have a significant relationship with Indonesian Education 2023 in all provinces in Indonesia. So, the number of observations in this study is 34.

Research variables are anything in the form of anything set by researchers to study or study so that information is obtained about it and conclusions are drawn [22]. This study consists of endogenous latent variables, namely Indonesian education with three indicators, and exogenous latent variables, namely school Outcome, Facilities and Infrastructure, and regional poverty.

Table 1. Research Variable

Latent Variable	Sub Variable	Indicator
Poverty	Severity	Poverty Severity
	Residents	Poor Population
	Depth	Depth of Poverty

Latent Variable	Sub Variable	Indicator
School Outcomes	Illiteracy-Rate	Illiteracy Rate
	Dropout	High school dropout rate
	Not-School	Number of children out of school
Facilities And Infrastructure	Student-Class	Student to Class Ratio
	Student-School	Student to School Ratio
	Class-Good	Class Ratio Good Condition of School
	Library	School Library Ratio

Latent Variable	Sub Variable	Indicator
Education	Average-Length	Average Years of Schooling
	Expectations	Expected Years of Schooling
	Literacy	Literacy Rate

3. RESULTS AND DISCUSSION

To analyze the influence of exogenous variables on endogenous variables, creating a path diagram explaining the relationship pattern between latent variables and their indicators was necessary.

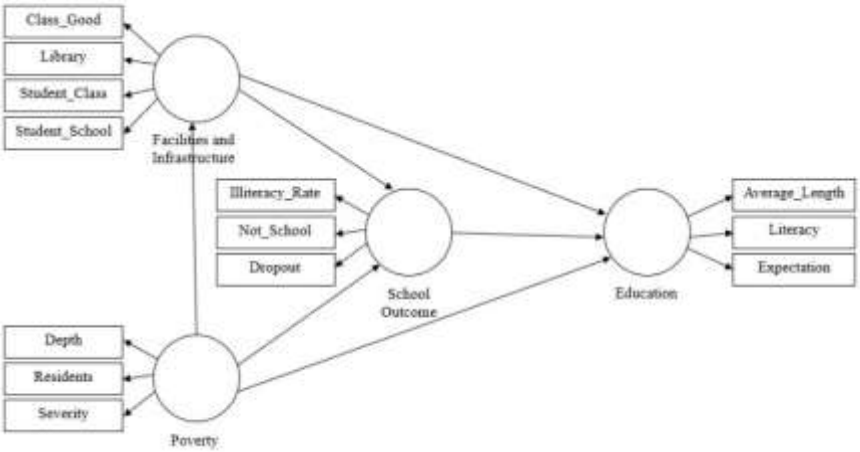

Figure 1. Initial conceptual model

Based on **Figure 1**, there are four latent variables consisting of 3 exogenous latent variables and one endogenous latent variable. The exogenous variables in this research are independent variables, such as school outcome, Facilities and Infrastructure, and regional poverty. The endogenous variable in this study is the dependent variable, namely education. The theoretical research above illustrates the relationship between education and the aspects that influence it. The relationship between these variables is presented in the path diagram with each factor loading value, as shown in **Figure 2**.

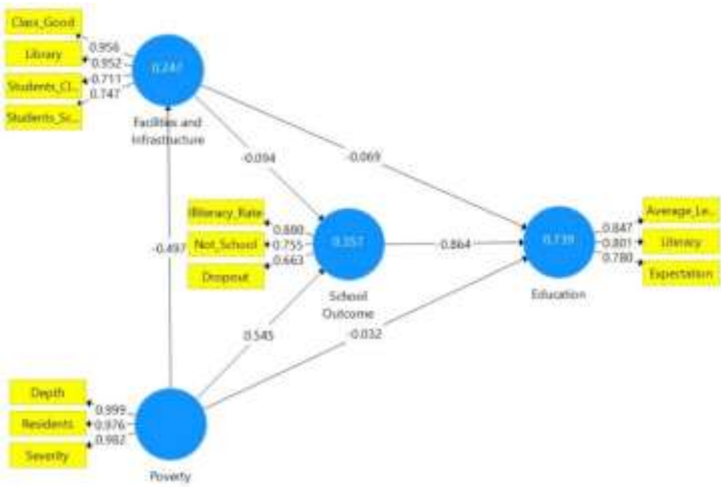

Figure 2. Path diagram construction model

According to Figure 2, the factor loading value of the thirteen items is more than 0.6. Based on the path diagram in Figure 2, each indicator in the four aspects of Facilities and Infrastructure is more than 71%, which means that the Facilities and Infrastructure variable can explain the variance of the four indicators by more than 71%. In the poverty aspect, each of the three indicators can explain the variance of the three indicators by more than 97%. Then, for the three aspects of school Outcome, it is more than 66%, which means that the school outcome variable can explain the variance of the three indicators more than 66%. In the educational aspect, the educational variable can also explain the variance of the three indicators by more than 78%. All latent variables have explained the variance of each indicator that measures above 60%. It shows that the convergent validity of the latent variables is fulfilled.

3.1 Outer Model

We evaluated the measurement model through outer loading testing to see the significance of each indicator. The results of the outer loading test are described in Table 2.

Table 2. Outer Loadings

	School Outcome	Poverty	Education	Facilities And Infrastructure	Information
Illiteracy_Rate	0.880				Valid
Dropout	0.663				Valid
Not_School	0.755				Valid
Depth		0.999			Valid
Severity		0.982			Valid
Residents		0.976			Valid
Expectations			0.780		Valid
Literacy			0.801		Valid
Average_Length			0.847		Valid
Class_Good				0.956	Valid
Student_Class				0.711	Valid
Student_School				0.747	Valid
Library				0.952	Valid

Based on the output of the loading factor value for the variable indicators of poverty, education, school outcome, facilities, and infrastructure, it can be seen that the loading factor values are > 0.6 for all indicators. It indicates that the indicators used are convergently valid. Furthermore, Table 3 shows the results of the validity testing.

Table 3. Validity testing

	Average Variance Extracted (AVE)	Information
Poverty	0.971	Valid
Facilities And Infrastructure	0.721	Valid
Education	0.656	Valid
School Outcome	0.595	Valid

Table 3 shows that the AVE values for the four latent variables are above 0.5, meaning that the four variables of poverty, Facilities and Infrastructure, school outcome, and education are categorized as valid. Next, the reliability test results are given in Table 4.

Table 4. Reliability testing

	Cronbach's Alpha	Composite Reliability	Information
Poverty	0.985	0.990	Reliable
Facilities And Infrastructure	0.891	0.910	Reliable
Education	0.743	0.851	Reliable
School Outcome	0.655	0.813	Reliable

Table 4, shows that the composite reliability value for the four latent variables already has a value above 0.7. So that all indicators used are reliable.

3.2 Inner Model

Evaluation of the inner model is done by bootstrapping with evaluation of the coefficient of determination R^2, the value of the t statistic, and the parameter coefficients are given in Table 5.

Table 5. Coefficient of determination

Endogenous Variable	R-Square	R-Square Adjusted
Education	0.739	0.713

Based on Table 5, the R-Square value for the education variable is 0.713, which means that the education variable can be explained by 71.3% of all exogenous construct variables. In comparison, the other 28.7% is influenced by factors not included in this research model. Furthermore, the bootstrapping results are given in Table 6.

Table 6. Significance of the relationship in the direct structural model

	Original Sample (O)	T Statistics (IO/STDEV)	P-Value	Information
School Outcome → Education	-0.864	6.009	0.000	Significant
Poverty → School Outcome	0.545	1.739	0.042	Significant
Poverty → Facilities and Infrastructure	-0.497	3.878	0.000	Significant

Based on Table 6, it can be concluded that school results hurt education with a coefficient of -0.864 and a significant p-value of 0.000. Then, poverty positively affects school achievement, as shown by a coefficient of 0.545 and a significant p-value of 0.042. Furthermore, poverty hurts Facilities and Infrastructure, as indicated by a coefficient of -0.497 and a significant p-value of 0.000. All significant values obtained were smaller than the alpha level of 0.05.

Table 7. Significance of the relationship in the indirect structural model (moderation)

	Original Sample (O)	T Statistics (IO/STDEV)	P-Value	Information
Poverty → School Outcome →Education	-0.476	1.662	0.049	Significant

Based on Table 7, it can be concluded that poverty negatively influences education through school results, as indicated by a coefficient of -0.476 and a p-value of 0.049, which is smaller than the alpha level of 0.05.

3.3 Clustering Analysis with Finite Mixture-Partial Least Square

Clustering is carried out with FIMIX-PLS by determining the best Segmentation value based on the BIC and CAIC values in Table 8.

Table 8. BIC and CAIC value criteria

Fit Indices	K = 2	K = 3	K = 4	K = 5	K = 6
BIC	248.048	228.904	251.767	237.377	247.073
CAIC	267.048	257.904	290.767	286.377	306.073

Based on Table 8, a comparison of fit indices is obtained. At k = 3, it has the lowest BIC and CAIC values, 228.904 and 257.904. So, it is concluded that the best Segment is 3 Segments. Education structure grouping is obtained based on the value of each Segment membership probability divided into 3 Segments. The number of Segment members in each Segment can be seen in Table 9, based on the percentage of segment size.

Table 9. Segment Size

	Segment 1	Segment 2	Segment 3
%	0.577	0.306	0.117

The overall members of each Segment are given in **Table 10**; each Segment is considered homogeneous because it has the same characteristics.

Table 10. Provincial grouping

Segment	Provincial
1	Aceh, Riau, Bengkulu, West Java, Central Java, East Java, Banten, East Nusa Tenggara, West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, Maluku North, West Papua
2	South Sumatra, Lampung, Bangka Belitung Islands, Riau Islands, DKI Jakarta, DI Yogyakarta, Bali, West Nusa Tenggara, South Kalimantan, South Sulawesi
3	North Sumatra, West Sumatra, Jambi, Papua

Table 11. Path coefficient of segmentation

Relationship between Latent Variable	Path Coefficient		
	Segment 1	Segment 2	Segment 3
School Outcome and Education	-0.857	0.697	14.678
Poverty and School Outcome	0.216	-0.481	0.441
Poverty and Facilities and Infrastructure	-0.267	-0.844	-0.987

Based on **Table 11**, the effect of the school outcome variable on education has the most significant impact in the first Segment consisting of 20 provinces. In the second Segment, the influence of the poverty variable on the Facilities and Infrastructure variable is greater than the influence of other variables. In the third segment, school outcome on education also has the most significant impact on the third segment compared to different variables.

4. CONCLUSIONS

According to the research findings, 13 significant indicators within the model structure influence education, indicating each indicator's ability to elucidate the latent variable. Within the structural equation framework, three significant relationship paths emerge: school outcome hurts education, regional poverty positively impacts school outcome, and the relationship between regional poverty and facilities and infrastructure. The moderation effect obtained is the significance of the indirect relationship. Namely, Regional Poverty hurts Education with School Outcomes as a moderating construct.

Utilizing the FIMIX-PLS method and considering BIC and CAIC criteria, the grouping yielded optimal outcomes, resulting in three segments. The first segment encompasses 57.7% of Indonesia's total provinces, with the second and third segments accounting for 30.6% and 11.7% of the total provinces, respectively. The analysis reveals that in the first segment, the influence of school outcome variables on education outweighs other factors. Conversely, in the second segment, the influence of poverty on Facilities and Infrastructure is more pronounced. Lastly, in the third segment, the impact of school outcome variables on education precedes other variables. In further research, adding more variables using different methods for more specific results is recommended.

REFERENCES

[1] M. R. Putri, G. S. Nugraha, and R. Dwiyansaputra, “Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering,” *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 7, no. 1, pp. 76–83, 2023.

[2] M. Saini, E. Sengupta, M. Singh, H. Singh, and J. Singh, “Sustainable Development Goal for Quality Education (SDG 4): A study on SDG 4 to extract the pattern of association among the indicators of SDG 4 employing a genetic algorithm,” *Educ Inf Technol (Dordr)*, vol. 28, no. 2, pp. 2031–2069, 2023.

[3] Bappenas, “Pendidikan Berkualitas,” Bappenas. Accessed: Feb. 06, 2024. [Online]. Available: https://sdgs.bappenas.go.id/17-goals/goal-4/

[4] R. Bali Swain and F. Yang-Wallentin, “Achieving sustainable development goals: predicaments and strategies,” *International Journal of Sustainable Development & World Ecology*, vol. 27, no. 2, pp. 96–106, 2020.

[5] H. Wiratama, “Analisis Persebaran dan Ketersediaan Sekolah Menengah Di Kota Tanjungbalai Tahun 2014,” 2015.

[6] C. F. Ramadani, “Penerapan K-Medoids dalam Penggerombolan Provinsi di Indonesia berdasarkan Indikator Pendidikan Jenjang SMA”.

[7] M. S. Putri, “Perbandingan Metode K-Means dan AG K-Means dalam Penggerombolan Provinsi di Indonesia Berdasarkan Indikator Pendidikan”.

[8] W. I. Putri, “Analisis Terhadap Indikator–Indikator yang Mencirikan Standar Nasional Pendidikan Sekolah Menengah Pertama di Indonesia”.

[9] E. D. Anggita, A. Hoyyi, and A. Rusgiyono, “Analisis Structural Equation Modelling Pendekatan Partial Least Square dan Pengelompokan dengan Finite Mixture PLS (FIMIX-PLS)(Studi Kasus: Kemiskinan Rumah Tangga di Indonesia 2017),” *Jurnal Gaussian*, vol. 8, no. 1, pp. 35–45, 2019.

[10] I. Ghozali, “Structural Equation Modelling Metode Alternatif dengan Partial Least Square Semarang,” *Universitas Diponegoro*, 2011.

[11] C. Hahn, M. D. Johnson, A. Herrmann, and F. Huber, “Capturing customer heterogeneity using a finite mixture PLS approach,” *Schmalenbach Business Review*, vol. 54, pp. 243–269, 2002.

[12] J. F. Hair, M. Sarstedt, and C. M. Ringle, “Rethinking some of the rethinking of partial least squares,” *Eur J Mark*, vol. 53, no. 4, pp. 566–584, 2019.

[13] T. Nur, “Pengaruh growth opportunity, profitabilitas dan struktur modal terhadap nilai perusahaan dengan dividen sebagai variabel intervening pada perusahaan manufaktur yang terdaftar di BEI pada periode 2014-2017,” *Jurnal Manajemen Dan Bisnis Indonesia*, vol. 5, no. 3, pp. 393–411, 2018.

[14] L. K. Harahap and M. Pd, “Analisis SEM (Structural Equation Modelling) dengan SMARTPLS (partial least square),” *Fakultas Sains Dan Teknologi Uin Walisongo Semarang*, vol. 1, no. 1, pp. 1–11, 2020.

[15] E. R. CHAIRANI, “Analisis Faktor Penentu Keputusan Konsumen Dalam Pembelian Produk Kecantikan Halal Dengan Metode Analisis Structural Equation Modeling-Partial Least Square (Sem-Pls),” 2022.

[16] J. F. Hair, C. M. Ringle, M. Sarstedt, and G. , T. M. Hult, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, vol. 3. Sage Publishing, 2021.

[17] A. Ramirez-Orellana, M. del Carmen Valls Martinez, and M. S. Grasso, “Using higher-order constructs to estimate health-disease status: the effect of health system performance and sustainability,” *Mathematics*, vol. 9, no. 11, p. 1228, 2021.

[18] J. Henseler, C. M. Ringle, and R. R. Sinkovics, “The use of partial least squares path modeling in international marketing,” in *New challenges to international marketing*, vol. 20, Emerald Group Publishing Limited, 2009, pp. 277–319.

[19] M. Sarstedt, J.-M. Becker, C. M. Ringle, and M. Schwaiger, “Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments?,” *Schmalenbach Business Review*, vol. 63, pp. 34–62, 2011.

[20] E. E. Rigdon, C. M. Ringle, and M. Sarstedt, “Structural modeling of heterogeneous data with partial least squares,” *Review of marketing research*, pp. 255–296, 2010.

[21] B. W. Otok, R. Sriningsih, and D. S. Dila, “Segmentation of toddler nutritional status using REBUS and FIMIX partial least square in Southeast Sulawesi,” *MethodsX*, vol. 12, p. 102515, 2024.

[22] Sugiyono, “Statistika untuk Penlitian,” *Alfa Beta*, 2019.