BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1515–1526
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1515-1526

MULTIVARIATE MULTILEVEL MODELLING TO ASSESS FACTORS AFFECTING THE QUALITY OF VOCATIONAL HIGH SCHOOLS IN SOUTH SULAWESI PROVINCE

Abdullah Pannu[1], **Hari Wijayanto**[2*], **Budi Susetyo**[3]

[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Jl. Raya Dramaga, Babakan, Dramaga 16680, Bogor, Indonesia

Corresponding author's e-mail: [2*] *hari@apps.ipb.ac.id*

Abstract. This study analyzes the quality of Vocational High Schools (VHS), which have a hierarchical data structure and have more than one response variable. Data gathered for this study are from the Basic Education Data (DAPODIK) in the form of raw data variables of several variables that characterize the quality of VHS and other independent variables in South Sulawesi for four years (2018 to 2021) from the Ministry of Finance Republic of Indonesia (KEMENKEU), and Statistics Indonesia (BPS). The explanatory variable at the regency level consists of four years (2018 to 2021), a multi-year and high-dimensional data structure. Therefore, Principal Component Analysis (PCA) is used to overcome this. The modelling is done by using multivariate multilevel modelling (MVMM) on one-level and two-level structures. This study aims to model the average National Examination and Accreditation scores of VHS in South Sulawesi using MVMM modelling that considers the regency/city level and identifies the factors that influence the average National Examination and Accreditation scores. The results showed that the two-level multivariate model with a random intercept as a hierarchical component was better than the one-level multilevel model based on a minor Deviance Information Criterion (DIC) value. Simultaneously, at the 5% level of significance, variables that contribute significantly to the quality of VHS in South Sulawesi Province are produced. The variables that have a significant effect on the quality of VHS at the school level are the ratio of the number of students/pupils per study group, the percentage of certified teachers to the number of teachers, the ratio of the number of students/pupils per number of toilets, the ratio of laboratory availability, and the ratio of the availability of supporting rooms. Meanwhile, at the regency level, it was found that the percentage of poverty and Gross Regional Domestic Product (GRDP) had a significant effect on the quality of VHS.

Keywords: *hierarchy, multivariate multilevel, PCA, random effect.*

Article info:

Submitted: 4th September 2022 *Accepted: 26th November 2022*

How to cite this article:
A. Pannu, H. Wijayanto and B. Susetyo, "MULTIVARIATE MULTILEVEL MODELLING TO ASSESS FACTORS AFFECTING THE QUALITY OF VOCATIONAL HIGH SCHOOLS IN SOUTH SULAWESI PROVINCE", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1515-1526, Dec., 2022.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Copyright © *2022* Author(s)

1 INTRODUCTION

Research development in various disciplines often involves a hierarchical data structure. The hierarchy population and sample data indicate levels in the data. In the hierarchical data structure, individuals in the same group tend to have the same characteristics, so observations at lower levels are not independent [1]. Modelling without paying attention to the hierarchy structure will violate the freedom assumption so that the conclusions obtained are less precise [2]. The statistical method that can solve this problem is multilevel modelling. This model can solve all the problems in the hierarchical data structure. Multilevel regression is one type of modelling that considers the impact of each data level on the diversity of responses [3]. In addition, multilevel regression can measure interactions between explanatory variables at different levels, response variables at the lowest level, and explanatory variables at each level [4].

In some research cases, the number of response variables in regression have more than one response variable, which in this regression analysis is called Multivariate Regression. The effect of several independent variables on multiple response variables simultaneously has not been able to be accommodated in a single variable multilevel modelling. Therefore, to overcome these problems, Multivariate Multilevel Modelling (MVMM) is used. Multivariate multilevel modelling is the development of multivariate analysis on hierarchy data where the response variable is more than one and between response variables has a correlation or covariance [5]. The authors in [6] mentioned the advantages of using a multivariate multilevel analysis approach compared to a single variable multilevel analysis, namely: (i) MVMM may compare the impact of explanatory variables on each dependent variable, (ii) derive inferences about the correlations between response variables at the individual and group levels, and (iii) MVMM can simultaneously test the significance of all response variables. Research related to MVMM, among others, by the authors in [7] used multilevel modelling of pupils achievement in the case studies of Trends in International Mathematics and Science Studies and Progress in International Reading Literacy Studies in Italy. The authors in [8] used MVMM to assess the achievement of students' cognitive abilities in the 2019 action survey.

In the field of education, the factors that influence the quality of Vocational High Schools (VHS) can be investigated. Vocational quality characteristics can be reflected through the achievement of students' National Examination (NE) scores in education units and school unit accreditation scores. These characteristics tend to be the same within regency/city and differ between regency/city. The quality of VHS is influenced by internal school factors that reflect the characteristics of the school, for example, the number of students, the adequacy of teachers, the percentage of graduate students, the number of study groups, the number of administrative staff, and the completeness of school facilities and infrastructure [9]. In addition, the quality of VHS is influenced by external factors, namely school location, community environment, regional economic variables, absorption of VHS graduates in the business/industry area, the unemployment rate for VHS graduates, and local government policies. Internal factors can be measured at the school level, and external factors are measured at the regency/city level so that, in this context, the data is obtained using a hierarchical structure. If the effect of regency/city is ignored, it can cause heteroscedasticity to appear in the residual, so it is more appropriate to use multilevel analysis than multiple regression analysis [10], [11].

VHS is a vocational school that plays a strategic role in developing human resources in the business/industry area [12]. However, the absorption capacity of VHS graduates in regional development has not been optimal. The open unemployment rate (TPT) for VHS graduates is still the highest compared to other educational TPT [13]. Data released by Statistics Indonesia [14] states that TPT in Indonesia is still dominated by VHS graduates for the last three years, namely 10.36%, 13.55%, and 11.13%, respectively. The same condition occurred in South Sulawesi Province, where TPT is dominated by the workforce of VHS graduates for the last three years, namely 9.06%, 10.96%, and 11.34%, respectively [15]. The high unemployment of VHS graduates is caused by the low quality of VHS graduates compared to other levels of education, which triggers strategic issues in the education sector. Departments, facilities, teachers, and the workforce are only a few of the variables affecting VHS quality; other factors also include the interaction between the government, educational institutions, and businesses [16].

This phenomenon is interesting to be studied further to see the school's internal and external factors that affect the quality of VHS. The quality of VHS can be seen from two points of view, namely, the achievement of students' NE scores in schools and the school accreditation scores. Multivariate multilevel regression with two response variables can be applied to this case. Therefore, this study will examine the application of Multivariate Multilevel Modelling to assess factors affecting the quality of Vocational High Schools in South Sulawesi Province. This study can pinpoint factors that impact VHS quality, particularly in

South Sulawesi, so that the government can utilize the findings as a standard to raise the quality of VHS. Additionally, high-dimensional data were employed in this study, so researchers could combine the MVMM and Principal Component Analysis (PCA) approaches to get more precise parameter estimations.

2 RESEARCH METHODS

2.1. Data

The data used in this study consist of 169 Public Vocational Schools from the Basic Education Data (DAPODIK) in the form of raw data variables of several variables that characterize the quality of VHS and other independent variables in South Sulawesi from 2018 to 2021 from Ministry of Finance Republic of Indonesia (KEMENKEU), and Statistics Indonesia (BPS). Table 1 shows the response in this study, the explanatory variable at the school level, and the explanatory variable at the regency/city level.

Table 1. School Level Variable and Regency/City Level Variable

Variable	Information
	School Level Variable
Y_1	Average National Exam Score
Y_2	Average School Accreditation Final Score
X_1	Percentage of graduate students/pupils
X_2	Percentage of students/pupils who drop out (DO)
X_3	The ratio of the number of students/pupils per study group
X_4	The ratio of the number of students/pupils per the number of teachers
X_5	Percentage of certified teachers
X_6	Percentage of undergraduate teachers
X_7	The ratio of administrative staff per number of groups
X_8	The ratio of the number of classrooms per group
X_9	The ratio of number of computers per number of students/pupils
X_{10}	The ratio of number of students/pupils per number of toilets
X_{11}	The ratio of laboratory availability
X_{12}	The ratio of support space availability
	Regency/City Level Variable
Z_1	Labor Force Participation Rate (TPAK)
Z_2	Education Budget
Z_3	Poverty Percentage
Z_4	Human Development Index (HDI)
Z_5	Net Participation Rate (NER)
Z_6	Assistance Budget of Smart Indonesia Program (PIP) for Vocational High Schools
Z_7	Open Unemployment Rate (TPT)
Z_8	Gross Regional Domestic Product (GDP)

2.2. Modelling Procedure

The data analysis procedure was carried out using R version 4.0.0 and SAS software with package library (*nlme*) and library (*lme4*) with the following stages:

a. Exploring data to obtain an overview of the variables to be analyzed.
b. Performing a multivariate regression modelling to detect variables affecting the average NE scores and school accreditation scores.
c. Performing a multivariate regression assumption test to detect any deviations in assumptions caused by multilevel data structures.
d. Checking multicollinearity by examining the correlation coefficient between the explanatory variables and the variance inflation factor (VIF). If multicollinearity exists, then the explanatory variable with the highest VIF value is reduced, and the multivariate regression modelling is performed again.
e. Conducting MVMM modelling to model the relation between the average score of the national examination and the final score of accreditation with the following model:

1). Develop an MVMM Model with a Multivariate Random Intercept Model without involving an intercept-only model or an MVMM.0 model then calculate the intra-class correlation (ICC) value for each response variable.
2). Develop the MVMM1.1 model by inputting the school-level explanatory variable in the MVMM.0 model.
3). Simplify or reduce the dimensions of multiyear data structures using Principal Component Analysis (PCA) on regency/city level variables.
4). Develop a model involving explanatory variables at the school level and regency/city level using the MVMM1.2 Model. Combined school-level model and regency/city-level model.
5). Calculate the deviance value for each model and perform a model comparison test.

f. Choosing the best model through the smallest deviance difference size, which is $diff = D_1 - D_2$.
g. Determining the variables that affect the quality of VHS with simultaneous test and partial test.
h. Interpreting variables that have a major impact on VHS quality.

Multilevel maximum likelihood estimation (MLE) is the model's chosen strategy for parameter estimation [17]. The specification of the model used is the intercept only (MVMM.0) multilevel model, which means the multivariate multilevel model without involving the independent/explanatory variables at the school level and regency level, the school-level multivariate multilevel model (MVMM1. 1), namely the multivariate multilevel model involving independent/explanatory variables at the school level, and (MVMM1.2), namely the general multivariate multilevel model from the school level multivariate multilevel models and the regency level multivariate multilevel model (MVMM1 .2) [5], [10], [18]. Furthermore, it is written as follows.

a. The model of h-th response dummy variable (lowest level) by including dummy variables in the model as follows.

$$Y_{hij} = \pi_{1ij} d_{1ij} + \pi_{2ij} d_{2ij} + \cdots + \pi_{pij} d_{pij} \tag{1}$$

where p is the response variable, Yhij is the response on the h-th measure, where i is individual at school level and j is individual at regency level.

b. School-level multivariate multilevel modelling (MVMM1.1)

$$\pi_{pij} = Y_{hij} = \beta_{p0j} + \sum_{k=1}^{K} \beta_{pkj} X_{kij} + e_{pij} \tag{2}$$

when mean vector 0 and covariance matrix Ω_e of the multivariate normal distribution are present for the $e_{1ij}, e_{2ij}, \dots, e_{pij}$.

c. Regency level multivariate modelling (MVMM1.2)

$$\begin{aligned}
\beta_{10j} &= \gamma_{100} + \gamma_{101} Z_{1j} + u_{10j} \\
\beta_{11j} &= \gamma_{110} + \gamma_{111} Z_{1j} + u_{11j} \\
\beta_{1kj} &= \gamma_{1ok} + \sum_{l=1}^{L} \gamma_{1lk} Z_{lj} + u_{1kj} \\
&\vdots \\
\beta_{20j} &= \gamma_{200} + \gamma_{201} Z_{2j} + u_{20j} \\
\beta_{21j} &= \gamma_{210} + \gamma_{211} Z_{2j} + u_{21j} \\
\beta_{pkj} &= \gamma_{pok} + \sum_{l=1}^{L} \gamma_{pkl} Z_{lj} + u_{pkj}
\end{aligned} \tag{3}$$

where mean vector 0 and covariance matrix $\boldsymbol{\Omega_u}$ of the multivariate normal distribution are present for the $u_{10j}, u_{p0j}, u_{1kj}, \dots, u_{pkj}$, with $h = 1,2, \dots, p$; $i = 1,2, \dots, n_j$; $j = 1,2, \dots, J$; $k = 1,2, \dots, K$; $l = 1,2, \dots, L$

Y_{hij} = response variables for the h-th measurement at the i-th school and j-th regency level
x_{kij} = explanatory variables for the h-th measurement at the i-th school and j-th regency level
β_{p0j}= random intercept of the p-th response variable at the regency-level j-th unit
γ_{p0j}= fixed intercept response variable of the p-th response variable at the regency-level j-th unit

e_{pij} = residual for the i-th unit at the school level in the j-th unit at the regency level, is assumed to be spread out N(0,σ_e^2).
u_{p0j}= random residual in the j-th unit at the regency level, assumed to be spread N(0,σ_u^2).

3 RESULTS AND DISCUSSION

3.1 Characteristics of Regency/City Based on National Examination and Accreditation Scores

This study was conducted to determine the variables affecting VHS quality in South Sulawesi. The VHS quality variables are the National Examination (NE) and Accreditation scores. The data characteristics of the National Examination (Y1) and the accreditation scores (Y2) of VHS of 169 Public VHS in South Sulawesi Province, consisting of 21 regencies and 3 cities from 2018 to 2020, are shown in Figure 1.

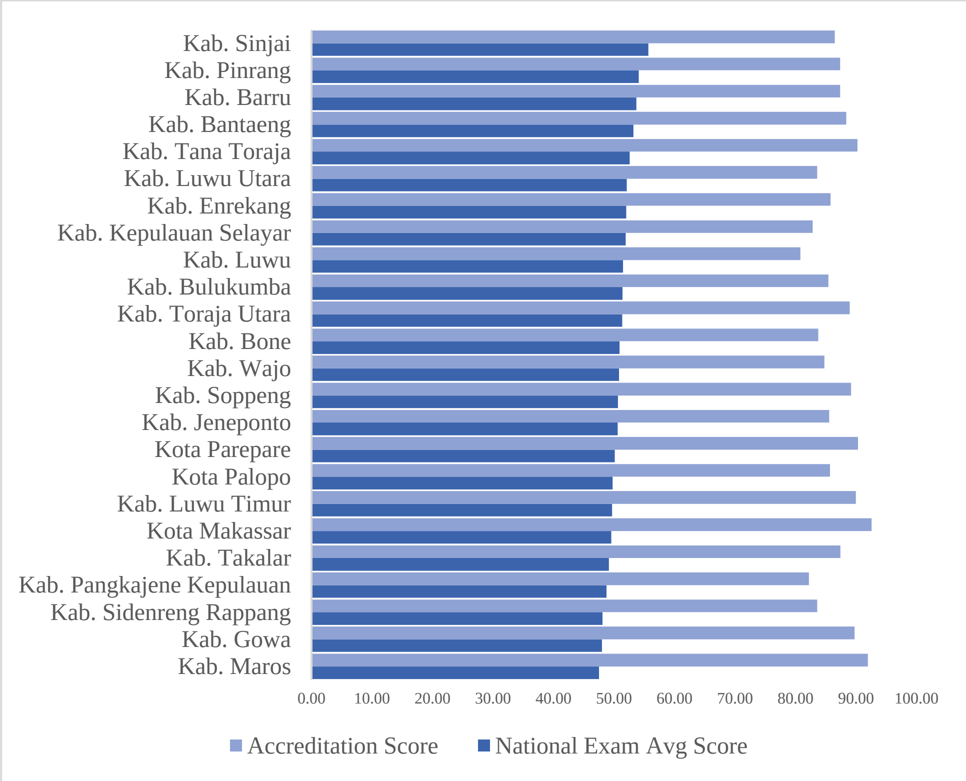

Figure 1. Distribution of National Examination and Accreditation Scores by Regency/City

Figure 1 shows that the highest average national exam score in Sinjai Regency is 64.207 and the lowest national exam score in Maros Regency is 35.02. The school obtained an average NE score of 50.984 with a standard deviation of 5.458. Meanwhile, the highest accreditation score in Makassar City is 99.00, and the lowest score in Luwu Regency is 72.00. In addition, the school received an average accreditation score of 86.156 with a standard deviation of 5.889. The achievement of the national examination scores and the VHS accreditation scores in Figure 1 shows that the characteristics of the average NE scores and the VHS Accreditation scores in South Sulawesi reflect inconsistent characteristics. These are outside with the government's expectation that better-quality schools will have higher accreditation and national examination scores [19]. One possibility is that the accreditation instrument is not yet accurate in measuring the quality of Vocational High Schools. In addition, based on the picture above, the average score for the national examination and the accreditation score in each regency tends to be different. These indicate a hierarchical structure in the VHS quality data, so it is more appropriate to use a multilevel model.

3.2 Multivariate Regression Modelling

In this modelling, regression modelling is carried out to determine the variables that affect the quality of VHS using a multivariate regression model and a multivariate multilevel regression model. The best model will be determined from the two models. This section will discuss the results of multivariate regression modelling and test the classical assumptions of the multivariate regression.

To analyze the occurrence of multicollinearity in multivariate regression analysis, the correlation between independent variables was examined. The results of the calculation show that the ratio variables of the number of students/pupils per group (X3) and the ratio of the number of students/pupils per number of teachers (X4) have a correlation coefficient of -0.76; the percentage of certified teachers to the number of teachers (X5) and the ratio of the availability of supporting space (X12) has a coefficient of 0.61. These variables indicate the existence of multicollinearity. However, the VIF value of the four variables is less than 10, so it can be stated that there is no multicollinearity [20]. Furthermore, a multivariate regression model was built to see the pattern of variable relationships simultaneously and the basis for comparison of subsequent modelling.

Model (4) is a model of the estimated impact of independent factors on the NE score, and Model (5) is a model of the estimated impact of independent variables on the accreditation score (Y2), which is described as follows.

$$\widehat{Y_{1i}} = 50.9841 + 0.0629\,X_{1i} - 0.0643\,X_{2i} + 0.4173X_{3i} + 0.7796X_{4i} - 1.2953X_{5i} + 0.4476X_{6i} - 0.4689X_{7i} - 0.1844X_{8i} + 0.7782X_{9i} - 0.2862X_{10i} - 0.4709X_{11i} + 0.6257X_{12i} \quad (4)$$

$$\widehat{Y_{2i}} = 86.1557 + 0.4600\,X_{1i} + 0.3096\,X_{2i} + 1.6579X_{3i} + 0.1714X_{4i} + 0.9721X_{5i} - 0.6089X_{6i} - 0.4150X_{7i} - 0.0513X_{8i} - 0.1891X_{9i} + 0.9914X_{10i} + 1.0952X_{11i} + 1.27021X_{12i} \quad (5)$$

where $i = 1, \ldots, n$; $h = 1,2$; Y_{hi} = the h-th response variable in the i-th school unit; X_{hi} = explanatory variables in the h-th response and the i-th school unit; and e_{hi} = residual for the h-th response of the i-school unit.

Table 2. Multivariate Test Results (Wilks Lambda)

Wilk Lambda test	Wilks value	*p-value*	Wilk Lambda test	Wilks value	*p-value*
Simultaneous (12 explanatory variables)	0.4676	0.000*	Partial X7	0.9849	0.3132
Partial X1	0.9454	0.3001	Partial X8	0.9978	0.8462
Partial X2	0.9951	0.6873	Partial X9	0.9812	0.2353
Partial X3	0.9454	0.014*	Partial X10	0.9615	0.049*
Partial X4	0.3771	0.4826	Partial X11	0.9529	0.049*
Partial X5	0.9353	0.026*	Partial X12	0.9390	0.008*
Partial X6	0.9779	0.1815			

(*) Significant at 5% degree (α: 5%)

Equations (4) and (5) were tested simultaneously to test the significance of the explanatory variables on the quality of VHS. Simultaneous test results indicate that there is at least one explanatory variable that affects the quality of VHS at the 5% level of significance. The result of the partial test in Table 2 shows that the variables, which are the ratio of the number of students per group (X3), the percentage of certified teachers per number of teachers (X5), the ratio of the number of students per number of toilets (X10), the ratio of laboratory availability (X11), and the ratio of the availability of supporting rooms (X12), has a significant effect on the quality of VHS at a significant level of 5%. However, the results of the multivariate regression analysis were unable to identify the effect of the regency/city on the quality of VHS and the possible interactions between the school-level variables and the variables describing the regency/city level. Therefore, the multivariate analysis was only carried out at the school level [3]. The results of the test of the assumption of freedom between residuals obtained a *p*-value of 0.00*, which indicates that the residuals are not independent (correlated). This means that there is a deviation in the assumption of multivariate regression, namely, the residuals in the regression model are not independent (correlated), so the multivariate regression model is less accurate in this modelling [5], [21].

3.3 Multivariate Multilevel Modelling

3.3.1 Random Intercept Multivariate Multilevel Modelling

In this study, multivariate multilevel modelling was carried out using three model specifications, namely the intercept-only multilevel model (MVMM.0), the school-level randomized intercept multivariate multilevel model (MVMM1.1), and the multivariate multilevel combined model of school-level randomized intercept and regency-level randomized intercept (MVMM1.2). The authors in [3], [22] mentioned that the MVMM.0 model is used to determine an indication of the proportion of diversity at the regency/city level, or it can be interpreted as the expected correlation between two randomly selected schools within the same regency/city called the intraclass correlation coefficient (ICC). The result of the calculation of the ICC value is shown in Table 3 as follows.

Table 3. Variance Between Response Variables

Outline	*StdDev*	ICC
National Exam Score	0.0189	0.0345
Accreditation Score	2.4171	0.3076
Residual	5.4414	

Table 3 shows the estimated intraclass correlation coefficient (ICC) for each response variable: the national exam score of 0.0345 and the accreditation score of 0.3076. This means the expected correlation estimate between two randomly selected schools within the same regency/city of the accreditation score is 30.76%. Meanwhile, the expected correlation between two schools taken randomly in the same regency/city with the average NE score is 3.50%. The ICC value shows that the proportion of diversity at the regency/city level is quite large and indicates that the diversity of the regency/city affects the quality of VHS, making it possible to use multivariate multilevel analysis [23], [24].

The MVMM1.1 model was built to obtain explanatory variables at the school level that affect the quality of VHS. Table 5 shows that the MVMM1.1 model is a significant model at the 5% significance level, which means that the addition of explanatory variables at the school level has a significant effect on the quality of VHS. The significant variables in the MVMM1.1 model are as follows: percentage of graduates per number of students/pupils (X1), the ratio of the number of students/pupils per group (X3), the ratio of the number of students/pupils per number of teachers (X4), percentage of certified teachers per number of teachers (X5), the ratio of the number of students/pupils per number of toilets (X10), the ratio of laboratory availability (X11), and the ratio of the availability of supporting rooms (X12). The MVMM1.2 model is a combined model to determine the effect of adding explanatory variables at the regency/city level on the quality of VHS.

The explanatory variables at the regency/city level consist of four years (from 2018 to 2021). This is a multi-year data structure, so the dimensions of the variable data structure are complex. One method that handles high-dimensional data is PCA [25]. The results of the regency level variable PCA scree plot are shown in Figures (2) as follows.

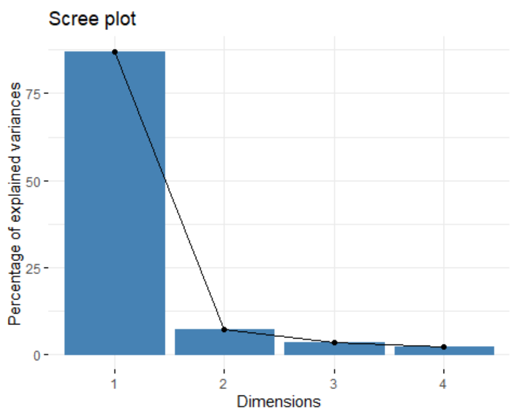

Figure 2. PCA scree plot of regency level variables

Figure (2) is a plot between *k* (sequential number of principal components) on the *x*-axis and the KU_k (*eigenvalue*$_k$) on the *y*-axis. The number of principal components (k) is taken based on the extreme position, where the curve line starts to slope or the point where there is no significant change in variance (steep) anymore. Based on the screen plot of Figure (2), the number of principal components k is 1 [26]. Thus, the characteristic root value (λ_i) is greater than one consisting of one main component with a cumulative proportion of 91.23%. This explains that 91.23% of information can be represented by explanatory variables at the regency/city level. The result of the deviance comparison test for the MVMM1.1 model is presented in Table 4 below.

Table 4. Comparison test of MVMM.0, MVMM1.1, and MVMM1.2 models

Model	*Df*	*Deviance*	LogLik	LRT	*p-value*
MVMM.0	8	2087.103	-1043.552		
MVMM1.1	22	1983.074	-991.5368	104.0297	0.001*
MVMM1.2	30	1967.387	-983.6934	15.6867	0.047*

*Significant at 5% degree (α: 5%)

Table 4 shows that the MVMM1.2 model is a significant model at the 5% level, meaning that the random intercept model involving explanatory variables at the school and regency/city levels is the best model at this stage. Thus, through the stepwise regression method, the fixed effect structure used at the school level is the percentage of graduate students/pupils (X1), the ratio of the number of students/pupils per group (X3), the ratio of the number of students/pupils per number of teachers (X4), the percentage of certified teachers per number of teachers (X5), the ratio of the number of students/pupils per number of toilets (X10), the ratio of laboratory availability (X11), and the ratio of the availability of supporting rooms (X12). Furthermore, the fixed effect structure used at the regency/city level is LFPR (Z1), poverty percentage (Z3), NPR (Z5), and GRDP (Z8).

3.3.2 Best Multivariate Multilevel Modelling

The MVMM1.2 model is the best based on the deviance value comparison test results. The estimation of fixed effects results uses the Maximum Likelihood approach as presented in Table 5.

Table 5. Best model estimation result (MVMM1.2)

	NE value		Accreditation			NE value		Accreditation	
Fixed Effect	Estimation	p-value	Estimation	p-value	Fixed Effect	Estimation	p-value	Estimation	p-value
Intercept	50.984	0.00*	86.157	0.000*	X11	-0.329	0.529	0.927	0.01*
X1	0.366	0.449	0.474	0.1917	X12	0.743	0.146	1.035	0.01*
X3	-0.088	0.898	1,370	0.008*	Z1	-0,125	0.660	0.099	0.643
X4	0.094	0.884	0.431	0.3750	Z3	0.023	0.938	0.645	0.00*
X5	-1.093	0.05*	1.699	0.001*	Z5	-0.164	0.485	-0.029	0.868
X10	-0.223	0.672	1.018	0.010*	Z8	0.225	0.422	0.712	0.00*

*Significant at 5% degree (α: 5%)

The estimation of the multivariate multilevel model that can be formed based on the estimated model parameters in Table 5 is as follows.

MVMM of School Level:

$$\hat{Y}_{1ij} = \beta_{10j} + 0{,}3658\,X_{1ij} - 0{,}0880X_{3ij} + 0{,}0940X_{4ij} - 0{,}9468^{*}X_{5ij} - 0{,}2234X_{10ij} - 0{,}3296X_{11ij} + 0{,}7435X_{12ij} \quad (6)$$

$$\hat{Y}_{2ij} = \beta_{20j} + 0{,}4742X_{1ij} + 1{,}3708X_{3ij} + 0{,}4317X_{4ij} + 0{,}5397^{*}X_{5ij} + 1{,}0183^{*}X_{10ij} + 0{,}9269^{*}X_{11ij} + 1{,}0350^{*}X_{12ij} \quad (7)$$

MVMM of Regency/City Level:

$$\hat{\beta}_{10j} = 50{,}9847 - 0{,}1251Z_{1j} + 0{,}0230Z_{3j} - 0{,}164Z_{5j} + 0{,}2253Z_{8j} + u_{10j} \quad (8)$$

$$\hat{\beta}_{20j} = 86{,}1557 + 0{,}0990Z_{1j} + 0{,}6448^{*}Z_{3j} - 0{,}0293Z_{5j} + 0{,}7122^{*}\,Z_{8j} + u_{20j} \qquad (9)$$

where, h =1,2; i=1, …n; j =1, …,p ; k=1,…,K; Y_{hij} = h-th response variables at the i-th school and j-th regency/city level; β_{p0j}= random intercept at the j-th unit at the regency/city level; X_{kij} = the k-th explanatory variable at the i-th school level and the j-th regency/city level; e_{pij} = residual for the i-th unit at the school level in the j-th unit at the regency/city level.

The estimation of the random effect that can be formed based on the multivariate multilevel model is shown in Table 6 below.

Table 6. The results of the best random effect estimation model

	NE		Accreditation	
Random Effect	**Estimator**	***p*-value**	**Estimator**	***p*-value**
Residual between regencies/cities				
$var(u_{h0j})$	0.01956	<.047	0.6676	<.047
$cov(u_{10j}, u_{20j})$	0.01095	<.047		
Residual within regencies/cities				
$var(e_{hij})$	0.4017	<.0001	1.0133	<.0001
$cov(e_{1ij}, e_{2ij})$	0.6342	<.0001		

(*)Significant at 5% degree (α: 5%)

3.3.3 Best Model Result Interpretation

This study conducted simultaneous and partial tests on the variables affecting VHS quality. Simultaneous testing of the MVMM1.2 model was carried out using the F-test. The F test in Table 5 reveals that at least one school level and regency/city level variables are significant at the 5% level. These variables have different effects on the quality variables of VHS. One of the factors is the characteristics of the national examination scores and VHS accreditation in South Sulawesi, which have a negative linear relation. The average NE score in the education unit does not reflect that the school's accreditation is better.

Estimation models (6), (7), (8), and (9) show that the variables at the school level that have a positive influence on accreditation are the ratio of the number of students/pupils per group (X3), the percentage of certified teachers (X5), the ratio of the number of students/pupils per number of toilets (X10), the ratio of laboratory availability (X11), and the ratio of the availability of supporting space (X12). This means that adding one unit of the variable can increase the quality of the school by the value of its coefficient. Meanwhile, the variable that has a negative effect on the quality of VHS is the percentage of certified teachers (X5). That means the addition of one unit of the variable will reduce the quality of the school by the value of its coefficient.

The ratio of the number of students/pupils per group increases (according to the standards of Permendikbud Number 34 of 2018 concerning Vocational Schools, which is in the range of 15-36 students/pupils per group) is supported by the percentage of certified teachers, and the availability of adequate facilities will improve on the school quality. The consequence of a large number of students/pupils is that schools must have qualified teachers and adequate facilities (including classroom conditions) so that the atmosphere of the learning process can run well, thus enabling the achievement of better quality as well [27]. Similarly, at the regency/city level, variables that have a positive value on school quality are obtained, namely the percentage of poverty (Z3) and GRDP (Z8). This means that the addition of one unit of the characteristic variable at the regency level will increase the quality of the school by the value of its coefficient. The expectation is that regencies/cities that have high GRDP values will have a positive influence on the quality of schools in their respective regencies/cities [28]. Meanwhile, it is not expected that the regional poverty level will affect the achievement of school quality. In addition, there are indications that students/pupils who have low economic ability prefer other education compared to VHS.

According to the estimation results of the random effect of the multivariate multilevel model in Table 6, it can be seen that there is at least one explanatory variable for the characteristics of the regency/city that has a significant impact on school quality. Characteristics of regency/city that significantly impact school quality according to the coefficient values in the two-level MVMM model in Table 6 are the percentage of poverty (Z3) and GRDP (Z8). Partially, these two variables have a positive effect on the accreditation score.

Table 6 shows that the variance that can be explained by these variables on the accreditation score is 66.67%. However, the contribution of the percentage of poverty (Z3) and GRDP (Z8) is low to the average diversity of NE scores between regencies/cities, which is 19.56%. Although the value of diversity shown is small, the average variance of NE scores in each regency is significant at the 5% level. Schools located in areas with higher GRDP tend to have better VHS quality. GRDP, in this context, reflects the regional economic structure and the level of economic growth in the regency/city [29]. Thus, it can be concluded that regencies/cities with high levels of economic growth, equitable community welfare, and the contribution of a large government role in education will improve school quality [30]. The achievement of a better quality of VHS in an area, indeed, is supported by the synergy of the role of the government, the role of schools, and the role of the community [31].

4 CONCLUSIONS

Multivariate multilevel modelling provides better modelling results than multivariate regression models. Multivariate multilevel modelling can show variations in the quality of VHS caused by the influence of the regency as the highest level. The regency-level multi-variable model with random intercept gives better modelling results than the school-level multi-variable model with the smallest deviance value. Simultaneous testing at the 5% significance level produces variables that contribute significantly to the quality of VHS in South Sulawesi Province. Partially, the percentage of certified teachers per number of teachers is one of the factors that significantly affects the NE score at the school level, while accreditation scores are the ratio of the number of students/pupils per group, the percentage of certified teachers per number of teachers, the ratio of the number of students/pupils per number of toilets, the ratio of laboratory availability, and the ratio of the availability of supporting space. Meanwhile, at the regency/city level, it was found that the percentage of poverty and GRDP had a significant effect on the accreditation score. Therefore, it can be interpreted that the progress of the education level of a region, besides being influenced by internal school factors, is also influenced by external school factors such as the community environment, regional economic structure, and the role of local government at the regency/city level.

AKNOWLEDGEMENT

The authors extend their gratitude to professor of statistics and science data, Mr Hari Wijayanto, and Mr Budi Susetyo for their guidance and direction.

REFERENCES

[1] Tom, A.B.S; Roel J.S, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling, 2nd Edition, London: SAGE Publications, 2012.

[2] J. Gill and J. A. Wommack , The Multilevel Model Framework, United Kingdom (UK): SAGE Publications Ltd, 2016.

[3] Hox J J; Moerbeek M; de Schoot R V, Multilevel Analysis Technique and Applications Third Edition, New York: Taylor & Francis, 2017.

[4] H. Aguinis and S. A. Culpepper, "An Expanded Decision Making Procedure for Examining Cross-Level Interaction Effects with Multilevel Modeling," *Organizational Research Methods*, vol. 18, no. 2, pp. 155-176, 19 January 2015.

[5] Snijders, TAB; Bosker, RJ, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling Second Edition, Los Angeles: SAGE, 2012.

[6] Tabachnick, B. G.; Fidell, L. S, Using Multivariate Statistics (6th ed.), Boston, MA: Pearson, 2013.

[7] Grilli L; Pennoni F; Rampichini C; Romeo I, "Exploiting TIMSS and PIRLS Combined Data: Multivariate Multilevel Modelling of Student Achievement.," *The Annals of Applied Statistics*, vol. Vol 10, no. No. 4, pp. 2405-2406, 2016.

[8] Rachmat H.,Susetyo B., Indahwati., "Applied Multivariate Multilevel Modelling for Student Cognitive Achievement Analysis in AKSI 2019 Survey," *IJSBAR*, vol. Vol. 55, no. No. 1, pp. ISSN2307-4531, 2021.

[9] Kementerian Pendidikan dan Kebudayaan, Pemenuhan Standar Nasional Pendidikan dan Mutu Satuan Pendidikan., Jakarta: Pusat Penelitian Kebijakan Pendidikan dan Kebudayaan, Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan, 2018.

[10] Goldstein H, Multilevel Statistical Model Internet Edition, London: Arnold Publishers, 1999.

[11] J. J. Hox and L. . M. Wijngaards-de, The multilevel regression model. In H. Best & C. Wolf (eds), TheSAGE handbook of regression analysis and causal inference., Thousand Oaks,CA: Sage, 2014.

[12] Khurniawan AW, Erda G, "Potret Tenaga Kerja Lulusan SMK pada Industri Manufaktur.," *Vocational Education Policy,* vol. 1, no. 8, 2019.

[13] Khurniawan, A. W. Sailah, I. Muliono, P. Maarif, M. S., Indriyanto, B., "Analysis of the Effect of School Governance and Total Quality Management on the Effectiveness of Vocational School-based Entreprise.," *International Journal of Management (IJM),* vol. Volume 11, no. Issue 9, pp. pp. 297-306, September 2020.

[14] Badan Pusat Statistik, Statistik Indonesia 2021, Jakarta: BPS, 2021.

[15] Badan Pusat Statistik Provinsi Sulawesi Selatan, Indikator Kesejahteraan Rakyat Sulawesi Selatan 2021, Makassar: BPS, 2021.

[16] Mukhlason A; Winanti T; Yundra E, "Analisa Indikator SMK Penyumbang Pengangguran di Provinsi Jawa Timur," vol. Vol 02, no. No 02, pp. p29-36, 2020..

[17] M. Turčičová, J. Mandel and K. Eben, "Multilevel maximum likelihood estimation with application to covariance matrices," *Communications in Statistics - Theory and Methods,* vol. 48, no. 4, pp. 909-925, 23 Jan 2018.

[18] Hox J. J. , Multilevel Analysis Techniques and Applications : Quantitative Methodology Series 2nd Edition, New York: Routledge, 2010.

[19] Badan Akreditasi Nasional Sekolah/Madrasah, Perangkat Akreditasi SMA/Sederajad, Jakarta (ID): Badan Akreditasi Nasional Sekolah/ Madrasah, 2021.

[20] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," *American Journal of Applied Mathematics and Statistics,* vol. 8, no. 2, pp. 39-42, 2020.

[21] C. Deeming and J. Kelvyn, "Social Policy and Economic Development in Europe: Investigating the Macro Determinants of Individual Health and Well-being in a Multilevel Multivariate Analysis of Thirty European Nations," *International Journal of Sociology,* vol. 45, no. 4, 2013.

[22] D. Liljequist, B. Elfving and K. S. Roaldsen, "Intraclass correlation – A discussion and demonstration of basic features," 22 July 2019.

[23] Sorra, J; Dyer, N, "Multilevel Psychometric Properties of the AHRQ Hospital Survey on Patient Safety Culture," *BMC Health Service Research,* Vols. Vol 10,199, pp. 3-13, 2010.

[24] T. K. Koo and Y. M. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *Journal of Chiropractic Medicine,* vol. 15, no. 2, pp. 155-163, 2016.

[25] Pramana S, Yuniarto B, Mariyah S, Santoso I, Nooraeni R, Data Mining dengan R: Konsep Serta Implementasi, Jakarta (ID): IN MEDIA-Anggota IKAPI No.250/JBA/2014, 2018.

[26] Johnson RA.,Wichern DW., Applied Multivariate Statistical Analysis, Sixth Edition, New Jersey:: Pearson Prentice Hall, 2007.

[27] S. N. Perdana, M. Handayani and J. Purnama, Analisis Hubungan Jumlah Rombongan Belajar dan Jumlah Peserta Didik per Rombongan Belajar Dengan Mutu Lulusan, Jakarta: Pusat Penelitian Kebijakan, Badan Penelitian dan Pengembangan dan, 2020.

[28] . P. Zulvia, A. Kurnia and M. A. Soleh , "Multilevel modelling and panel data analysis in educational research (Case study: National examination data senior high school in West Java)," in *Cite as: AIP Conference Proceedings*, 2017.

[29] Nurmayanti W.P, Notodiputro K.A, Indahwati, "Multilevel Modeling of Indonesian Student Achievement in Mathematics Based on TIMSS Data," *International Journal of Scientific & Engineering Research,* vol. 8, no. 2, February 2017.

[30] Sayuti M., Mujiarto, "Employability skills in vocational high school context: An analysis of the KTSP curriculum," *Journal of Vocational Education Studies,* vol. 1, no. 2, pp. 33-44. , 28 11 2018.

[31] Boer, L.F; Wijayanto, H; Indahwati, I, "Multilevel Modelling with Eigenvector Spatial Filtering and its Application to UN Score Data in Kendari.," *International Journal of Sciences: Basic and Applied Research (IJSBAR).,* vol. Vol 38, no. No.2, pp. 24-33, 03 May 2018..