JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 9 (ISSUE 1) JUNE 2025: 22 – 36
https://doi.org/10.21009/JSA.09103

IDENTIFICATION OF PRIMARY SCHOOL LITERACY ACHIEVEMENT FACTORS IN PROVINCE X USING ORDINAL STEPWISE LOGISTIC

Siti Nur Azizah[1*], Dela Gustiara[2], Anwar Fitrianto[3], Erfiani[4], Pika Silvianti[5]

[1,2,3,4,5] Department of Statistics, IPB University, Kampus IPB, Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, 16680, Indonesia

Corresponding author's e-mail: * ***24delagustiara@apps.ipb.ac.id***

ABSTRACT

Literacy is a foundational skill that underpins students' academic success and lifelong opportunities. Low literacy skills can result in long-term disadvantages such as limited access to higher education, low productivity, and social inequality. Indonesia continues to face challenges in improving students' literacy achievement, particularly at the primary school level. According to the PISA 2022 results, Indonesia ranked 69th out of 81 countries, indicating that students' literacy levels remain relatively low. This study aims to identify the factors that influence the literacy achievement of primary school students in Province X. The analytical method employed is ordinal logistic regression with a backward stepwise approach. The dependent variable is the level of literacy achievement (categorized as low, moderate, and good), while the independent variables include learning quality, teacher reflection and improvement, instructional leadership, school climate (including safety, diversity, and inclusiveness), and curriculum type. The results show that the final selected model follows the partial proportional odds assumption and includes only the significant predictors identified through backward stepwise elimination. Variables that positively influence literacy achievement include safety climate, diversity, inclusiveness, curriculum type, and teachers' reflection and improvement of learning. Model evaluation using AIC, BIC, and accuracy measures indicates good predictive performance. These findings offer valuable insights for policymakers in designing strategies to enhance literacy through strengthening school climate and improving the quality of teaching and learning.

Article History:

Received: May 25, 2025
Revised: June 28, 2025
Accepted: June 29, 2025
Published: June 20, 2025
Available online.

Keywords:

Ordinal Logistic Regression, Backward Stepwise, National Assessment, Literacy, School Climate

How to cite this article:
S. N. Azizah, D. Gustiara, A. Fitrianto, Erfiani, P. Silvianti, "IDENTIFICATION OF PRIMARY SCHOOL LITERACY ACHIEVEMENT FACTORS IN PROVINCE X USING ORDINAL STEPWISE LOGISTIC", Jurnal Statistika dan Aplikasinya, vol. 9, iss. 1, pp. 22 – 36, June 2025

Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Literacy is a fundamental pillar of lifelong learning and a core competency that must be developed from an early age [1]. In the context of primary education, literacy plays a crucial role as it forms the foundation for learning across all subjects. However, Indonesia continues to face significant challenges in improving literacy outcomes, particularly at the primary school level. Hafizah and Rakhmania's research [2] emphasizes that this issue requires urgent attention, as the 2022 PISA results placed Indonesia 69th out of 81 participating countries, with an average reading literacy score of 359, far below Singapore, which ranked first with a score of 543. This stark gap reflects the persistently low quality of students' literacy skills and highlights the need for systematic efforts to improve them. Low literacy is associated with long-term disadvantages, including limited access to higher education, reduced productivity, and increased social inequality. As literacy is not developed instantly but rather through a long-term process, early interventions in primary education are essential. Research by the OECD [3] shows that school climate, including discipline, safety, and teacher support, has a strong correlation with students' academic achievement. Therefore, understanding the key school-level factors associated with literacy outcomes is essential for designing targeted interventions.

To support these efforts, appropriate analytical methods are required to examine the relationship between literacy achievement and contextual predictors. Given that literacy in the National Assessment is categorized into three levels (low, moderate, and good), ordinal logistic regression is considered an appropriate method for analysis. Ordinal logistic regression is a statistical method used to analyze the relationship between one response variable that has an order (ordinal) with one or more predictor variables [4]. Among the available models, the cumulative logit model is the most widely used form of ordinal logistic regression, as it estimates the cumulative probability of a response falling within or below a given category. To ensure the model is both statistically efficient and interpretable, a variable selection procedure is necessary. One widely used approach is backward stepwise selection, which begins with a full model that includes all candidate predictors and sequentially removes the least significant variables based on p-values. This process continues until only statistically significant predictors remain [5]. Compared to forward selection, backward stepwise selection is considered more stable as it reduces the risk of prematurely excluding important variables [5]. In addition to statistical significance, model selection should also consider a balance between goodness-of-fit and model simplicity. Evaluation criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are commonly used to assess model performance. However, simpler models tend to be easier to interpret and have less risk of overfitting, as long as it does not significantly compromise prediction accuracy [5], [6], [7]. This is in line with the findings of Khikmah et al. [8], who used the backward stepwise method to obtain the best model for identifying factors affecting the population rate in Java. The analysis method is carried out in stages by eliminating insignificant variables until the best model with the smallest AIC is obtained. Thus, the significant variables based on the smallest AIC are consumption level, gender empowerment index, and area.

Ordinal logistic regression has been widely applied across various domains, including health and education. For instance, Mayawi et al. [9] showed that ordinal logistic regression can identify significant factors that affect blood sugar levels in diabetic patients, such as age, BMI, cholesterol, LDL, HDL, and TCH hormone. In the field of education, Wang [10] used an ordinal logistic regression model to analyze the effectiveness of ACUE teaching practices and found that students who took classes with the ACUE approach had a 1.92 times greater chance of graduating as well as significantly participating actively in class. In the Indonesian context, the National Assessment was introduced in 2021 as a more comprehensive evaluation of the education system, replacing the national exam. One of its core components is the Minimum Competency Assessment (Asesmen Kompetensi Minimum or AKM), which includes literacy as a key domain. Literacy in AKM has a broader scope, which is defined as the ability to understand, evaluate, and reflect on written information in various contexts, not just the ability to read [11]. Based on this background, the present study aims to identify the factors that influence literacy achievement among primary school students in Province X. The analysis is based on secondary data from the Ministry of Primary and Secondary Education and employs ordinal logistic regression with a backward stepwise selection approach. The resulting model is expected to be statistically robust, practically interpretable, and useful for informing evidence-based policies to improve literacy outcomes in primary education.

2. METHODS

Material and Data

This research uses secondary data obtained from the Ministry of Education, Culture, Research, and Technology (currently the Ministry of Primary and Secondary Education). The data used is the result of the 2023 National Assessment for Elementary School Level in one of the provinces in Indonesia. The amount of data in Province X is 9,433 schools, with the variables used presented in Table 1.

Table 1. List of variables used

Variable	Description	Data Type	Data Scale
Y	Literacy	Ordinal	1: Low 2: Moderate 3: Good
X_1	Quality Learning	Numeric	Score 0-100
X_2	Teacher Reflection and Learning Improvement	Numeric	Score 0-100
X_3	Instructional Leadership	Numeric	Score 0-100
X_4	Safety Climate	Numeric	Score 0-100
X_5	Diversity Climate	Numeric	Score 0-100
X_6	Inclusiveness Climate	Numeric	Score 0-100
X_7	Curriculum	Nominal	Curriculum 2013 1-Year *KM* 2-Year *KM* 3-Year *KM*

Note: KM = Kurikulum Merdeka

To maintain confidentiality and protect regional identity, the names of the regions analyzed are not included in this publication. The data used in this analysis includes elementary school-level data, which is under the auspices of the Ministry of Primary and Secondary Education (Kemendikdasmen). Religious or equality-based schools are not included in the scope of the analysis, because the research focus is directed at formal education units that are directly under the authority of the Ministry of Basic Education.

Research Method

1. Data preparation/preprocessing data

Data preparation is the initial process before carrying out analysis. In this stage, a data cleaning process is carried out, namely, removing variables that are not needed in the analysis, for example, removing educational units other than elementary schools, and only taking adequate and representative data, handling missing data (missing values), for example, by deleting, replacing, or inputting values.

2. Data Exploration

Data exploration was carried out by analyzing the distribution of response and predictor variables, identifying outliers, and basic descriptive statistics.

3. Splitting/ data splitting

In evaluating model performance objectively and avoiding overfitting problems, namely the condition when the model is too suitable for the training data and fails to make good predictions on new data. The data needs to be divided into two subsets, namely training data and test data. This division is an approach commonly used in predictive analysis because it allows the model to be evaluated against data that is not used in the training process, thus providing an idea of the model's generalization capabilities [12]. The proportion of data sharing of 80% for training (training data) and 20% for testing (testing data) is a common practice in predictive analysis because it is considered to provide a good balance between sufficient data to build a model and representative data for evaluating model performance [13], [14].

Before separating the data, it is important to first check the distribution of the response variable on a categorical scale. This examination aims to determine the proportion of each category in the data. Balancing proportions between categories is a crucial aspect in the modeling process. If the proportion of data is not balanced, it can cause problems such as the model being biased in prediction results [15]. The bias in question is that when one category dominates, the model tends to be more accurate in predicting the majority category but is less able to recognize the minority category correctly.

4. Check assumptions

The assumptions that usually need to be taken into account in ordinal logistic regression are that there is no multicollinearity, no outliers, and a sufficient sample size. According to Gujarati and Porter [16], multicollinearity can occur if there is a linear relationship between independent variables. Multicollinearity can be tested by calculating the value of VIF. If the value is more than 10, the data indicates multicollinearity, which will make the regression coefficient difficult to interpret accurately. For this reason, if the model being built contains multicollinearity, it needs to be handled first, for example, using PCA, ridge regression, or other relevant statistical techniques.

The second assumption is the absence of outliers. This outlier detection is important because it can affect the model results. Fitrianto and Xin's research [17] shows that the presence of outliers can cause model results to not represent the data as a whole. So, outlier detection and handling need to be done to produce a model that is accurate and meets the assumption requirements.

Next, the third assumption relates to sample size. In ordinal logistic regression, inadequate sample size can affect the accuracy of parameter estimates and the validity of statistical tests. Research by Bujang et al. [18] shows that a logistic regression model built with a sample size of 500 produces more stable and accurate parameter estimates compared to a small sample size, such as 100.

Harrell emphasizes that in logistic regression, especially when the response variable is on an ordinal scale, the amount of statistical information obtained from the data depends not only on the number of observations but also on the distribution of categories and the complexity of the model [19]. Therefore, it is recommended that the number of parameters that are free in the model (including non-linear interactions and transformations) should be smaller than $m/10$ or $m/20$, where m is the number of events in the minor category. Thus, there is a balance between the complexity of the model and the amount of information available. Because even though a large sample is available, estimates can still be unstable or biased if the model contains too many parameters.

In addition, for ordinal variables, the effective amount of information can also be calculated using the following formula:

$$n_{effective} = n - \frac{1}{n^2}\sum_{i=1}^{k}({n_i}^3) \quad (1)$$

where n_i is the number of observations in the i-category and k is the number of categories in the response variable. This formula is used to assess the efficiency of the ordinal category distribution of the total available information. Thus, a balance between sample size, category distribution, and number of parameters is important so that the model does not suffer from overfitting or estimation bias.

5. Modelling

Ordinal logistic regression used in modeling is a statistical method used to analyze the relationship between a response variable that has an order (ordinal) and one or more predictor variables [4]. The general model that is most widely used in ordinal logistic regression is the model of *cumulative logit*, because it can model or present cumulative probability in a particular category.

The following is the equation of the logit model for ordinal logistic regression:

$$logit\ P(x_i) = g_k(x_i) = ln\left[\frac{P(Y\leq k\,|\,x_i)}{P(Y>k\,|\,x_i)}\right] = \theta_k + \sum_{i=1}^{p}\beta_j x_{ij} \tag{2}$$

Where $k = 1,2,..,q - 1$ and q are the number of response variables while $j = 1,2,\dots,p$ and p are the number of predictor variables. The cumulative probability of the k response variable is obtained using the following equation:

$$P(x_i) = \pi_k(x_i) = \frac{exp\ (\theta_k+\sum_{j=1}^{p}\beta_j x_{ij})}{1+exp\ (\theta_k+\sum_{j=1}^{p}\beta_j x_{ij})} \tag{3}$$

The Maximum Likelihood Estimation (MLE) method is used to estimate parameters θ and β, where is likelihood function obtained based on the product of the joint probability function of $Y_1, Y_2, \dots, Y_n$. Here is the functional equation of likelihood:

$$L(\theta,\beta) = \prod_{i=1}^{n}\left[\pi_1(x_{ip})^{yi1}\pi_2(x_{ip})^{yi2}\dots\pi_q(x_{ip})^{yiq}\right] \tag{4}$$

to obtain estimated values θ and β can optimally use the function likelihood, which is maximized in the form of a natural logarithm known as a function ln-likelihood. Here is the functional equation ln-likelihood:

$$L(\theta,\beta) = \sum_{i=1}^{n} y_{i1}ln\left[\pi_1(x_{ip})\right] + y_{i2}ln\left[\pi_2(x_{ip})\right] + \cdots + y_{iq}ln\left[\pi_q(x_{ip})\right] \tag{5}$$

Ordinal logistic regression modeling was carried out in several stages. First, the initial model (full model) was compiled by entering all predictor variables into an ordinal logistic regression model. In ordinal logistic regression with cumulative logit, there are important assumptions to fulfill, namely assumption of proportional odds. According to Liu, test assumptions proportional odds (PO) is used to ensure that the influence of predictor variables is consistent (proportional) across all response variable categories [20]. Several methods that can be used in this test are score test, forest test, and likelihood ratio test. Second, according to Peterson and Harrell [21], if there are predictors that do not meet the PO assumption, then it is necessary to use other alternative models such as partial proportional odds (PPO) or a multinomial model. The PPO model is used when PO assumptions are not met. This assumption allows some predictors in the model do not meet the PO assumption. Third, carry out variable selection using the backward stepwise method. Backward stepwise is one of the variable selection methods in regression modeling. This procedure starts from a complex model (full model) that includes all predictor variables, then gradually eliminates variables that are not significant in stages based on the statistical significance value or Akaike Information Criterion (AIC). Deletion of variables is carried out one by one until only the variables remain that make a significant contribution to the response variable [6].

The significance value can be seen from the simultaneous test or the partial test. The simultaneous test is a test used to find out whether the predictor variables together have a significant effect on the response variable. If the p-value of the simultaneous test is less than 0.05, it can be concluded that there is at least one predictor that influences the response variable statistically [9]. The partial test is a test used to determine the effect of each predictor variable on the response variable. If the p-value of the partial test is less than 0.05, it can be concluded that the predictor variable has a statistical effect on the response variable [22].

6. Choose the best model

According to Agresti, selecting the best model can be done by comparing the selected models based on their deviance values, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) by considering the balance between model suitability and complexity [6]. Mark deviance is a measure of the model's suitability to the data. The smaller the deviance value, the better the model's suitability to the data. The following is the equation for deviance:

$$Deviance = -2(log - likelihood\ model) \tag{6}$$

AIC is used to compare the quality of several ordinal logistic regression models. AIC assesses how well the model explains the data without being too complex (overfitting). The smaller the AIC value, the better the model is at explaining the data. The following is the AIC equation:

$$AIC = 2k - 2ln(L) \quad (7)$$

with k is the number of categories in the response variable, and L is the value of likelihood.

Meanwhile, BIC is a statistical model selection method with a Bayesian approach whose aim is to find the model that is most likely to represent the actual data. BIC in selecting the best model tends to choose simpler models because it imposes a greater penalty on model complexity, especially when the sample size is large. So, the smaller the BIC value, the better the model will be in explaining the data. The following is the BIC equation:

$$BIC = -2ln(L) + k\ ln(n) \quad (8)$$

with k is the number of categories in the response variable, and n is the number of observations.

Besides deviance, AIC, and BIC, the supporting criteria that can be used in selecting the best model are the pseudo R-square values. A higher pseudo R-squared value indicates the model better explains the data. Pseudo R-square is a statistical measure used to measure how well an independent variable can explain variations in the dependent variable in a regression model. Pseudo R-square ranges from 0 to 1, where if the value is higher or closer to 1, it indicates the model's ability to make predictions is getting better [23].

7. Evaluate model accuracy

After obtaining the best model based on information criteria (AIC, BIC, and deviance) on the training data, the performance of the model is evaluated on the test data to assess its generalization ability. This evaluation is important because high accuracy on training data does not guarantee good performance on test data. Performance evaluation on test data is a key indicator in assessing a model's ability to predict data from a wider population. One method commonly used to assess the performance of classification models is the analysis of the confusion matrix, which describes the level of accuracy and patterns of classification errors [24]. The evaluation metrics used in assessing the performance of the classification model can be seen in Table 2.

Table 2. Confusion Matrix

	Positive	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

In evaluating the model, there are several measures that can be used, namely: Sensitivity/recall, Specificity, and Accuracy. Sensitivity/recall is a measure of how precise the model is in predicting all true positive classes, with the formula:

$$Sensitivity = \frac{TP}{TP+FN} \quad (9)$$

Specificity is a measure of how precise the model is in predicting all true negative classes, with the formula:

$$Specificity = \frac{TN}{TN+FP} \quad (10)$$

Accuracy is a measure of how precisely the model is in correctly predicting the entire dataset, with the formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

For a fair evaluation of classification models, especially if the data is imbalanced, it is necessary to consider the values of all three, sensitivity, specificity, and accuracy. This use is carried out to provide a comprehensive overview and results of the model performance in each class [25]. Thus, all model evaluations can be used to recognize and predict minority and majority classes fairly.

8. Perform parameter estimates and interpret the final model

Parameter estimation is done by calculating the odds ratio value obtained from the exponent of each predictor coefficient. Then the interpretation of each predictor is done by interpreting the odds ratio value.

3. RESULTS

Descriptive Analysis

Before conducting further analysis, a descriptive analysis was first performed to obtain a general overview of the 2023 National Assessment data at the elementary school level in one of the provinces in Indonesia. The following is a descriptive analysis of each numerical predictor variable using a boxplot.

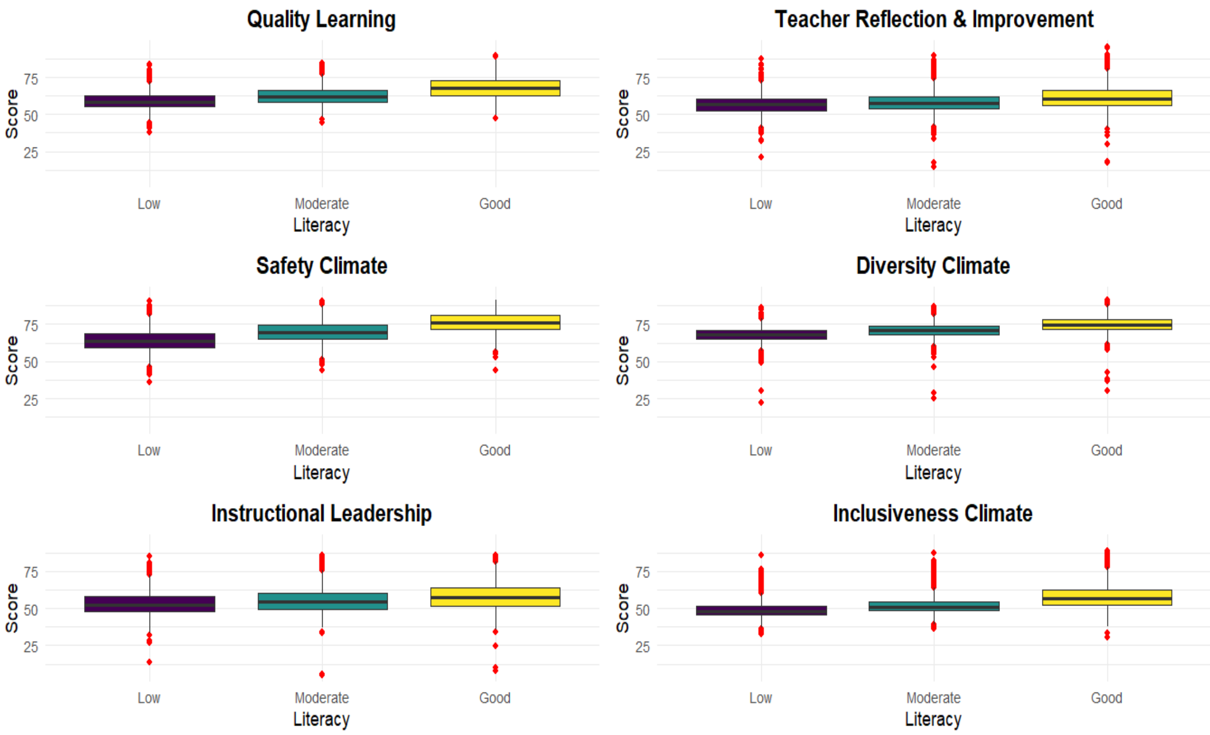

Figure 1. Boxplot of predictor with outlier identification

Figure 1 shows a boxplot that describes the relationship between literacy achievement and six predictor variables, namely learning quality, teacher reflection and learning improvement, safety climate, diversity climate, and inclusiveness climate. From this visualization, it appears that the highest median value for all predictor variables is in the literacy achievement category “Good.” This pattern shows a tendency that the higher the score on the predictor variables, the greater the likelihood that the school is in a higher literacy achievement category. Thus, these findings indicate a positive association between the literacy achievement category and each of these predictor variables.

Table 3 presents the results of Spearman's correlation between predictor variables and elementary school literacy achievement. The results of the analysis show that all predictor variables have a positive and significant relationship with literacy achievement. A positive correlation coefficient indicates that schools with higher scores on variables such as learning quality, teacher reflection and learning improvement, instructional leadership, safety climate, diversity climate, and inclusiveness climate tend to have better literacy outcomes. These findings indicate a positive association between these characteristics and literacy achievement.

Table 3. Correlation between predictor variables and literacy outcomes

Variable	Correlation
Safety Climate	0.5704958*
Inclusiveness Climate	0.5346132*
Learning Quality	0.5066289*
Teacher Reflection and Learning Improvement	0.2468498*
Instructional Leadership	0.2342630*
Diversity Climate	0.5073118*

*Signif.: * (0.05)*

Previously, the visualization boxplot in Figure 1 shows a trend of increasing predictor variable values as the literacy achievement category increases. This pattern is in line with the results of Spearman's correlation analysis, which indicates a positive and significant relationship between each predictor variable and literacy achievement. Among the variables analyzed, safety climate and inclusiveness climate showed relatively stronger associations, thus potentially becoming important predictors in the subsequent modeling process. Apart from the relationship between variables, the visualization also shows the existence of outlier values for all predictor variables. The existence of these outliers indicates a potential violation of assumptions that needs to be addressed first before proceeding to the modeling stage.

Apart from carrying out descriptive analysis of the numerical scale predictor variables, it is also important to review the distribution of categories in the response variable. In this case, the distribution of literacy achievement is as follows: 23.30% in the "low" category, 34.08% in the "moderate" category, and 42.62% in the "good" category. This proportion indicates a relatively balanced distribution across categories, with no significant indication of class imbalance. As a result, potential bias in the modeling process due to unequal category distribution can be minimized.

Assumptions Testing

Before testing assumptions, the data is first divided into training data and testing data. This separation is important to avoid bias in model evaluation. Several studies suggest that the entire modeling process, including transformation, variable selection, and assumption checking, be carried out on training data so that the estimated model performance on testing data remains valid and is not influenced by information from the testing data (data leakage) [12], [19], [26]. This data separation ensures that the model is tested on data that it has never actually seen, so that the evaluation results more closely reflect the actual performance of the model. The proportion used in data separation is 80% for testing data and 20% for training data. Thus, the amount of data in the training data is 7,547, and the testing data is 1,886.

In ordinal logistic regression, some of the main assumptions that must be met are the absence of outliers, absence of multicollinearity, and adequate sample size. First, the assumption regarding the absence of outliers in the predictor variables needs to be considered. Based on the visualization boxplot, outliers are identified in all predictor variables (see Figure 2), which could potentially influence modeling results. According to Ghozali [23], deleting outlier data is permitted because it is one way to overcome symptoms of normality in the data. Jim [27] also mentioned in his article that outlier removal is acceptable if it is done systematically and does not throw away important information. In this research, outliers are overcome by removing them because the model estimation results become less reasonable when outliers are not removed. The method used to handle outliers is the Interquartile Range (IQR), where data that falls outside the lower and upper limits of the IQR is deleted. Handling outliers using IQR removes 472 schools from a total of 7,547 schools in the training data, or around 6.25% of the data. Removing outlier data to produce more accurate estimates was also carried out by Asriyana et al. [28]. In his research, deletion was carried out on 4 identified outlier data points, where the outliers had caused symptoms of heteroscedasticity and disrupted the validity of the model estimates. Data visualization before and after handling outliers is presented in Figure 2.

The visualization in Figure 2 shows the distribution of data after handling outliers using the IQR method. Identified outliers have been removed, resulting in cleaner data distribution without extreme values that could potentially influence modeling results. Even though outliers are removed, the main distribution of the data does not change substantially. The pattern of relationship between literacy level and predictor variables remains consistent, with the highest median value still found in the literacy category “Good.” This indicates that even though outliers have been removed, the positive association between literacy achievement and the predictor variables is maintained.

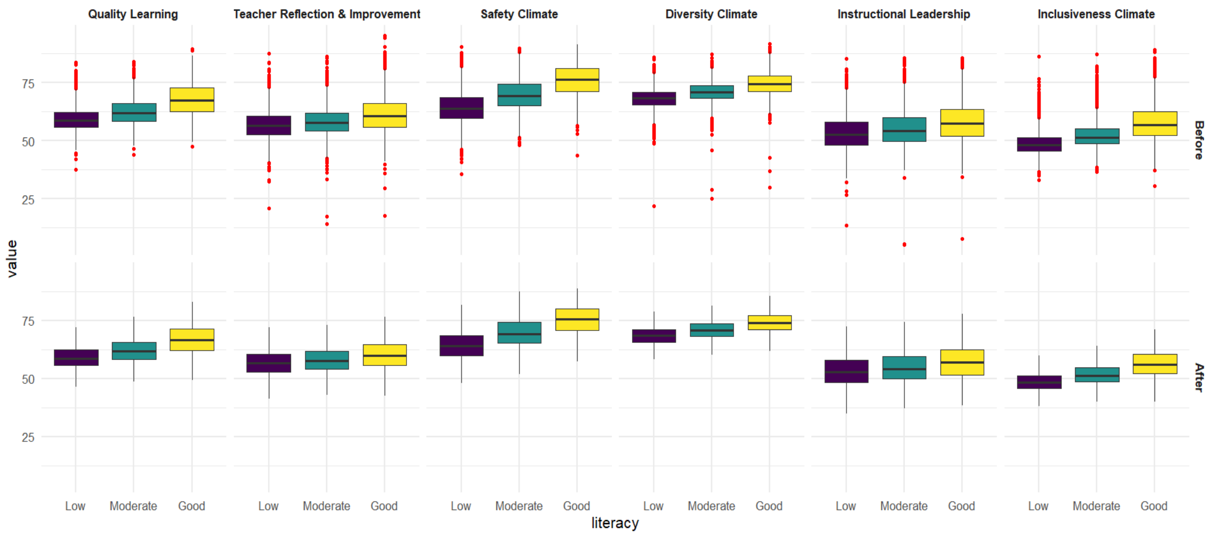

Figure 2. Boxplot comparison before and after outlier treatment for each predictor

Second, to verify that there is no multicollinearity among the predictor variables. The multicollinearity test is carried out by looking at the Variance Inflation Factor (VIF) values. Based on Table 4, all predictors have VIF values below 5, which indicates there is no multicollinearity problem in the data.

Table 4. VIF values for each numerical predictor

Predictor Variable	VIF
Safety Climate	1.563850
Inclusiveness Climate	1.380994
Quality Learning	1.643995
Teacher Reflection and Learning Improvement	1.800875
Instructional Leadership	1.771971
Diversity Climate	1.461699
Curriculum	1.003668

Third, a sufficient sample size. After handling outliers, 7,075 valid data points were obtained with a literacy achievement distribution of 23.31% for the “low” category, 34.08% for the “moderate” category, and 42.61% for the “good” category. Referring to Long [29] and Bujang et al. [18], the sample size used in this study exceeds the minimum recommended threshold of 500, which is considered sufficient to produce stable and significant parameter estimates in logistic regression. In addition, it is important to ensure that the number of parameters used remains proportional to the number of events in the minor category to prevent overfitting. Based on the formula proposed by Harrell [29], namely $n - (1/n^2) \times \sum(n_i^3)$, where n_i is the number of observations in each category, an estimated minimum sample size of 6,157 is obtained. Since the amount of data available (7,075) exceeds this value, the ordinal logistic regression modeling in this study can be said to be strong enough to produce valid and reliable parameter estimates.

After testing and handling assumptions, the next step is to select the best model using the backward stepwise method. This method is used to identify the most significant variables in explaining the relationship between variables, as well as to reduce model complexity by eliminating variables that are not significant.

Selection of the Best Model

In this section, several models will be discussed, namely the full ordinal logistic regression model and the backward stepwise logistic regression model. The full ordinal logistic regression model is modeled by including all predictor variables in the model. Ordinal logistic regression uses logit cumulative odds, with the important assumption of proportional odds (PO). After modeling with the full model, PO assumptions are checked using the Brant test (one of the score test methods), and the results show that there is one variable that does not meet the PO assumptions, namely the variable reflection and improvement of learning by the teacher, because of grades p-value < 0.05. Therefore, we cannot assume that all predictors have the same probability for each category of the response variable. An alternative to handle this is to use partial proportional odds (PPO). The PPO assumption allows some predictors not to meet the PO assumption, while predictors that meet the PO assumption are retained. The implication is that predictor variables that do not meet the PO assumption do not have the same probability for each category of literacy achievement.

Table 5. Results of backward stepwise PPO model analysis

	AIC	LRT	P-value
Backward stepwise PPO model - Step 1			
Intercept	12.222		
Quality Learning	12.221	0.002	0.9658
Safety Climate	12.503	282.223	$< 2.2 \times 10^{-16***}$
Diversity Climate	12.413	192.197	$< 2.2 \times 10^{-16***}$
Instructional Leadership	12.223	2.158	0.1418
Inclusiveness Climate	12.330	109.202	$< 2.2 \times 10^{-16***}$
Curriculum	12.281	64.294	$7.1 \times 10^{-14***}$
Backward stepwise PPO model - Step 2			
Intercept	12.221		
Instructional Leadership	12.221	2.18	0.1398
Safety Climate	12.578	359.94	$< 2.2 \times 10^{-16***}$
Diversity Climate	12.416	197.74	$< 2.2 \times 10^{-16***}$
Inclusiveness Climate	12.339	120.87	$< 2.2 \times 10^{-16***}$
Curriculum	12.279	64.58	$6.2 \times 10^{-14***}$
Backward stepwise PPO model - Step 3			
Intercept	12.221		
Safety Climate	12.585	366.51	$< 2.2 \times 10^{-16***}$
Diversity Climate	12.418	199.71	$< 2.2 \times 10^{-16***}$
Inclusiveness Climate	12.340	121.79	$< 2.2 \times 10^{-16***}$
Curriculum	12.279	64.21	$7.4 \times 10^{-14***}$
Backward stepwise PPO model - Step 4			
model_null → Literacy~Safety Climate+Diversity Climate+Inclusiveness Climate + Curriculum			
model_final → Literacy ~ Teacher Reflection and Learning Improvement+Safety Climate + Diversity Climate+Inclusiveness Climate+Curriculum			
model_null	12.305		
model_final	12.221	88.756	$< 2.2 \times 10^{-16***}$

*Signif.: 0 '***' , 0.001 '**' , 0.01 '*' , 0.05 '.' , 0.1 ' ' , 1*

In the first step, the learning quality variable shows the highest p-value (> 0.05), and its presence in the model does not provide a significant decrease in the AIC value. Therefore, this variable was eliminated in the second step. Next, in the second step, the instructional leadership variable is also not significant and does not contribute to the decrease in AIC, so it is excluded in the third step. After deletion of these two variables, all remaining variables in the model showed statistical significance, signaling the end of the selection process backward.

Evaluation of variables that do not meet PO assumptions, namely teacher reflection and learning improvement, is carried out through tests, the Likelihood Ratio Test (LRT), by comparing models with and without non-parallel variables of these variables. The test results shown in Table 5, step four, show that the model with the variable reflection and improvement of learning by teachers has a p-value < 0.05, which means that the non-parallel effects of these variables significantly improve the fit of the model. Thus, this variable is worthy of being retained in the model. In addition, the AIC value of the model with variables for teacher reflection and learning improvement is lower than the model without these variables, thus supporting its existence in the PPO model. Thus, the best PPO model is procedure-based, backward stepwise, and teacher reflection and learning improvement, security climate, diversity climate, inclusiveness climate, and type of school curriculum.

Before choosing the best model based on the criteria mentioned previously, it is important to compare the model's accuracy level on training data and testing data. This comparison is carried out to evaluate how well the built model can generalize to new data. The results comparing the accuracy of training data and testing data are presented in Table 6. Based on Table 6, the level of accuracy between training data and testing data is almost the same, which shows that the model built does not experience underfitting or overfitting. In other words, this model can be used well not only on training data but also on test data.

Table 6. Accuracy comparison between training and testing data

Logistic Regression Model	Training Data	Testing Data
Full PO Model	0.5937	0.5938
Full PPO Model	0.5934	0.5941
Backward Stepwise PPO Model	0.5925	0.5924

The results of the comparison of three models, namely the full PO model, the full PPO model, and backward stepwise PPO, can be seen in Table 7. Based on the AIC and BIC values, the PPO model shows the best performance compared to the other two models. However, if you look at the standard deviation, pseudo R^2, and accuracy on test data, the results are inconsistent. Specifically, standard deviation and the smallest and largest pseudo R^2 were found in the full PPO model, while the highest accuracy was found in the full PO model. However, these differences are relatively small. For deviance, the difference between PPO and full models backward stepwise PPO is less than 2 points, while for pseudo R^2 it is almost identical; the value for the full PPO model is 0.3579, and for backward stepwise PPO it is 0.3577. Likewise, for accuracy on testing data, the difference between the two models is only about 0.42 points.

Table 7. Comparison of regression models

Criteria	Full PO	Full PPO	Backward Stepwise PPO
Deviance	9759.567	9757.464	9759.466
AIC	9781.567	9781.464	9779.466
BIC	9854.620	9861.158	9845.878
Accuracy	59.378	59.406	59.237
Sensitivity	57.517	57.602	57.409
Specificity	78.929	78.944	78.865
Pseudo R^2	0.197	0.197	0.197

Apart from the dimensions already mentioned, selecting the best model also considers aspects of model complexity and simplicity. Therefore, the backward stepwise PPO model is selected as the best model, as it focuses only on variables that significantly affect literacy achievement. This model not only accommodates violations of the basic assumptions of ordinal regression but also presents a simpler structure with predictive performance that remains competitive compared to the full PPO model.

Best Model Estimation and Interpretation

Results of the analysis of the ordinal logistic regression model with backward stepwise PPO are presented in Table 8. All predictor variables retained in the model show a positive relationship with literacy achievement. Safety climate is defined as the school community's perception of feeling safe in the school environment from physical and psychological violence. The odds ratio of 1.10 indicates that each one-point increase in the safety climate score is associated with a 10% increase in the odds for a school to be in a higher literacy category, compared to the same or lower category, after controlling for other variables in the model. In accordance with research by Kustyuruba et al. [30], who emphasized that a safe school environment allows students to feel comfortable and focused on learning. Learning conditions that are free from disturbances, such as violence or intimidation, create an optimal environment, which will have a positive impact on increasing students' literacy achievements.

Likewise, the diversity climate and the inclusiveness climate have odds ratios of 1.12 and 1.07, respectively, indicating that each one-point increase in each of these variables is associated with a 12% (diversity climate) and 7% (inclusiveness climate) increase in the odds for a school to be in a higher literacy category, compared to the same or lower category, after controlling for other variables in the model. Diversity climate refers to the extent to which a school facilitates interaction between students from diverse backgrounds and acceptance of diversity. Research shows that education by implementing diversity has been proven to increase tolerant behavior among students. Through a learning environment that prioritizes tolerant behavior, students not only learn to live side by side in harmony but also develop critical thinking towards human and national values [31]. A climate of inclusivity measures the extent to which a school creates an environment that supports the involvement of all students, including those who have special needs or are from minority groups. Research shows that effective inclusive education can improve the adjustment abilities of students with special needs, thereby helping them socialize better in diverse school environments. Good adaptability has an important role in supporting the learning process, which in the end can optimally increase achievement in both academic and non-academic fields for students with special needs [32].

Table 8. Parameter estimation of predictor variables of backward stepwise PPO model

Predictor Variable	Coefficient	Standard Error	Odds ratio
Intercept (Low|Moderate)	15.259146	0.477710	
Intercept (Moderate|Good)	16.834653	0.452550	
Safety Climate	0.100082	0.005326	1.10
Diversity Climate	0.116482	0.008347	1.12
Inclusiveness Climate	0.068608	0.006288	1.07
1-Year *KM*	0.186811	0.053799	1.20
2-Year *KM*	0.536013	0.070707	1.70
3-Year *KM*	0.932168	0.308849	2.54
Teacher Reflection(Low|Moderate)	0.040049	0.006077	1.04
Teacher Reflection(Moderate|Good)	0.049823	0.005560	1.05

Curriculum Type indicates the implementation status of the Independent Curriculum (KM) in school. This curriculum is categorized into: (i) non-*KM* (still using the 2013 curriculum), (ii) 1-year KM, (iii) 2-year *KM*, and (iv) 3-year *KM*. The results of the analysis show that schools that have implemented the Independent Curriculum have higher odds of literacy achievement than schools that have not implemented it. Compared to non-*KM* schools, schools that implemented *KM* for 1 year had an odds ratio of 1.20 (20% increase in odds), 1.70 for 2 years of KM (70% increase in odds), and 2.54 for 3 years of *KM*, which shows that the odds of literacy achievement are more than double compared to schools that still use the 2013 Curriculum. These findings are consistent with the studies by Zamjani et al. [33] and Aditomo et al. [34], which indicate that the implementation of the Independent Curriculum has a positive impact on improving student learning outcomes, especially in the literacy aspect.

Reflection and improvement of learning by teachers describe the extent to which teachers implement active, meaningful learning and reflect on learning outcomes. This variable does not meet the assumptions of proportional odds, so it is modeled using the non-parallel approach. At the first cut point (comparing the “low” vs. “moderate” or “good” category of literacy), the odds ratio of 1.04 indicates that every one-point increase in reflection and learning improvement scores is associated with a 4% increase in the odds of being in the “moderate” or “good” category. Meanwhile, at the second cut point (comparing the “moderate” vs. “good” category of literacy), the odds ratio of 1.05 indicates an increase in odds of 5% to reach the “good” literacy category. These findings indicate that the quality of reflection and improvement of learning by teachers consistently has a positive impact on school literacy outcomes. These results are in line with previous studies, which emphasize that teacher reflection, both internally and through external feedback, plays an important role in improving the learning process. Teachers who regularly reflect are able to identify strengths and weaknesses in their teaching practice, so they can improve learning methods that are more active, meaningful, and focused, which ultimately contributes to improving student learning outcomes [35].

In this research, there are two predictor variables that do not have a significant influence when building the model, namely learning quality and instructional leadership. Quality of learning is a key factor that influences student learning outcomes. Teachers who implement teaching with appropriate methods, relevant materials, and approaches that can stimulate students' interest and understanding will improve their understanding and learning outcomes. The quality of teaching is not only influenced by teacher competence but also by how lesson material is delivered and how interactions in the classroom take place. However, although teaching quality tends to have a positive relationship with learning outcomes, its insignificance in this study could occur due to the presence of other factors that are more dominant in a particular context, such as student characteristics, school policies, and parental support.

Instructional leadership is an important factor in encouraging improvements in the quality of education in schools. School principals who are effective in carrying out their instructional roles can create a conducive learning environment and encourage teacher performance towards improving student learning outcomes. However, in the context of this research, instructional leadership variables do not emerge as significant predictors of literacy achievements. This alleged insignificance occurs because the influence of instructional leadership may be indirect on student learning outcomes but rather is mediated by factors such as teacher motivation and commitment. This is in line with the findings of Leithwood et al. [36], which state that the contribution of leadership to student learning outcomes is generally indirect and only becomes apparent when analyzed through a complex chain of variables in school organizations. Research by Khan et al. [37] also showed that instructional leadership does not directly influence student academic achievement but rather works through increasing teacher commitment. In addition, Vahidov also emphasized that teacher professional development is an important bridge connecting instructional leadership with student achievement [38]. Therefore, although instructional leadership is a key element in school management, its impact on student learning outcomes is often not visible directly but rather through intermediary roles in the school education ecosystem.

In general, based on the results of data analysis, these findings provide insight that both dimensions of school climate (safety, diversity, and inclusiveness) and learning implementation (reflection and improvement of learning and the type of curriculum used) have a positive influence on school literacy achievements. This shows that a socially supportive learning environment, as well as implementing quality learning, plays an important role in improving school literacy achievements.

4. CONCLUSION

Based on the research results, it can be concluded that the use of the ordinal logistic regression method, based on the backward stepwise approach with partial proportional odds (PPO), is an effective technique for identifying factors that influence student literacy achievement at the elementary school level. Significant variables in the model include reflection and improvement of learning by teachers, security climate, diversity climate, inclusiveness climate, and the type of curriculum implemented. This model is accurate and meets statistical criteria with low AIC and BIC values. The results of the analysis show that schools that have a better climate, such as a climate of security, diversity, and inclusiveness, as well as schools that implement the Independent Curriculum, have a higher odds ratio of achieving literacy in the “good” category. Apart from that, reflection and improvement in learning carried out by teachers also "positively contribute to increasing literacy, although the effect is different in each achievement category.

This research recommends that policymakers focus on strengthening school climate and improving the quality of learning to support better literacy achievement in elementary schools. Although the variables of learning quality and principal leadership do not show a significant influence in the model, this shows the need for a more in-depth approach in evaluating and developing these factors. Policy makers are expected to consider steps to increase teacher competency and empower school principals, especially in aspects that are more directly related to increasing literacy.

5. REFERENCES

[1] A. Aditomo, Rahmawati, N. Felicia, N. Shihab, and F. Handayani, "Kajian Akademik dan Rekomendasi Reformasi Sistem Asesmen Nasional," *Badan Penelit. dan Pengemb. dan Perbukuan Kementeri. Pendidik. dan Kebud.*, pp. 1–26, 2019.

[2] N. Hafizha and R. Rakhmania, "Dampak Program Penguatan Literasi pada Hasil Asesmen Kompetensi Minimum di Sekolah Dasar," *J. Basicedu*, vol. 8, no. 1, pp. 171–179, 2024, doi: 10.31004/basicedu.v8i1.6907.

[3] OECD, *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. Paris: OECD Publishing, 2019. doi: 10.1787/acd78851-en.

[4] D. W. J. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, New Jersey: John Wiley & Sons, 2013.

[5] A. Agresti, "Categorical Data Analysis," John Wiley & Sons, 2002. doi: 10.1002/0471249688.

[6] A. Agresti, *Analysis of Ordinal Categorical Data*. John Wiley & Sons, 2010. doi: 10.1002/9780470594001.

[7] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. John Wiley & Sons, 2000. doi: 10.1002/0471722146.

[8] K. N. Khikmah, I. Indahwati, A. Fitrianto, E. Erfiani, and R. Amelia, "Backwards Stepwise Binary Logistic Regression for Determination Population Growth Rate Factor in Java Island," *Jambura J. Math.*, vol. 4, no. 2, pp. 177–187, 2022, doi: 10.34312/jjom.v4i2.13529.

[9] M. Mayawi, N. Nurhayati, T. Talib, A. W. Bustan, and N. S. Laamena, "Ordinal Logistic Regression Analysis of Factors that Affecting the Blood Sugar Levels Diabetes Mellitus Patients," *Pattimura Int. J. Math.*, vol. 2, no. 1, pp. 33–42, 2023, doi: 10.30598/pijmathvol2iss1pp33-42.

[10] H. Wang, "Ordinal Logistic Regression Analysis in Effective Teaching Practices," *J. Math. Stat.*, vol. 20, no. 1, pp. 13–17, 2024, doi: 10.3844/jmssp.2024.13.17.

[11] Kementerian Pendidikan dan Kebudayaan, *Framework asesmen kompetensi minimum dan implikasinya pada pembelajaran*. 2020. [Online]. Available: https://pusmendik.kemdikbud.go.id

[12] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning (2nd ed.)*. Springer, 2021.

[13] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. O'Reilly Media, 2019.

[14] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York: Springer, 2013. doi: 10.1007/978-1-4614-6849-3.

[15] V. Kumar *et al.*, "Addressing Binary Classification over Class Imbalanced Clinical Datasets Using Computationally Intelligent Techniques," *Healthc.*, vol. 10, no. 7, pp. 1–28, 2022, doi: 10.3390/healthcare10071293.

[16] D. Gujarati and D. Porter, *Basic Econometrics*. New York: McGraw-Hill, 2009.

[17] A. Fitrianto and S. H. Xin, "Comparisons Between Robust Regression Approaches in the Presence of Outliers and High Leverage Points," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 1, pp. 243–252, 2022, doi: 10.30598/barekengvol16iss1pp241-250.

[18] M. A. Bujang, N. Sa'at, T. M. I. Tg Abu Bakar Sidik, and L. C. Joo, "Sample size guidelines for logistic regression from observational studies with large population," *Malaysian J. Med. Sci.*, vol. 25, no. 4, pp. 122–130, 2018, doi: 10.21315/mjms2018.25.4.12.

[19] F. E. Harrell, “Regression Modeling Strategies,” 2nd ed., New York: Springer, 2015. doi: 10.1007/978-3-319-19425-7.

[20] A. Liu, H. He, X. M. Tu, and W. Tang, “On testing proportional odds assumptions for proportional odds models,” *Gen. Psychiatry*, vol. 36, no. 3, pp. 1–6, 2023, doi: 10.1136/gpsych-2023-101048.

[21] B. Peterson and F. E. Harrell, “Partial Proportional Odds Models for Ordinal Response Variables,” *Appl. Stat.*, vol. 39, no. 2, p. 205, 1990, doi: 10.2307/2347760.

[22] Miftha Delinda and Devni Prima Sari, “University Election Analysis Logistic Regression Approach with Dummy and Ordinal Variables,” *Math. J. Model. Forecast.*, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.24036/mjmf.v1i2.13.

[23] I. Ghozali, “Aplikasi analisis multivariate dengan program IBM SPSS 25 Edisi 9,” in *Aplikasi analisis multivariate dengan program IBM SPSS 25 Edisi 9*, Semarang: Badan Penerbit Universitas Diponegoro, 2018.

[24] B. Santoso, H. Wijayanto, K. A. Notodiputro, and B. Sartono, “Synthetic Over Sampling Methods for Handling Class Imbalanced Problems : A review,” *IOP Conf. Ser. Earth Environ. Sci. 58 (*, 2017, doi: 10.1088/1755-1315/5.

[25] M. Ghanem *et al.*, “Limitations in Evaluating Machine Learning Models for Imbalanced Binary Outcome Classification in Spine Surgery: A Systematic Review,” *Brain Sci.*, vol. 13, no. 12, 2023, doi: 10.3390/brainsci13121723.

[26] B. Sartomo and H. Dharmawan, *Pemodelan Prediktif Berbasis Pohon Klasifikasi*. PT Penerbit IPB Press, 2023.

[27] F. Jim, “Guidelines for Removing and Handling Outliers in Data.” https://statisticsbyjim.com/basics/remove-outliers/

[28] N. A. Asriyana, R. Purbayati, and B. Barnas, “Determinan Capital Adequacy Ratio Sebagai Indikator Ketahanan Modal Pada Bank Umum Syariah di Indonesia,” *Indones. J. Econ. Manag.*, vol. 2, no. 2, pp. 277–295, 2022, doi: 10.35313/ijem.v2i2.3697.

[29] J. . Long, *Regression Models for Categorical and Limited Dependent Variables*. London: SAGE, 1997.

[30] B. Kutsyuruba, D. A. Klinger, and A. Hussain, “Relationships among school climate, school safety, and student achievement and well-being: a review of the literature,” *Rev. Educ.*, vol. 3, no. 2, pp. 103–135, 2015, doi: 10.1002/rev3.3043.

[31] A. Simanungkalit and S. Habeahan, “Pengaruh Implementasi Nilai Kebinekaan Terhadap Pembentukan Perilaku Toleransi Siswa,” *J. Pendidik. Islam Anak Usia Dini*, vol. 6, pp. 1147–1161, 2024, doi: 10.36088/assabiqun.v6i6.5414.

[32] A. Rafikayati, L. I. Badiah, and B. Soedarmadji, “Pengaruh Implementasi Layanan Bimbingan Pribadi Sosial Terhadap Penyesuaian Diri Anak Berkebutuhan Khusus (Abk) Di Sekolah Inklusif Sman 10 Surabaya,” *Buana Pendidik. J. Fak. Kegur. dan Ilmu Pendidik.*, vol. 14, no. 26, pp. 151–157, 2018, doi: 10.36456/bp.vol14.no26.a1680.

[33] I. Zamjani *et al.*, *Implementasi Kurikulum Merdeka dan Peningkatan Kemampuan Literasi dan Numerasi Peserta Didik*. Pusat Standar dan Kebijakan Pendidikan (PSKP), 2024.

[34] A. Aditomo *et al.*, *Pulih Bersama Pulih Lebih Kuat: Tren dan Tantangan Pemerataan Hasil Belajar di Indonesia 2021–2023*. Pusat Standar dan Kebijakan Pendidikan (PSKP), 2025.

[35] S. S. Pratiwi, J. H. G. Purwasih, D. S. Rozakiyah, and R. A. Rahma, *Teacher Self-reflection for Learning Improvement*, no. Iceiss. Atlantis Press SARL, 2023. doi: 10.2991/978-2-38476-190-6_19.

[36] K. Leithwood, K. Seashore Louis, S. Anderson, and K. Wahlstrom, *How Leadership Influences Student Learning*. The Wallace Foundation, 2004.

[37] A. A. Khan *et al.*, “Instructional leadership and students academic performance: Mediating effects of teacher’s organizational commitment,” *Int. J. Learn. Teach. Educ. Res.*, vol. 19, no. 10, pp. 233–247, 2020, doi: 10.26803/IJLTER.19.10.13.

[38] H. Vahidov, “The Impact of Instructional Leadership on Teacher Professional Development and Its Effect on Student Learning Outcomes,” vol. 2, no. 12, pp. 16–25, 2024.