



 **1 J E M S**

**INDONESIAN
JOURNAL OF
EDUCATION &
MATHEMATICAL
SCIENCE**

**UNIVERSITAS MUHAMMADIYAH
SUMATERA UTARA**



ISSN 2715 - 985X
(ONLINE)

ISSN
2715-985X

Editor-in-Chief
Akrim

Managing Editor
Asrar Aspia Manurung

Editorial Board

Kerry Kennedy
Mohd. Roslan bin Mohd. Nor
Nurulhuda Abd Rahman
Terry Lovat
Ali Mahmudi
Syahriani Sirait
M. Romi Syahputra
Yumira Simamora
Tua Halomoan Harahap.

Reviewers

Irdyanti binti Mat Nashir
Lilla Adulyasa
Anton Abdulbasah Kamil
Balamuralithara Balakrishnan
Muhammad Zain Musa
Herman Mawengkang
Tatang Herman
Dian Armanto
Wahyu Widada
Ahmad
Marah Doly Nasution
Ella Andhany
Henra Saputra Tanjung

Publisher
Universitas Muhammadiyah
Sumatera Utara

Vol 7 No 3 2026

Table of Contents

215	Evaluation of Gaussian Process Regression, Support Vector Regression, and K-Nearest Neighbors for Predicting Concrete Compressive Strength at Various Curing Ages Elmanani Simamora, Abil Mansyur, Muhammad Badzlan Darari, Suvriadi Panggabean, Rizki Habibi
231	The Effectiveness of PjBL Learning on Interest and Mathematical Creative Thinking in Determining the Number of Construction Workers for 11th Grade Students in the Construction Department Mozard Aryan Hukunala, Rizky Fatmalasari L., Agung K. Henaulu, Tri Siwi Nasrulyati
242	Enhancing Fifth Graders' Flat Geometry Problem-Solving Skills through Stem-Based 'Squid Math' Gamified Learning Media Yuni Mariyati, Intan Dwi Hastuti, Reny Widya Astuti
254	Wordwall Based Learning Media in Earth Structure Learning: Student Engagement and Students' Perceptions Dede Saepuloh, Lina Mafazi Salsabila, Hurriyah Salasih F., Hairunisa Lestari, Herman Supriyatna, Yuli Nurmawati Dewi, Supardi U.S., Dasmo Dasmo
262	MaturiKKA-SD: Development of a Maturity Assessment Instrument for CAI Implementation in Indonesian Primary Schools Mirlanda Diaz Pratiwi, Cepi Riyana, Riche Cynthia Johan
276	Evaluating the Accuracy of ARIMA vs. Holt's Double Exponential Smoothing in Predicting Indonesia's Gold Price Trends Septi Ika Apriliani, Laila Fitriana
290	Mathematics in the Nyadran Tradition and How Students Solve Permutation Problems Based on It Fitria Sulistyowati, Al Ainaa Almardhiyyah, Pardimin Pardimin, Wikan Budi Utami, Luc Bergman
305	Development of a Physics Diagnostic Assessment Framework based on Solo Taxonomy for Senior High School Student Sabani, Irham Ramadhani, Wawan Bunawan, Seila Agustina, Elang Mulia Lesmana, Dyah Inggit Murtiningrum
319	Vocational and Mathematical Knowledge: Designing Hybrid Literacy Tasks in the Context of Food Security Ai Tusi Fatimah, Agus Yuniawan Isyanto, Euis Erlin
329	AI-Assisted Ethnomathematics Gamification through Traditional Stilt Games for Geometry Learning in Elementary Education Tasya Fortuna Aurora, Beti Istanti Suwandayani

ISSN
2715-985X

Editor-in-Chief
Akrim

Managing Editor
Asrar Aspia Manurung

Editorial Board

Kerry Kennedy
Mohd. Roslan bin Mohd. Nor
Nurulhuda Abd Rahman
Terry Lovat
Ali Mahmudi
Syahriani Sirait
M. Romi Syahputra
Yumira Simamora
Tua Halomoan Harahap.

Reviewers

Irdyanti binti Mat Nashir
Lilla Adulyasa
Anton Abdulbasah Kamil
Balamuralithara Balakrishnan
Muhammad Zain Musa
Herman Mawengkang
Tatang Herman
Dian Armanto
Wahyu Widada
Ahmad
Marah Doly Nasution
Ella Andhany
Henra Saputra Tanjung

Publisher
Universitas Muhammadiyah
Sumatera Utara

Vol 7 No 3 2026

Table of Contents

339	Development of a Flipbook-Based E-Module for Junior High School Geometry Learning: Its Effectiveness in Enhancing Student Engagement and Learning Motivation Sitti Zuhaerah Thalhah, Dwy Risky Arifanti, Sumardin Raupu, Irma T.
354	Spiritual Economic Dynamics: A Nonlinear ODE Model of Giving, Faith, and Wealth Growth with Threshold Transition from Give to Take Sutanto, Deni Setiawan
363	A Data-Driven Travel-Time Scenario Framework for the Preliminary Analysis of School Meal Distribution as Education Support Muliawan Firdaus, Marah Doly Nasution
376	Ethnomodeling Indigenous Knowledge for Decolonizing Mathematics Education: Insights from The Passaungang Tau Heritage Site Muhammad Nurhusain, Rizky Ramadhana, Rezki Novianti, Abdul Hadi
390	The Muffled Echoes of Reflection: Mapping the Fading Resonance of Mathematics Reflective Thinking Among Anxious Students Muntazhimah, Indah Widiati, Joko Dwi Prasetyo

Evaluation of Gaussian Process Regression, Support Vector Regression, and K-Nearest Neighbors for Predicting Concrete Compressive Strength at Various Curing Ages

Elmanani Simamora¹, Abil Mansyur², Muhammad Badzlan Darari³, Suvriadi Panggabean⁴, Rizki Habibi⁵

^{1,2,3,4,5}Department of Mathematics, Faculty of Mathematics, Universitas Negeri Medan, Medan, Indonesia

¹elmanani_simamora@unimed.ac.id

ABSTRACT

Accurate prediction of concrete compressive strength is essential for quality control, structural evaluation, and concrete mix optimization. However, the relationships among mixture composition, curing age, and compressive strength are often nonlinear and heterogeneous. This study evaluates four prediction models—Native Gaussian Process Regression (NGPR), Gaussian Process Regression with Conformal Intervals (GPR-CI), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN)—for predicting concrete compressive strength at various curing ages. The predictor set includes mixture-composition variables, curing age, and material-relevant derived features. Model performance is assessed using Leave-One-Age-Out validation and repeated stratified K-fold cross-validation by age, based on root mean square error, mean absolute error, coefficient of determination, empirical coverage, and average interval length. The results show that NGPR provides the most consistent overall performance across both evaluation schemes. Under the Leave-One-Age-Out scheme, NGPR achieves better point-prediction accuracy and higher empirical coverage than GPR-CI, whereas under repeated age-stratified cross-validation, GPR-CI achieves slightly higher coverage at the cost of wider prediction intervals. Overall, NGPR provides the most favorable balance among point-prediction accuracy, interval calibration, and interval efficiency for heterogeneous concrete compressive-strength data across curing ages.

Keywords: Concrete Compressive Strength, Gaussian Process Regression, Support Vector Regression, K-Nearest Neighbors, Prediction Interval, Curing Age



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Elmanani Simamora,
Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Universitas Negeri Medan
Jalan William Iskandar Ps. V, Medan, Indonesia, 20221
elmanani_simamora@unimed.ac.id

1. INTRODUCTION

One of the most important properties of concrete in structural design, quality control, and material performance evaluation is its compressive strength. This measure is used in the profession of civil engineering to assess the acceptability of the mixture, compliance with standards, and to make technical decisions in production and execution. However, the relationship between mixture composition and compressive strength is not simple. Cement, water, slag, fly ash, superplasticizer, aggregate, and curing age work together, so that the resulting response tends to be nonlinear and influenced by interactions between materials (Yeh, 1998; Yeh, 2006; Feng et al., 2020; Kaloop et al., 2020; Beskopylny et al., 2022). Classical and recent literature both show that curing age, water content, water-to-binder ratio, and additional cementitious materials remain the dominant variables in the formation of compressive strength (Yeh, 1998; Yeh, 2006; Beskopylny et al., 2022; Liu & Sun, 2023).

These developments have driven a shift from conventional empirical models to machine learning-based approaches. Various studies have used ensemble models, interpretable models, and kernel methods to predict concrete compressive strength (Feng et al., 2020; Kaloop et al., 2020; Beskopylny et al., 2022; Liu & Sun, 2023; Elhishi et al., 2023; Zhang et al., 2024; Angiulli et al., 2024; Miao et al., 2025). Boosting approaches can capture complex nonlinear relationships with high accuracy (Feng et al., 2020; Kaloop et al., 2020). Interpretable models such as the Explainable Boosting Machine and SHAP-based approaches have also shown that good predictions can still be achieved with explanations

of variable contributions (Liu & Sun, 2023; Elhishi et al., 2023; Zhang et al., 2024). On the other hand, comparative studies using concrete data still rank Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) as important benchmarks because they represent two distinct strategies for interpreting nonlinear data structures (Beskopylny et al., 2022; Angiulli et al., 2024). Thus, the main question is no longer which model has the smallest error. The more appropriate question is which model remains stable when the data structure changes, and which model provides a reasonable balance between point accuracy and interval quality.

In recent years, attention has also shifted towards models that are not only accurate but also more informative in terms of prediction uncertainty. At this point, Gaussian Process Regression (GPR) becomes relevant. GPR is not only flexible in modeling nonlinear relationships but also naturally generates predictive distributions (Miao et al., 2025; Xie et al., 2025). This characteristic is important when the analysis objective is not limited to point predictions, but also includes prediction intervals and actual coverage. Recent literature on multi-binder concrete, non-destructive testing, and uncertainty-aware compressive strength estimation shows that uncertainty is increasingly difficult to ignore (Miao et al., 2025; Xie et al., 2025; Tamuly & Nava, 2025; Fu et al., 2025). Correspondingly, conformal prediction is increasingly used to generate more reliable intervals without relying heavily on distributional assumptions about the model (Tamuly & Nava, 2025; Fu et al., 2025).

However, one point is still often not discussed explicitly. Many studies report Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and other metrics using a common validation scheme, but do not always treat the curing age as an explicit evaluation criterion. In fact, in concrete compressive strength data, age is often recorded at discrete points, and the number of observations is not evenly distributed across age groups. Recent literature on high-performance concrete, high-strength concrete, and multi-dataset analysis also shows that curing age remains a dominant variable, and its distribution can affect the model's behavior (Zhang et al., 2024; Shaaban et al., 2025; Mobasheri et al., 2025; Xiao et al., 2025; Olvera-Mayorga et al., 2025). If so, the usual random evaluation is likely too optimistic, especially when the dominant age group consistently appears in both the training and test data.

The need is particularly relevant to the data in this study. The exploratory results indicate that the compressive strength distribution is unimodal and right-skewed, while variation with curing age is not homogeneous. Median compressive strength tends to increase from early to intermediate ages, particularly up to approximately 28 days, but after that, the pattern of increase is no longer consistent across age groups. Furthermore, the ages in the data are recorded at discrete points rather than on a continuous scale, so the age–compressive strength relationship is better understood as gradual and clustered. The 28-day age group is particularly significant because it has the largest number of observations and is often regarded as the primary reference for assessing concrete compressive strength in civil engineering practice. But some of the earlier ages have significantly less data. The circumstances indicate that age in the study is not only a predictor column but also part of the evaluation structure that needs to be expressly addressed.

Based on this, we focus the comparison on Native GPR (NGPR), GPR with Conformal Intervals (GPR-CI), SVR, and KNN. These four models were chosen because they represent different approaches to nonlinear regression. The GPR family was placed more firmly from the outset because it is substantively closest to the needs of this study. This model is flexible to nonlinear relationships and supports the formation of prediction distributions. In this study, the GPR family is divided into two forms: NGPR and GPR-CI, because the mechanisms for forming prediction intervals differ. Meanwhile, SVR is a well-established kernel-based comparator in nonlinear regression, and KNN is a simple local proximity-based comparator that remains relevant to tabular data (Beskopylny et al., 2022; Miao et al., 2025; Xie et al., 2025; Tamuly & Nava, 2025; Fu et al., 2025).

The research problem is formulated in two directions. First, whether different prediction models will show the same stability when the data is evaluated by explicitly taking into account the curing age. Second, whether a model that is good in terms of point predictions also remains good in terms of prediction intervals. To solve these problems, we conduct assessments using two complementary frameworks: Leave-One-Age-Out (LOAO) and repeated stratified K-fold by Age. LOAO is used to examine the generalization capacity when one age group is not involved in the training at all. Conversely, repeated stratified K-fold by Age is used to obtain a more empirically stable picture of performance

because all age groups are still represented in each fold. This choice is in line with this study, as the results show that model performance varies with the age structure encountered, and that the quality of the interval cannot be assessed solely from RMSE or R^2 .

In the context of prediction intervals, this study also starts from the view that the best model cannot be judged solely by the narrowest interval or the highest coverage. Intervals that are too narrow may appear sharp but fail to adequately cover the actual values. Conversely, intervals that are too wide are indeed safe, but less efficient. The literature on probabilistic prediction emphasizes that interval quality needs to be assessed from two perspectives simultaneously: calibration and sharpness (Gneiting et al., 2007; Tamuly & Nava, 2025; Dong & Zhang, 2025; Almutairi, 2025). The principle is very relevant in this study, because the results show a clear trade-off between interval length and actual coverage, both in the LOAO scheme and in the repeated stratified K-fold by Age.

Based on these factors, the purpose of this work is to evaluate the performance of NGPR, GPR-CI, SVR, and KNN in predicting concrete compressive strength at various curing ages, with an emphasis on two aspects: point prediction accuracy and prediction interval quality. The proposed working hypothesis is that the GPR family will display more stability than SVR and KNN when evaluated on a scheme sensitive to age-group shifts. At the same time, the difference between NGPR and GPR-CI will predominantly show as in trade-off between interval efficiency and coverage conservatism. The expected result is a stronger empirical basis for selecting concrete compressive strength prediction models that are not only good in aggregate but also remain reliable as the curing age structure changes and when prediction uncertainty needs to be read explicitly.

The novelty of this study lies not in the four algorithms themselves, which are well established, but in the integrated age-structured framework used to evaluate both predictive accuracy and uncertainty. The framework combines Leave-One-Age-Out validation to assess generalization to curing-age groups excluded from model training with repeated age-stratified cross-validation to assess performance across resampled partitions that preserve the curing-age distribution as far as the group sizes permit. This complementary design enables the four models to be compared under two distinct age-based validation settings in terms of point-prediction accuracy, interval calibration as measured by empirical coverage, and interval sharpness.

2. RESEARCH METHOD

A. Data and predictor formation

We use concrete compressive strength data that have also been used in several regression analyses, both in multiple linear regression and in robust regression and bootstrap studies (Rana et al., 2012; Simamora et al., 2025), with one response variable, namely concrete compressive strength, and several predictor variables representing the mixture composition and curing age. In general, the data is expressed as $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, where y_i represents the concrete compressive strength at the i -th observation, while $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents the predictor vector. The relationship between the response and predictors is written in equation (2.1) as

$$y_i = f(x_i) + \varepsilon_i, \quad (2.1)$$

with $f(\cdot)$ the unknown regression function and ε_i representing random errors. Curing age is retained as the main predictor. In addition to the basic variables, materially relevant derived features are also formed, namely total binder, water to binder ratio, indicators of the presence of slag, fly ash, and superplasticizer, and $\log(\text{age})$ transformation. The formation of these features is used to enrich the representation of the nonlinear relationship between mixture composition, curing age, and concrete compressive strength.

Before modeling, data exploration was conducted using compressive strength histograms, boxplots by curing age, and scatterplots between curing age and compressive strength. This stage was used to interpret the response distribution, the imbalance in the number of observations between ages, and the initial pattern of the relationship between age and compressive strength. This arrangement was maintained because it was consistent with the research results, which interpreted the data based on the global distribution, variation between ages, and the pattern of the relationship between age and compressive strength.

B. Preprocessing and validation design

All predictors are standardized using formula (2.2) so that models are compared on the same input scale. Standardization is performed only based on the training data statistics for each partition, so that information from the test data does not enter the training stage. If X_{tr} and X_{te} represent the training and test predictor matrices, respectively, then the transformation is performed with

$$Z_{tr} = \frac{X_{tr} - \bar{X}_{tr}}{s_{tr}}, \quad Z_{te} = \frac{X_{te} - \bar{X}_{tr}}{s_{tr}}, \quad (2.2)$$

where $\bar{X}_{tr}$ and s_{tr} denote the mean vector and standard deviation vector computed from the training predictors. The same training statistics are then applied to the test predictors to prevent information leakage from the test data. The evaluation is conducted within a multilevel validation framework. Two schemes are used in the outer layer. The first scheme is LOAO. In this scheme, all observations from a single curing age are removed from the training data and used as test data. If the set of unique ages is written as $\mathcal{A} = \{a_1, \dots, a_G\}$, then at the g -th iteration,

$$\mathcal{D}_{test}^{(g)} = \{(x_i, y_i) : t_i = a_g\}, \quad \mathcal{D}_{train}^{(g)} = \mathcal{D} \setminus \mathcal{D}_{test}^{(g)} \quad (2.3)$$

where t_i denotes the curing age of the i -th observation. The scheme in (2.3) was chosen because it provides a rigorous generalization test when the model has to predict age groups that do not appear at all in the training.

The second scheme is a repeated stratified K-fold by Age. In this approach, the data is divided into five folds, stratified by curing age, and then the division is repeated five times with different random partitions. Within each outer partition, five-fold cross-validation is performed on the training data to select the best hyperparameters. This design is appropriate for the data structure with an unbalanced age distribution. It is also in line with research findings, as stratified evaluation by Age yields a more stable performance picture than LOAO.

C. Prediction models and prediction intervals

The study compares four models: NGPR, GPR-CI, SVR, and KNN. Conceptually, these four models represent three families of approaches. The Gaussian process family is divided into two because the mechanisms for forming prediction intervals differ. In NGPR, point predictions and prediction intervals are both derived from the model's predictive distribution. In contrast, in GPR-CI, fixed-point predictions are obtained from a Gaussian process model, whereas prediction intervals are formed via a conformal approach. Meanwhile, SVR and KNN are used as comparison models to assess the stability of the kernel- and local-proximity-based approaches on concrete compressive strength data.

In the GPR model, the latent function $f_{GP}(x)$ in equation (2.4) is assumed to follow a Gaussian process,

$$f_{GP}(x) \sim GP(\mu_{GP}(x), k(x, x')), \quad (2.4)$$

where $\mu_{GP}(x)$ represents the mean function and $k(x, x')$ represents the covariance or kernel function. The mean function is assumed to be constant. The observation model is then expressed in equation (2.5) as

$$y_i = f_{GP}(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_n^2), \quad i = 1, 2, \dots, n, \quad (2.5)$$

where y_i denotes the concrete compressive strength at the i -th observation, $f_{GP}(x_i)$ denotes the latent function at point x_i and σ_n^2 denotes the variance of the observation noise. For a new prediction point x_* , the posterior mean of the latent function is given by

$$\hat{\mu}_{f|\mathcal{D}}(x_*) = \mu_{GP}(x_*) + k_*^\top (K + \sigma_n^2 I)^{-1} (y - \mu), \quad (2.6)$$

and the posterior variance of the latent function is given by

$$\hat{\sigma}_{f|\mathcal{D}}^2(x_*) = k(x_*, x_*) - k_*^\top (K + \sigma_n^2 I)^{-1} k_*. \quad (2.7)$$

Because this study evaluates the prediction interval against the actual observed value, the predictive variance for the new observation y_* is written as

$$\hat{\sigma}_{y|\mathcal{D}}^2(x_*) = \hat{\sigma}_{f|\mathcal{D}}^2(x_*) + \sigma_n^2. \quad (2.8)$$

In Equations (2.6)–(2.8), $K = [k(x_i, x_j)]_{i,j=1}^n$ is the $n \times n$ covariance matrix containing the kernel values for all pairs of training points, while $k_* = (k(x_1, x_*), k(x_2, x_*), \dots, k(x_n, x_*))^T$ denotes the covariance vector between the new point x_* and all training points. Furthermore, $y = (y_1, y_2, \dots, y_n)^T$ denotes the observed response vector and $\mu = (\mu_{GP}(x_1), \mu_{GP}(x_2), \dots, \mu_{GP}(x_n))^T$ denotes the mean prior vector over all training points. Thus, $\hat{\mu}_{f|D}(x_*)$ in Equation (2.6) represents the point prediction, while $\hat{\sigma}_{f|D}^2(x_*)$ and $\hat{\sigma}_{y|D}^2(x_*)$ in Equations (2.7) and (2.8) represent the prediction uncertainty for the latent function and the new observation.

During model tuning, several standard kernels were considered, namely the Squared Exponential (SE), Matérn 3/2, Matérn 5/2, and Rational Quadratic (RQ) kernels. For two predictor vectors x and x' , and $r = \|x - x'\|$, the four kernels are defined as

$$\begin{aligned} k_{SE}(x, x') &= \sigma_k^2 \exp\left(-\frac{r^2}{2\ell^2}\right), \\ k_{Mat\ 3/2}(x, x') &= \sigma_k^2 \left(1 + \frac{\sqrt{3}r}{\ell}\right) \exp\left(-\frac{\sqrt{3}r}{\ell}\right), \\ k_{Mat\ 5/2}(x, x') &= \sigma_k^2 \left(1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5}r}{\ell}\right), \text{ and} \\ k_{RQ}(x, x') &= \sigma_k^2 \left(1 + \frac{r^2}{2\gamma\ell^2}\right)^{-\gamma}. \end{aligned} \quad (2.9)$$

In Equation (2.9), σ_k^2 represents the kernel amplitude parameter that governs the magnitude of the latent function variation, ℓ is the scale length parameter, and γ is the shape parameter in the rational quadratic kernel. The squared exponential kernel produces a very smooth function. The Matérn 3/2 and Matérn 5/2 kernels provide flexibility with a more moderate level of smoothness, while the rational quadratic kernel can accommodate variations in scale length in the data. The best kernel is then selected based on the smallest average RMSE value in the inner cross-validation.

In NGPR, point predictions and prediction intervals are both derived from the model's predictive distribution. For a nominal confidence level of $100(1 - \alpha)\%$, the prediction interval for a new observation y_* is written as

$$\hat{\mu}_{f|D}(x_*) \pm z_{1-\alpha/2} \hat{\sigma}_{y|D}(x_*), \text{ where } \hat{\sigma}_{y|D}(x_*) = \sqrt{\hat{\sigma}_{f|D}^2(x_*)}, \quad (2.10)$$

with α representing the miscoverage level of the prediction interval.

In contrast, in GPR-CI, point predictions are derived from a Gaussian process, whereas prediction intervals are formed via split conformal prediction. Suppose the external training data is divided into the main training data and the calibration data. If the absolute residuals on the calibration data are expressed as

$$s_j = |y_j^{\text{cal}} - \hat{y}_j^{\text{cal}}|, \quad j = 1, 2, \dots, m, \quad (2.11)$$

then the conformal quantile is determined as

$$q_{1-\alpha} = S_{(\lceil(m+1)(1-\alpha)\rceil)}, \quad (2.12)$$

with m representing the number of observations in the calibration data, α denoting the miscoverage level and $s_{(1)} \leq s_{(2)} \leq \dots \leq s_{(m)}$ are the ordered calibration residuals. The conformal prediction interval for the new point x_* is then written as

$$\hat{y}(x_*) \pm q_{1-\alpha} \quad (2.13)$$

Thus, the difference between NGPR and GPR-CI lies in the source of the prediction intervals. In the first model, the intervals are derived from the Gaussian process predictive distribution as in equation (2.10). In the second model, the intervals are derived from residual calibration through conformal prediction as in equation (2.13).

SVR is used as a kernel-based comparator. In general, SVR forms a prediction function

$$\hat{f}_{SVR}(x_*) = w^\top \phi(x) + b, \quad (2.14)$$

where $\phi(x)$ represents the feature transformation, w the coefficient vector, and b a constant. The penalty parameters, kernel scale, and ε_{SVR} , which controls the width of the epsilon-insensitive tube in SVR, are selected through inner cross-validation. Since the SVR in equation (2.14) does not produce a native predictive distribution, its prediction interval is formed using the same conformal procedure as in Equations (2.11)–(2.13).

KNN is used as a local proximity-based comparator. If weighting is used, the point prediction for a new point x_* is written as

$$\hat{f}_{KNN}(x_*) = \sum_{j=1}^{k_{KNN}} w_j y_{(j)}, \quad (2.15)$$

with $y_{(j)}$ denoting the response of the j -th neighbor and the weight

$$w_j = \frac{1/d_j}{\sum_{h=1}^k 1/d_h}. \quad (2.16)$$

In equation (2.16), the calculation of d_j is the distance between x and the j -th neighbor. Meanwhile, the number of neighbors k_{KNN} , the distance metric, and the use of weights are selected via inner cross-validation. Just like SVR, the KNN prediction interval in equation (2.15) is formed using the conformal procedure in equations (2.11)–(2.13).

D. Performance measures

Model performance is assessed from two perspectives: point prediction and prediction interval. For point prediction, RMSE, MAE, and R^2 are used:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.17)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (2.19)$$

For the prediction interval, coverage and Average Interval Length (AIL) are used. If the prediction interval of the i -th observation is $[L_i, U_i]$, then

$$\text{Coverage} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(L_i \leq y_i \leq U_i), \quad (2.20)$$

$$\text{AIL} = \frac{1}{n} \sum_{i=1}^n (U_i - L_i). \quad (2.21)$$

where $\mathbb{I}$ is the indicator function, L and U are the lower and upper prediction limits, respectively. RMSE, MAE, and R^2 were selected because they provide complementary information about point-prediction performance for a continuous engineering response measured in MPa. RMSE assigns a larger penalty to large prediction errors because the residuals are squared before averaging. This property is relevant to concrete compressive-strength prediction because substantial overestimation or underestimation may have more serious implications for quality-control and engineering decisions. MAE measures the typical absolute prediction error in the original MPa scale and is less strongly influenced by a small number of large residuals. Therefore, reporting both RMSE and MAE helps distinguish models that perform consistently from models whose average performance is affected by occasional large errors. The coefficient of determination, R^2 , measures the proportion of the observed variation in compressive strength explained by the model relative to a mean-based prediction. However, R^2 is interpreted jointly with RMSE and MAE rather than as a stand-alone measure, particularly for curing-age groups with small sample sizes, where R^2 may be unstable or negative. These three metrics are widely used in machine-learning studies of concrete compressive-strength prediction because they jointly describe error magnitude and explanatory performance (Beskopylny et al., 2022; Zhang et al., 2024).

Because point-prediction accuracy alone does not quantify predictive uncertainty, coverage and Average Interval Length (AIL) were also evaluated. Coverage measures the empirical proportion of observed compressive-strength values contained within the prediction intervals and therefore indicates interval calibration relative to the nominal coverage level. AIL measures interval sharpness or efficiency, with shorter intervals providing more precise predictions, provided that adequate coverage is maintained. Coverage and AIL must consequently be interpreted together: a narrow interval is not useful when it systematically undercovers the observations, whereas a very wide interval may achieve high coverage but provide limited practical precision. This joint assessment follows the calibration–sharpness principle for probabilistic prediction, according to which prediction intervals should be as narrow as possible subject to satisfactory calibration (Gneiting et al., 2007). The combined use of RMSE, MAE, R^2 , coverage, and AIL is therefore appropriate for this study because it evaluates both the accuracy of predicted concrete compressive strength and the reliability and efficiency of the associated prediction intervals.

E. Analysis procedures

The analysis was conducted in six stages. First, the data was cleaned, and a predictor set was formed. Second, data exploration was conducted to examine the distribution of compressive strength, variation with curing age, and the initial pattern of the relationship between curing age and compressive strength. Third, for each outer partition, the data were split into training and testing sets, and the predictors were standardized using the training data, as in Equation (2.2). Fourth, cross-validation was performed on the inner layer to select the best hyperparameters from NGPR, GPR-CI, SVR, and KNN. Fifth, the best model was retrained on the full outer training data and used to form point predictions and prediction intervals on the test data. Sixth, the prediction results were summarized into RMSE, MAE, R^2 , coverage, and AIL as defined in equations (2.17)–(2.21), both overall and by curing age group.

3. RESULTS AND DISCUSSION

A. Results

Figure 1 presents the distribution of concrete compressive strength, its distribution by curing age, and the relationship between curing age and concrete compressive strength. This visualization provides an initial overview of the data characteristics, variations between age groups, and the trend of the relationship patterns that underlie further analysis. Figure 1(a) shows a unimodal distribution of compressive strength with a rightward skew. The majority of observations fall within the 20–50 MPa range, with the highest concentration around 30–40 MPa. Very low values are present but not dominant, while high values above 60 MPa are relatively few and form a long right tail up to around 80 MPa. This means that the data does not appear perfectly symmetrical and does not appear roughly normal. There is considerable variation, but the main mass remains in the center. Thus, descriptively, this dataset is dominated by medium-strength concrete, with a small proportion of high-strength mixtures.

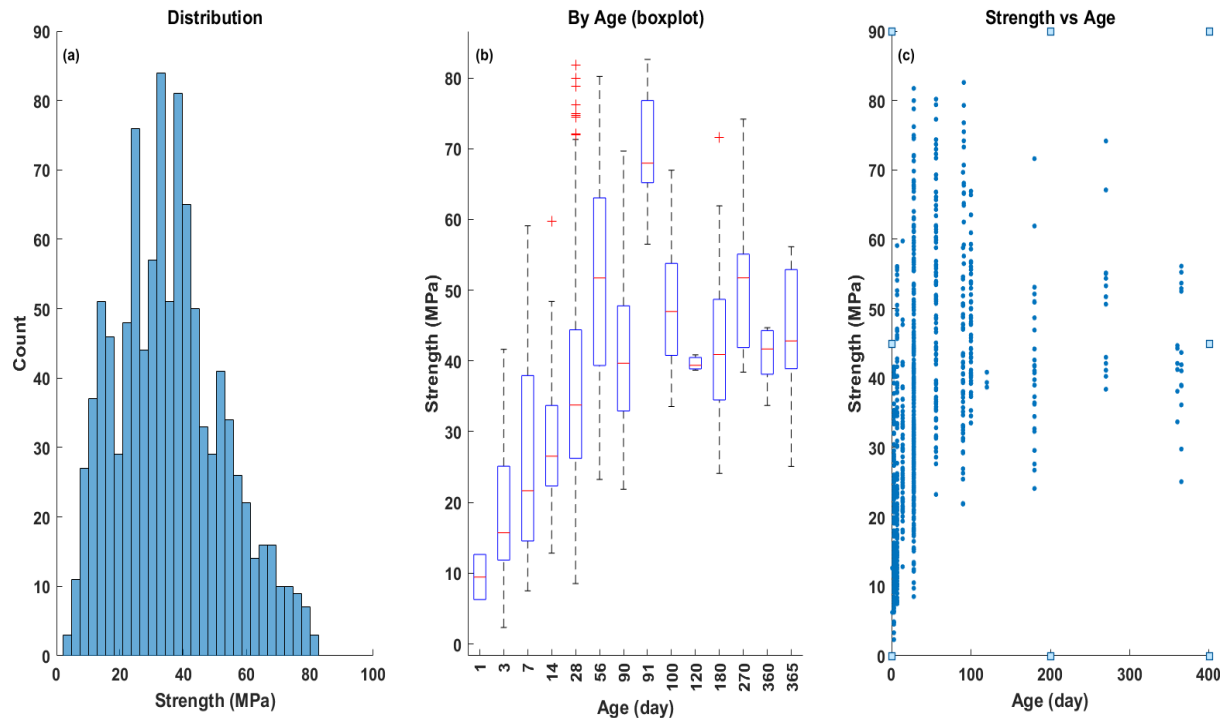


Figure 1: Characteristics of concrete compressive strength distribution and its relationship to curing age.

Figure 1(b) presents the distribution of concrete compressive strength across curing-age groups. The median strength generally increases from early to intermediate curing ages, particularly at 1, 3, 7, 14, and 28 days. However, the distribution within each age group is characterized not only by its median but also by its dispersion, range, and sample size. The comparatively wider interquartile ranges and longer whiskers observed at 28, 56, and 91 days indicate greater within-age variability, whereas the narrow distribution at 120 days should be interpreted cautiously because this age group contains only three observations. Points located beyond the boxplot whiskers represent unusually high or low compressive-strength observations and should not automatically be interpreted as statistically significant differences. Moreover, the observations at different curing ages do not represent repeated measurements of a single concrete mixture. Consequently, the differences among age groups reflect the combined effects of curing age, mixture composition, and unequal group sizes rather than the isolated effect of curing age alone.

Figure 1(c) shows that the relationship between age and compressive strength is positive but not linear. At low ages, particularly up to about 28 days, increasing age corresponds to an increase in the center of the compressive strength distribution. After that, the points do not form a neat ascending line, but rather a vertical distribution at discrete ages. This shows that the ages in this dataset were measured on certain test days rather than across a period. The link between age and compressive strength is more realistically described as a clustered, progressive relationship rather than a single smooth curve. Also, the point density is much higher for the younger ages, while the older ages (180, 270, 360, and 365 days) are much less frequent. Hence, the statistical information across large age groups is quite weak.

Beyond the median patterns shown in Figure 1, Table 1 summarizes the location, dispersion, and range of all variables in the dataset. The substantial differences in scale and variability across variables support the use of fold-specific predictor standardization and confirm that the data cover a broad range of mixture compositions and curing ages.

Table 1. Descriptive statistics of the concrete compressive strength dataset

Variable	Unit	Mean	SD	Minimum	Maximum
Cement	kg/m ³	281.17	104.51	102.00	540.00
Blast-furnace slag	kg/m ³	73.90	86.28	0.00	359.40
Fly ash	kg/m ³	54.19	64.00	0.00	200.10
Water	kg/m ³	181.57	21.36	121.75	247.00
Superplasticizer	kg/m ³	6.20	5.97	0.00	32.20
Coarse aggregate	kg/m ³	972.92	77.75	801.00	1,145.00
Fine aggregate	kg/m ³	773.58	80.18	594.00	992.60
Curing age	day	45.66	63.17	1.00	365.00
Compressive strength	MPa	35.82	16.71	2.33	82.60

The concrete compressive strength has a mean of 35.82 MPa and a standard deviation of 16.71 MPa, with observed values ranging from 2.33 to 82.60 MPa. Curing age has a mean of 45.66 days and a standard deviation of 63.17 days, with values ranging from 1 to 365 days. The broad ranges and relatively large variations observed for several variables indicate substantial heterogeneity in mixture composition and curing-age structure. Pearson correlation analysis shows that compressive strength is positively associated with cement content ($r = 0.498$), superplasticizer ($r = 0.366$), curing age ($r = 0.329$), and blast-furnace slag ($r = 0.135$). Negative correlations are observed for water ($r = -0.290$), fine aggregate ($r = -0.167$), coarse aggregate ($r = -0.165$), and fly ash ($r = -0.106$). All correlations between the predictors and compressive strength were statistically significant ($p < 0.001$).

Among the predictors, the largest absolute pairwise correlations occur between water and superplasticizer ($r = -0.657$), water and fine aggregate ($r = -0.451$), cement and fly ash ($r = -0.397$), and fly ash and superplasticizer ($r = 0.377$). Nevertheless, no predictor pair has an absolute correlation coefficient exceeding 0.80, indicating the absence of extremely high pairwise linear correlation. These findings suggest that some mixture components exhibit moderate interrelationships and that their associations with compressive strength should be evaluated jointly within a multivariable nonlinear modeling framework. Pearson correlations represent marginal linear associations only and should not be interpreted as isolated causal effects or as complete descriptions of potentially nonlinear and interactive relationships.

Table 2 shows the overall evaluation results of the LOAO scheme. In this scheme, NGPR achieves the best performance among the four models compared, with RMSE = 5.6259, MAE = 3.9271, and $R^2 = 0.8865$. These values are slightly better than GPR-CI, which produces RMSE = 5.7086, MAE = 4.0063, and $R^2 = 0.8831$. Meanwhile, SVR and KNN show a more obvious performance decline, with RMSE of 8.1760 and 9.7415, respectively. In terms of prediction intervals, NGPR achieved coverage of 0.9350 with an average interval length of 21.2339, while GPR-CI produced a shorter interval, namely 16.8252, but coverage dropped to 0.8379. The results in Table 2 show that in this most stringent evaluation, NGPR provides a better balance between point prediction accuracy and interval coverage.

Table 2. Overall performance of the concrete compressive strength prediction model in the LOAO scheme.

Model	RMSE	MAE	R^2	Coverage	AIL
NGPR	5.625882979	3.927106634	0.886479325	0.934951456	21.23388036
GPR-CI	5.708635369	4.006284048	0.883115161	0.837864078	16.82524068
SVR	8.176031413	5.774550776	0.760238924	0.831067961	21.10809974
KNN	9.741517809	7.646083052	0.65963348	0.842718447	27.80885379

Table 3 shows that the LOAO results are not completely uniform across age groups. At 3 days, GPR-CI slightly outperforms NGPR, with an RMSE of 3.9543 versus 4.1322 and an R^2 of 0.8380 versus 0.8231. At 7 days, the two GPR models remain relatively close, while SVR and KNN are weaker. At 14 days, NGPR again outperforms SVR and KNN, although the difference with respect to GPR-CI is not significant. The 28-day age group, containing 425 observations, shows that NGPR remains

superior with an RMSE of 7.3793 and an R^2 of 0.7478, while the other models produce larger errors. At 56 days, GPR-CI slightly outperforms NGPR. At 90 days, GPR-CI also produces a smaller RMSE. Interestingly, at 180 days, KNN actually produces the highest R^2 , at 0.8271. The pattern in Table 3 shows that in the LOAO scheme, model performance is highly influenced by the age excluded from the training data.

Table 3. Performance of concrete compressive strength prediction model according to curing age in the LOAO scheme.

Age	Model	n	RMSE	MAE	R^2	Coverage	AIL
1	NGPR	2	6.400516256	6.343538972	-3.03745553	1	22.50918255
1	GPR-CI	2	7.519051348	7.499915119	-4.57190679	1	17.70379403
1	SVR	2	8.859054561	8.826766236	-6.73486066	1	20.11373996
1	KNN	2	18.91734521	17.59533536	-34.2694039	0.5	21.32541504
3	NGPR	134	4.132208113	3.051034952	0.823140308	1	24.99697648
3	GPR-CI	134	3.954289854	2.992640319	0.83804234	0.992537313	21.5595886
3	SVR	134	11.82054693	10.74994536	-0.44723692	0.686567164	25.37258378
3	KNN	134	12.3938025	10.97089942	-0.59101265	0.686567164	27.88975092
7	NGPR	126	4.694078886	3.004009566	0.895562246	0.952380952	17.77026331
7	GPR-CI	126	4.719718925	3.092082261	0.894418208	0.888888889	14.84231081
7	SVR	126	5.194697675	3.682498143	0.872098007	0.952380952	21.64695142
7	KNN	126	8.066518471	6.329659075	0.691589595	0.928571429	29.82957525
14	NGPR	62	3.601492561	2.338695673	0.823324116	0.935483871	15.21554608
14	GPR-CI	62	3.714520134	2.467831452	0.812060676	0.983870968	24.09361069
14	SVR	62	5.596681991	4.089793951	0.573348291	0.935483871	26.50410707
14	KNN	62	7.005962587	5.720680464	0.331428506	0.951612903	27.06714355
28	NGPR	425	7.379342969	5.541208422	0.747790392	0.875294118	22.55796799
28	GPR-CI	425	7.462315962	5.608726656	0.742086839	0.670588235	13.97951047
28	SVR	425	9.470698955	6.51654082	0.584577227	0.741176471	17.40244768
28	KNN	425	11.12693054	8.783810373	0.426574515	0.795294118	28.37514101
56	NGPR	91	3.221612006	2.499508566	0.948742494	0.978021978	15.85505454
56	GPR-CI	91	3.180838439	2.539028575	0.95003174	1	18.17260401
56	SVR	91	3.823191862	2.847826368	0.927812303	0.989010989	25.53891775
56	KNN	91	6.748646595	5.197067696	0.775071559	0.978021978	31.02860142
90	NGPR	54	3.153289236	2.474682148	0.894911879	1	21.59861262
90	GPR-CI	54	2.969166158	2.24123507	0.906825941	1	19.32221929
90	SVR	54	4.459627234	3.13236889	0.789804639	0.962962963	21.1863094
90	KNN	54	8.617634442	6.921105357	0.21512202	0.759259259	24.51574625
91	NGPR	22	2.737659695	2.180084417	0.867486536	1	19.66742811
91	GPR-CI	22	4.758917957	3.082166707	0.599578379	0.954545455	19.45172484
91	SVR	22	4.240409697	3.669158899	0.682080826	1	21.49752152
91	KNN	22	4.83279205	4.248715643	0.587050162	1	22.81423444
100	NGPR	52	3.611419687	2.859559809	0.811615935	1	20.46160645
100	GPR-CI	52	4.088141321	3.241494292	0.758598457	0.942307692	15.15699509
100	SVR	52	4.215264474	3.531477459	0.743351992	1	25.76873389
100	KNN	52	6.810538797	5.823809469	0.330036436	0.942307692	24.88612772
120	NGPR	3	2.205798157	1.615609408	-4.99847246	1	14.37497339
120	GPR-CI	3	2.163190842	1.568814044	-4.76897692	1	17.98186435
120	SVR	3	1.991846265	1.656231849	-3.89126014	1	19.03680728
120	KNN	3	1.766010558	1.627551376	-2.84499449	1	28.79251776
180	NGPR	26	4.975081908	3.349038748	0.784516049	0.923076923	23.50176905
180	GPR-CI	26	4.809464801	3.177578452	0.798623884	0.884615385	20.41129718
180	SVR	26	6.683170419	5.056489368	0.611152303	0.884615385	19.54922251
180	KNN	26	4.456103032	3.400025481	0.8271279	1	21.99888091
270	NGPR	13	4.757151418	3.001782079	0.783632525	1	25.25367921

Age	Model	n	RMSE	MAE	R ²	Coverage	AIL
270	GPR-CI	13	4.955724042	3.442736112	0.765192342	0.846153846	20.74053654
270	SVR	13	7.717243649	4.91143938	0.43059346	0.846153846	21.8554607
270	KNN	13	5.082671417	3.787306578	0.75300845	0.923076923	20.32563229
360	NGPR	6	2.521886956	2.06945403	0.560944722	1	27.81737742
360	GPR-CI	6	2.6864393	2.372149648	0.501778992	1	22.24125934
360	SVR	6	4.471023338	4.351320496	-0.38000876	1	23.13387127
360	KNN	6	4.283804117	4.070400486	-0.26685578	1	24.62128823
365	NGPR	14	3.696772201	3.12623899	0.840973889	1	32.25250892
365	GPR-CI	14	4.526943186	3.52761279	0.761530332	0.857142857	14.24686546
365	SVR	14	7.169589991	5.964053642	0.401847807	0.714285714	18.93156876
365	KNN	14	6.167385349	5.350879898	0.557385917	0.928571429	25.24024364

Overall, LOAO provides a fairly rigorous evaluation because the model was tested on age groups not used in training. Substantively, these results indicate that GPR-based models are more adaptable to shifts in the curing age structure, while SVR and especially KNN are more prone to accuracy losses when certain age groups are completely removed from the training process. Thus, the results in Tables 2 and 3 indicate that NGPR provides the most consistent overall performance under the stringent LOAO generalization scheme.

The next evaluation was conducted using repeated stratified K-fold by Age. Unlike LOAO, this scheme provides a more empirically stable performance because all age groups are still represented in the fold distribution. Table 4 shows that in the overall evaluation, NGPR is again the best model with RMSE = 4.2931, MAE = 2.8890, and R² = 0.9339. This result is better than GPR-CI, which has RMSE = 4.5314 and R² = 0.9264, and were considerably better than the results obtained by SVR and KNN. In terms of prediction interval, NGPR also produces the shortest interval, namely 16.2586, with coverage = 0.9437. In comparison, GPR-CI has the widest coverage, 0.9631, but it comes with a wider interval, i.e., 23.3587. SVR and KNN still have longer gaps with more prediction errors. The results showed that overall, NGPR was the most efficient, but GPR-CI was more conservative in maintaining interval coverage.

Table 4. Overall performance of the concrete compressive strength prediction model in the repeated stratified K-fold by Age scheme.

Model	RMSE	MAE	R ²	Coverage	AIL
NGPR	4.293053597	2.889005007	0.933896235	0.94368932	16.25859249
GPR-CI	4.531445577	3.066529298	0.926350956	0.963106796	23.35873683
SVR	5.272841968	3.283625019	0.900279812	0.957281553	26.58946668
KNN	6.084619556	4.164952494	0.867211509	0.949514563	27.38362591

The age-specific results in Table 5 are largely consistent with the overall performance pattern reported in Table 4. NGPR resulted in an RMSE of 3.4744 and an R² of 0.8750 at 3 days, which was somewhat better than GPR-CI. NGPR fared better again at 7 days, with an RMSE of 4.6671 and an R² of 0.8968. At 14 days, GPR-CI produced a lower RMSE of 3.3742 than NGPR, which produced an RMSE of 3.5778. The most representative findings were observed in the 28-day age group (425 observations): NGPR achieved an RMSE of 5.0611, an MAE of 3.3804, and an R² of 0.8814, the best among all models. But the coverage of this model was 0.9129, slightly below the theoretical 95% objective, with an interval length of 16.1021. SVR obtains coverage = 0.9624 and an interval length of 29.3704, whereas GPR-CI is average with coverage = 0.9412 and an interval length of 24.5457. Hence, there is a trade-off between the sharpness and the coverage of the interval at 28 days.

Table 5. Performance of concrete compressive strength prediction model according to curing age in the repeated stratified K-fold by Age scheme.

Age	Model	n	RMSE	MAE	R ²	Coverage	AIL
1	NGPR	2	5.836028421	5.790401551	-2.35670007	1	22.37556045

Age	Model	n	RMSE	MAE	R ²	Coverage	AIL
1	GPR-CI	2	6.334354751	6.213020728	-2.95441738	1	22.45900519
1	SVR	2	11.46222335	11.46039498	-11.9483811	0.5	21.42375693
1	KNN	2	17.83164518	16.69141132	-30.3372274	0.5	25.14554067
3	NGPR	134	3.474448142	2.546921972	0.874963687	0.97761194	17.00544122
3	GPR-CI	134	3.579946096	2.578511439	0.867255215	0.985074627	19.44695182
3	SVR	134	4.733539701	3.245754621	0.767920628	0.910447761	19.50417411
3	KNN	134	6.608379495	4.854100708	0.54767099	0.962686567	31.80203254
7	NGPR	126	4.667082205	2.904653906	0.896760081	0.936507937	15.6272964
7	GPR-CI	126	4.913944815	3.088551887	0.885549602	0.976190476	32.79586784
7	SVR	126	5.254090889	3.236714138	0.869156571	0.952380952	32.7412578
7	KNN	126	6.022365915	3.933360646	0.828094	0.952380952	26.85072975
14	NGPR	62	3.577792482	2.641966387	0.825641742	0.935483871	14.66353836
14	GPR-CI	62	3.374205788	2.517562768	0.844920154	0.983870968	20.73363325
14	SVR	62	3.589837346	2.477996558	0.824465789	0.983870968	23.76332447
14	KNN	62	5.515843742	4.249832857	0.58558436	0.983870968	25.43800489
28	NGPR	425	5.061110032	3.380401153	0.881363645	0.912941176	16.10205715
28	GPR-CI	425	5.311994609	3.581071408	0.869310262	0.941176471	24.54569296
28	SVR	425	6.082973928	3.617408947	0.828620786	0.962352941	29.37040572
28	KNN	425	6.395414353	4.044290243	0.810563522	0.929411765	28.11547887
56	NGPR	91	3.439117467	2.653613515	0.941587604	0.967032967	15.2008838
56	GPR-CI	91	3.898021172	2.950704614	0.924958866	0.956043956	17.79923511
56	SVR	91	4.330422081	3.271502315	0.907387105	0.945054945	23.76336901
56	KNN	91	5.460526549	4.357173724	0.85274161	0.956043956	23.70487497
90	NGPR	54	2.847551926	1.866293	0.914302271	1	17.15167246
90	GPR-CI	54	3.297501139	2.131285725	0.88507992	0.962962963	20.84762437
90	SVR	54	2.744053578	2.005837192	0.920418672	1	23.70412175
90	KNN	54	5.103265915	3.842006267	0.724753119	0.981481481	25.35984981
91	NGPR	22	3.509094668	2.376101779	0.782283546	0.954545455	18.28407647
91	GPR-CI	22	3.681874406	2.634174001	0.760316011	1	20.65284691
91	SVR	22	5.39793418	3.841180073	0.484823253	0.954545455	23.2646945
91	KNN	22	4.682392867	2.995859796	0.612352683	1	25.26992619
100	NGPR	52	3.924832743	3.040242893	0.777499726	0.961538462	15.7749814
100	GPR-CI	52	3.994584019	3.248920883	0.769521	1	20.82891111
100	SVR	52	4.179880465	3.256219067	0.747642646	0.980769231	23.55110862
100	KNN	52	6.167200713	4.912548973	0.450630674	0.942307692	25.40804999
120	NGPR	3	1.98611982	1.492976586	-3.86317638	1	13.9404428
120	GPR-CI	3	1.958097511	1.555201445	-3.72691465	1	19.63350878
120	SVR	3	1.767124435	1.289921819	-2.84984633	1	25.11116653
120	KNN	3	1.848550484	1.826097033	-3.21280892	1	25.60055682
180	NGPR	26	2.021643391	1.55180396	0.964418557	1	17.72870964
180	GPR-CI	26	3.423724816	2.142046056	0.897950199	0.961538462	20.58941814
180	SVR	26	2.89015598	2.064527797	0.927279472	1	23.49514001
180	KNN	26	4.432033845	3.342881841	0.828990359	1	25.54046688
270	NGPR	13	2.170557896	1.775340683	0.954955603	1	18.95099663
270	GPR-CI	13	2.217977519	1.804021656	0.952965958	1	20.50289229
270	SVR	13	9.840905892	4.517163429	0.074091258	0.923076923	23.97657896
270	KNN	13	5.233875221	3.450071467	0.738094417	0.923076923	25.46055185
360	NGPR	6	1.061546727	0.899888944	0.922205982	1	21.67988276
360	GPR-CI	6	1.230239436	0.949446992	0.895516609	1	20.00845415
360	SVR	6	1.725376888	1.336416609	0.794488597	1	23.02493109
360	KNN	6	3.347220707	2.946722351	0.226542115	1	25.94685528
365	NGPR	14	2.6839525	2.172706781	0.916175165	1	20.72053718
365	GPR-CI	14	2.982361465	2.383556917	0.896499221	1	21.08738729

Age	Model	n	RMSE	MAE	R ²	Coverage	AIL
365	SVR	14	3.801746221	2.948636112	0.831814208	1	23.89817623
365	KNN	14	4.48844503	3.42108774	0.765569086	1	25.19213396

At intermediate ages, the dominance of NGPR remains quite clear. At 56 days, the model yields an RMSE of 3.4391, an R² of 0.9416, and a mean interval of 15.2009. At 90 days, there is an interesting exception: the SVR yields the best results for point predictions in the previous analysis, but overall, the GPR family remains more stable. At 91 and 100 days, NGPR again yields the best or most efficient results. Meanwhile, at later ages, such as 180, 270, 360, and 365 days, NGPR also remains strong, although some small-sample groups require caution in interpretation. The 1-day, 120-day, and 360-day age groups, for example, have very limited observations, making the R² values less stable. Therefore, the main conclusions are more appropriately emphasized on age groups with sufficient data, especially 3, 7, 14, 28, 56, 90, 100, and 180 days.

B. DISCUSSION

The results obtained in the previous section reveal that predicting concrete compressive strength from this data cannot be described by a simple linear relationship. The curing age is certainly important, but it is not independent of other factors. The pattern of age-compressive strength is more likely to be a gradual, nonlinear relationship rather than a clean, rising line from start to finish, depending on the mix composition, water-cement ratio, slag content, fly ash, and interactions among materials. This has long been noted in the data-driven concrete literature, and more recent studies have shown the same thing: age, water/cement ratio, slag, and water continue to emerge as the most influential variables when more flexible and interpretable models are used.

In line with the character of the relationship, the Matérn 3/2 kernel is the best GPR kernel in this study because it is most often selected in the inner cross-validation, namely 92.857% in the LOAO scheme and 100% in the repeated stratified K-fold by Age scheme. This finding indicates that the relationship among mix composition, curing age, and concrete compressive strength is better modeled with a flexible kernel and exhibits a moderate level of smoothness.

In that context, NGPR's comparatively consistent performance across the two evaluation schemes is consistent with both its probabilistic formulation and the empirical results. As a Gaussian process model, NGPR is flexible in representing nonlinear relationships and naturally provides a predictive distribution, allowing point-prediction accuracy and prediction uncertainty to be evaluated jointly (Rasmussen & Williams, 2006; Miao et al., 2025; Xie et al., 2025). This advantage is particularly relevant for concrete compressive-strength data, in which the relationships among mixture composition, curing age, and strength are complex and nonlinear. The LOAO scheme provides a stringent generalization assessment because the model must predict an entire curing-age group that is not represented in the training data. Under this setting, NGPR achieved empirical coverage of 0.9350 with an average interval length of 21.2339, whereas GPR-CI produced a shorter average interval length of 16.8252 but substantially lower coverage of 0.8379. These results indicate that NGPR provided a better empirical balance between point-prediction performance, interval calibration, and interval efficiency in the LOAO analysis. Thus, a narrower prediction interval should not be considered superior when it is accompanied by substantial undercoverage.

The same pattern is again observed in the repeated stratified K-fold analysis by Age. However, the results appear more stable because each fold still contains representatives from all age groups. In this evaluation, NGPR again demonstrated the most efficient performance with an average interval length of 16.2586. On the other hand, GPR-CI provided higher coverage (0.9631) but wider intervals (23.3587). These results demonstrate a common trade-off in evaluating prediction intervals. High coverage is usually associated with longer intervals, while too narrow intervals tend to reduce coverage. Therefore, the best model should not be judged solely by the shortest interval or the highest coverage, but by its ability to balance calibration and interval sharpness. In these results, this balance is more often demonstrated by NGPR.

The 28-day age group is of exceptional interest because it has the largest number of observations. Also, in civil engineering practice, the 28-day age is often selected as the main reference for determining concrete compressive strength. In this group, NGPR has the shortest interval (16.1021), but its coverage is 0.9129, just below the nominal 95% limit. In contrast, SVR produces a much wider interval, at

29.3704, while GPR-CI falls in the middle with an interval length of 24.5457. These results indicate that at the 28-day age, the issue is not simply choosing the most accurate model, but rather choosing one that strikes a reasonable balance between predictive efficiency and inferential caution. From an application perspective, these findings provide a fairly clear direction. If the primary goal is to obtain efficient yet informative predictions, then NGPR is a better priority. However, if more emphasis is placed on more conservative intervals, GPR-CI remains a reasonable choice.

From an industrial perspective, the proposed prediction framework can support preliminary concrete mix screening, quality-control planning, and the prioritization of specimens requiring confirmatory laboratory testing. By providing both predicted compressive strength and the associated prediction interval, the framework enables engineers to distinguish between mixtures that are likely to satisfy the required strength, mixtures that are unlikely to meet the requirement, and mixtures whose performance remains uncertain. Laboratory resources can therefore be directed more efficiently toward mixtures with high predictive uncertainty or predicted strength close to the specified threshold. In this context, the model may reduce unnecessary preliminary and repetitive laboratory testing, shorten the initial mix-evaluation process, and support more efficient use of materials, testing equipment, and technical personnel. Nevertheless, the model is intended as a decision-support tool rather than a replacement for mandatory standardized compressive-strength testing, particularly for final mix approval, regulatory compliance, and safety-critical structural decisions. Moreover, because the present study did not include a direct cost-effectiveness analysis or prospective industrial implementation, actual reductions in laboratory testing costs have not yet been empirically demonstrated.

Another important consideration is how to interpret the results for age groups with small data sets. At 1 day, 120 days, and 360 days, for example, the limited sample size makes performance measures more susceptible to change. Under such conditions, a model that appears superior in one age group may not necessarily be superior in general. Therefore, the main conclusions are more appropriate for age groups supported by more adequate data sets, specifically 3, 7, 14, 28, 56, 90, 100, and 180 days. Based on this, the results of this study generally indicate that the GPR family, particularly NGPR, is better able to adapt to the heterogeneity of concrete compressive strength data across various curing ages than SVR and KNN. This finding aligns with recent literature emphasizing that concrete compressive strength modeling should combine nonlinear flexibility, rigorous evaluation, and well-explained interpretation.

4. CONCLUSION

The analysis results show that the compressive strength of concrete at various curing ages exhibits a non-linear, inhomogeneous pattern. The response distribution tends to be right-skewed; variations across age groups differ, and the relationship between curing age and compressive strength is better understood as a gradual pattern rather than a single linear relationship. The condition emphasizes that predicting concrete compressive strength requires a flexible model and an evaluation design that explicitly considers the curing age structure.

The GPR family showed more consistent overall performance than SVR and KNN across both evaluation schemes. Under LOAO, NGPR achieved the lowest RMSE (5.6259) and MAE (3.9271), the highest R^2 (0.8865), and empirical coverage of 0.9350 with an AIL of 21.2339. In comparison, GPR-CI produced a slightly higher RMSE (5.7086) and a shorter AIL (16.8252), but its coverage decreased to 0.8379, while SVR and KNN yielded substantially higher RMSE values of 8.1760 and 9.7415, respectively. Under repeated age-stratified K-fold cross-validation, NGPR again achieved the lowest RMSE (4.2931) and MAE (2.8890), the highest R^2 (0.9339), and the shortest AIL (16.2586), while maintaining coverage of 0.9437. GPR-CI achieved slightly higher coverage (0.9631), but with a wider AIL (23.3587). These findings indicate that NGPR provides the most favorable overall balance among point-prediction accuracy, interval calibration, and interval sharpness, whereas GPR-CI provides more conservative intervals under repeated age-stratified validation.

Further work could explore the use of external data, more balanced distributions of curing age, and comparisons with other uncertainty techniques, such as more adaptive conformal models, bootstrapping, or richer Bayesian models. Extension to more particular concrete kinds or to shortfall risk analysis is also crucial to provide broader relevance of the results to engineering contexts.

ACKNOWLEDGEMENTS

This research was funded by the PNPB Fund of Universitas Negeri Medan through the 2023 Applied Research Scheme, Contract Number: 0038/UN33.8/PPKM/PPT/2023.

REFERENCES

- Almutairi, A. (2025). Explainable machine learning-based prediction of compressive strength in sustainable recycled aggregate self-compacting concrete using SHAP analysis. *Sustainability*, 17(24), 11334. <https://doi.org/10.3390/su172411334>
- Angiulli, G., Calcagno, S., La Foresta, F., & Versaci, M. (2024). Concrete compressive strength prediction using combined non-destructive methods: A calibration procedure using preexisting conversion models based on Gaussian process regression. *Journal of Composites Science*, 8(8), 300. <https://doi.org/10.3390/jcs8080300>
- Beskovpylny, A. N., Stel'makh, S. A., Shcherban', E. M., Mailyan, L. R., Meskhi, B., Razveeva, I., Chernil'nik, A., & Beskovpylny, N. (2022). Concrete strength prediction using machine learning methods CatBoost, k-nearest neighbors, support vector regression. *Applied Sciences*, 12(21), Article 10864. <https://doi.org/10.3390/app122110864>
- Dong, S., & Zhang, Z. (2025). Hybrid deep learning with conformal prediction for recycled aggregate self-compacting concrete strength prediction. *Buildings*, 15(24), 4419. <https://doi.org/10.3390/buildings15244419>
- Elhishi, S., Elashry, A. M., & El-Metwally, S. (2023). Unboxing machine learning models for concrete strength prediction using XAI. *Scientific Reports*, 13, 19892. <https://doi.org/10.1038/s41598-023-47169-7>
- Feng, D.-C., Liu, Z.-T., Wang, X.-D., Chen, Y., Chang, J.-Q., Wei, D.-F., & Jiang, Z.-M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000. <https://doi.org/10.1016/j.conbuildmat.2019.117000>
- Fu, H., Zhou, X., Xu, P., & Sun, D. (2025). Prediction of compressive strength of concrete using explainable machine learning models. *Materials*, 18(21), 5009. <https://doi.org/10.3390/ma18215009>
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2), 243–268. <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- Kaloop, M. R., Kumar, D., Samui, P., Hu, J. W., & Kim, D. (2020). Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Construction and Building Materials*, 264, 120198. <https://doi.org/10.1016/j.conbuildmat.2020.120198>
- Liu, G., & Sun, B. (2023). Concrete compressive strength prediction using an explainable boosting machine model. *Case Studies in Construction Materials*, 18, e01845. <https://doi.org/10.1016/j.cscm.2023.e01845>
- Miao, Q., Gao, Z., Zhu, K., Guo, Z., Sun, Q., & Zhou, L. (2025). Interpretable machine learning model for compressive strength prediction of self-compacting concrete with recycled concrete aggregates and SCMs. *Journal of Building Engineering*, 108, 112965. <https://doi.org/10.1016/j.jobbe.2025.112965>
- Mobasheri, F., Hosseinpour, M., Yahia, A., & Pourkamali-Anaraki, F. (2025). Predicting the compressive strength of self-consolidating concrete using machine learning and conformal inference. *Computer Modeling in Engineering & Sciences*, 145(3), 3309–3347. <https://doi.org/10.32604/cmes.2025.072271>
- Olvera-Mayorga, C. E., López-Martínez, M. d. J., Rodríguez-Rodríguez, J. A., Vázquez-Reyes, S., Solís-Sánchez, L. O., de la Rosa-Vargas, J. I., Duarte-Correa, D., González-Aviña, J. V., & Olvera-Olvera, C. A. (2025). AI-based inference system for concrete compressive strength: Multi-dataset analysis of optimized machine learning algorithms. *Applied Sciences*, 15(23), Article 12383. <https://doi.org/10.3390/app152312383>
- Rana, S., Midi, H., & Imon, A. H. M. R. (2012). Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presence of outliers. *Mathematical Problems in Engineering*, 2012, Article 730328. <https://doi.org/10.1155/2012/730328>
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press. <https://doi.org/10.7551/mitpress/3206.001.0001>
- Shaaban, M., Amin, M., Selim, S., & Riad, I. M. (2025). Machine learning approaches for forecasting compressive strength of high-strength concrete. *Scientific Reports*, 15, 25567. <https://doi.org/10.1038/s41598-025-10342-1>
- Simamora, E., Mansyur, A., Firdaus, M., & Habibi, R. (2025). Residual bootstrap analysis on multiple linear regression: Case study of concrete compressive strength. *AIP Conference Proceedings*, 3316(1), 040003. <https://doi.org/10.1063/5.0290027>
- Tamuly, P., & Nava, V. (2025). Machine learning based conformal predictors for uncertainty-aware compressive strength estimation of concrete. *Construction and Building Materials*, 487, 141844. <https://doi.org/10.1016/j.conbuildmat.2025.141844>

- Xiao, K., Zhang, H., Wei, S., Zhu, C., He, J., Zhu, S., & Yang, X. (2025). Prediction of compressive strength of high-performance concrete based on multiple machine learning models. *Frontiers in Materials*, 12, 1698248. <https://doi.org/10.3389/fmats.2025.1698248>
- Xie, L., He, J., Wang, C., Zhu, S., Zhang, H., & Yang, X. (2025). Enhanced Bayesian Gaussian process regression for compressive strength prediction of multi-binder concrete. *Journal of Building Engineering*, 111, Article 113308. <https://doi.org/10.1016/j.jobbe.2025.113308>
- Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12), 1797–1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3)
- Yeh, I.-C. (2006). Analysis of strength of concrete using design of experiments and neural networks. *Journal of Materials in Civil Engineering*, 18(4), 597–604. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2006\)18:4\(597\)](https://doi.org/10.1061/(ASCE)0899-1561(2006)18:4(597))
- Zhang, Y., Ren, W., Chen, Y., Mi, Y., Lei, J., & Sun, L. (2024). Predicting the compressive strength of high-performance concrete using an interpretable machine learning model. *Scientific Reports*, 14, 28346. <https://doi.org/10.1038/s41598-024-79502-z>

The Effectiveness of PjBL Learning on Interest and Mathematical Creative Thinking in Determining the Number of Construction Workers for 11th Grade Students in the Construction Department

Mozard Aryan Hukunala¹, Rizky Fatmalasari L², Agung K. Henaulu³, Tri Siwi Nasrulyati⁴

¹Department of Mathematic Education, Universitas Darussalam Ambon, Ambon, Indonesia

²Department of Physics Education, Universitas Darussalam Ambon, Ambon, Indonesia

³Department of Industrial Engineering, Universitas Darussalam Ambon, Ambon, Indonesia

⁴Department of Mechanical Engineering, Universitas Darussalam Ambon, Ambon, Indonesia

mozard_hukunala@yahoo.com, rizkyfatmalasari@gmail.com, agung@unidar.ac.id, trisiwi@unidar.ac.id

ABSTRACT

This study examined the effectiveness of the Project-Based Learning (PjBL) model in improving students' learning interest and mathematical creative thinking skills in spatial geometry at SMK Negeri 4 Ambon. It also investigated the implementation of project-based learning in the Building Engineering Program through a project to determine the number of construction workers based on the Indonesian Standard Work Unit Price Analysis (AHSP-SNI). This quasi-experimental study involved 69 eleventh-grade students. Data were collected through observations, learning interest questionnaires, creative thinking tests, project portfolios, assessment rubrics, and project presentations, and analyzed using descriptive statistics. The number of construction workers was calculated using the AHSP-SNI formula: $(8 \text{ working hours} \times \text{building area}) \div \text{desired project completion time}$. The results showed that students' average learning interest reached 97.10%, while mathematical creative thinking reached 87.00%. Students' ability to identify building components was categorized as good, whereas their ability to determine the required number of construction workers was good for a one-month duration and very good for two- and three-month durations. The findings indicate that PjBL effectively improves learning interest, creative thinking, and the application of mathematics in vocational construction education.

Keywords: Project-Based Learning (PjBL), learning interest, mathematical creative thinking, spatial geometry.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Mozard Aryan Hukunala,
Department of Mathematic Education,
Universitas Darussalam Ambon,
Jalan Waehakila Puncak Wara Ambon, 97128, Indonesia.
mozard_hukunala@yahoo.com

1. INTRODUCTION

Mathematics is a fundamental subject that plays a vital role in various areas of life, including vocational work, particularly in building construction. However, many students in vocational high schools (SMK) still view mathematics as an abstract and difficult subject that is irrelevant to their practical needs (Rahim et al., 2023). This results in low interest in learning and a lack of creative thinking skills in solving real-world problems related to their fields of expertise (Oktaviastuti et al., 2021a); (Qurohman & Sungkar, 2018a).

The process of learning mathematics in vocational high schools (SMK) has characteristics that are different from learning mathematics in general high schools (SMU). The difference lies in the implementation of the teaching material, which must be integrated with productive subjects (Qurohman & Sungkar, 2018b). One math topic that is useful for students is solid geometry. This material provides an understanding of the elements of solid figures, the practice of measuring length, width, and angles, calculating the area and volume of structures, as well as applying it to determine the number of construction workers based on a specified time (Oktaviastuti et al., 2021b)(Murad et al., 2024).

SMK Negeri 4 Ambon, especially in the building department, mathematics should serve as an important foundation for students to support vocational skills. One of the basic competencies taught is calculating the number of construction workers needed to complete a construction project ("Implementasi Model Pembelajaran Berbasis Proyek Dalam Penanaman Konsep Matematika Pada Siswa Sekolah Menengah," n.d.). This calculation not only requires an understanding of mathematical concepts but also creative thinking skills, such as generating various alternative solutions, using different strategies, and detailing answers thoroughly. In construction practice, a worker or technician

is often faced with situations that cannot always be resolved through routine methods. For example, when determining the number of construction workers required, it is not enough to simply use a basic formula; students also need to consider real factors such as labor productivity, project deadlines, types of work, and the efficiency of task distribution.

Unfortunately, spatial geometry learning in grade 11 of the building department at SMK Negeri 4 Ambon is still quite conventional, where teachers only provide examples of problems and students only imitate the steps to solve them. This condition is in line with the findings of (Rifai et al., 2025) who stated that mathematics learning in vocational schools is still teacher-centered and dominated by structured practice problems so that students' opportunities to build conceptual understanding and explore problem-solving strategies are still limited. This pattern makes students accustomed to single, passive answers and rarely connects spatial geometry concepts with real-world problems, resulting in underdeveloped creative thinking skills, even though these skills are very much needed in the construction workforce (Sabaruddin et al., 2018).

One alternative to addressing the challenges of mathematics learning in vocational schools is the implementation of the Project-Based Learning (PjBL) model. This model emphasizes active student involvement in completing authentic and contextual projects, encouraging students to build knowledge through hands-on learning experiences. In the Building Engineering Expertise Program, PjBL provides students with the opportunity to connect spatial geometry concepts with real-world problems in the construction sector. Research on the application of PjBL to spatial geometry materials has been conducted by (Sundari et al., 2024) which shows that PjBL can improve creative thinking skills through entrepreneurship-based projects. However, this research was conducted on university students and did not integrate mathematics learning with the context of vocational education, specifically the Building Engineering Expertise Program in vocational schools, or the application of the Work Unit Price Analysis (AHSP-SNI) standard in determining construction labor requirements. Therefore, this study offers a novelty in the form of the application of PjBL to spatial geometry learning integrated with authentic AHSP-SNI-based projects, so that students not only understand spatial geometry concepts but are also able to apply mathematical concepts to determine the number of construction workers according to project needs. The application of this model is expected to increase interest in learning mathematics while developing students' creative thinking skills (Laili et al., 2025); (Giro et al., 2024)

Based on the above explanation, this study aims to determine the effectiveness of the Project Based Learning (PjBL) model on interest and creative thinking ability in determining the number of construction workers among 11th-grade students in the Building Department at SMK Negeri 4 Ambon.

2. RESEARCH METHOD

A. Research Type and Design

This study uses a quantitative approach with a quasi-experimental type (quasi-experimental research) using a Pretest-Posttest Control Group Design (Sugiyono, 2022). This design was chosen to determine the effectiveness of the implementation of the Project-Based Learning (PjBL) model on students' learning interest and mathematical creative thinking abilities compared to conventional learning. The study was conducted in class XI of the Building Engineering Expertise Program of SMK Negeri 4 Ambon in May 2025. The research sample consisted of 69 students consisting of 34 students in class XI.A as the experimental class and 35 students in class XI.B as the control class. The independent variable of the study is the Project-Based Learning (PjBL) model, while the dependent variables include students' learning interest and mathematical creative thinking abilities.

B. Research Stages

1. Preparation Stage

At this stage, the following activities were carried out:

- a. Conducting a literature study on Project-Based Learning (PjBL), learning interest, mathematical creative thinking skills, spatial geometry, and Work Unit Price Analysis (AHSP-SNI).
- b. Conducting initial observations and interviews with mathematics teachers to identify learning problems.
- c. Determining the experimental and control classes.
- d. Developing learning materials, including:
 - Teaching Modules;

- Project-based Student Worksheets;
 - Pretest and posttest instruments;
 - Learning interest questionnaire;
 - Project assessment rubric;
 - Student activity observation sheets.
- e. Validating the instruments by experts.
- f. Conducting a pilot test of the instruments to determine validity and reliability.

2. Implementation Stage

The implementation stage is carried out over several meetings.

a. Experimental Class (Project-Based Learning)

Learning activities follow the PjBL syntax according to the Ministry of Education and Culture, namely:

1. Determine the fundamental question (Start with the Essential Question).
The teacher presents a real-life problem in the form of planning a simple house construction, asking students to determine labor requirements based on the building area.
2. Students are divided into groups and develop steps to complete the project.
3. Develop an activity schedule. Each group develops a project schedule.
4. Monitor project implementation.
The teacher guides the discussion process, calculates spatial geometry, identifies building components, and calculates the number of workers using AHSP-SNI.
5. Test the project results. Each group presents their project results.
6. Evaluate the learning experience.

The teacher and students reflect on the learning process and outcomes.

b. Control Class

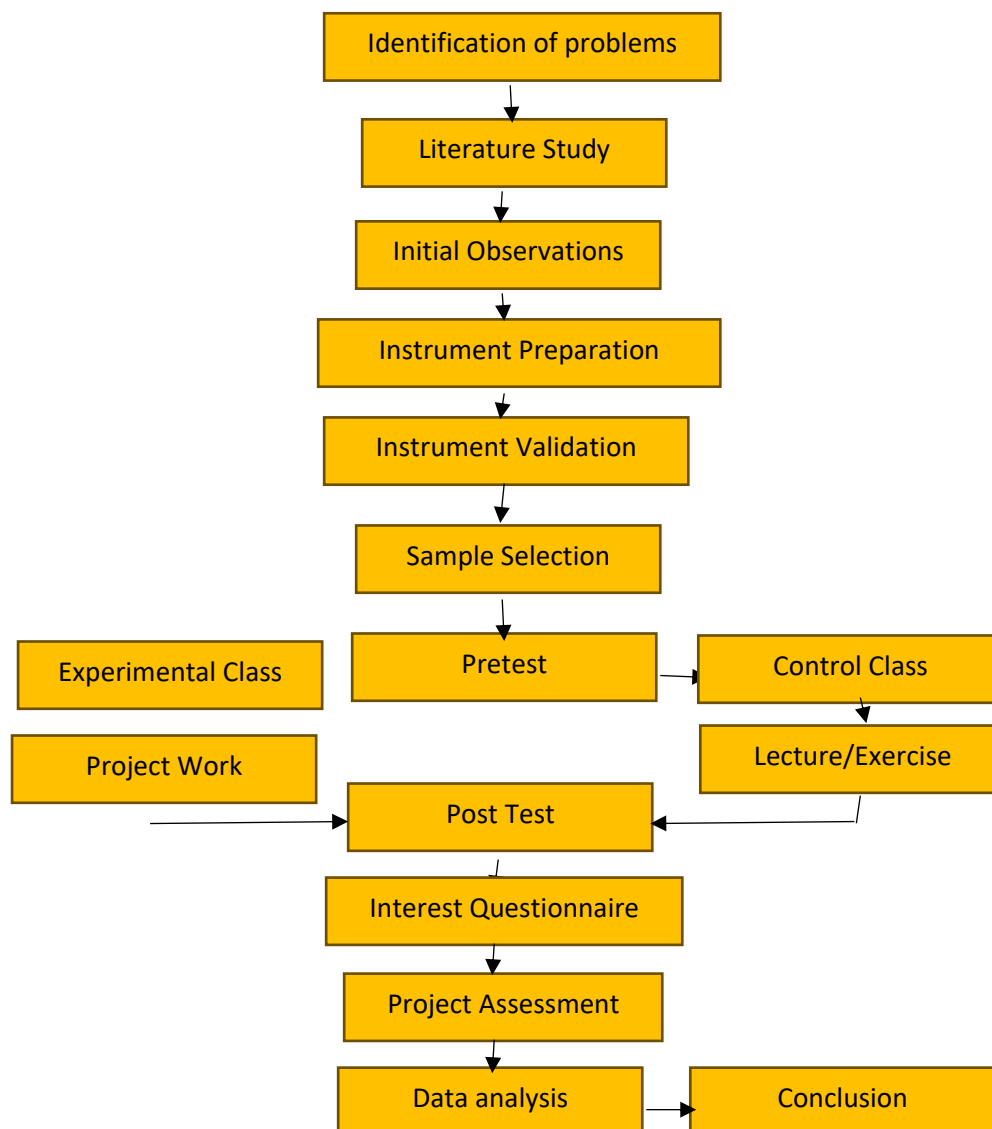
In the control class, learning was conducted using conventional methods such as lectures, questions and answers, example problems, and individual practice without project activities.

3. Evaluation Stage

Evaluation was conducted after the entire learning process was completed. The evaluation included:

- administering a post-test on mathematical creative thinking skills;
- completing a learning interest questionnaire;
- assessing project outcomes using a rubric;
- observing student activities;
- documenting learning outcomes.

All data was then analyzed to determine the effectiveness of the Project-Based Learning implementation. The following is the research flow carried out by the researcher:



(Sugiyono, 2022)

C. Data Collection Techniques

Data collection techniques include:

1. Observation. Conducted to determine teacher and student activities during learning.
2. Questionnaires. Used to measure student learning interest.
3. Tests. Open-ended essay questions to measure mathematical creative thinking skills.
4. Project Assessment.

Used to assess students' ability to determine building components and the number of construction workers based on AHSP-SNI

D. Data Analysis Techniques

Data analysis was conducted through the following stages:

1. Descriptive analysis of learning interest, creative thinking skills, and project outcomes.
2. Instrument validity and reliability testing.
3. Data normality testing and homogeneity of variance testing
(Henaulu & Ardian, 2020; Wulansari et al., 2025).
4. Hypothesis testing using an Independent Sample t-test to compare learning outcomes in the experimental and control classes.
5. Project outcome analysis using the calculation of the number of workers based on AHSP-SNI with the formula:

Number of Workers = $(8 \text{ Working Hours} \times \text{Building Area}) \div \text{Implementation Time}$

6. Drawing conclusions based on the statistical analysis results

3. RESULTS AND DISCUSSION

A. Research Result

Initial Data Analysis

This study was conducted in the eleventh grade at SMK Negeri 4 Ambon by implementing the PjBL Learning Model. In order to obtain initial data regarding students' learning interest in the eleventh grade at SMK Negeri 4 Ambon, a learning interest questionnaire was distributed with the following indicators: (1) Enjoyment in learning mathematics, (2) Student involvement in learning mathematics, (3) Interest in the material presented, and (4) Students' attention to the material being taught. The results were analyzed and summarized for two sample classes, XIA and XIB, and can be described as follows

Table 1. Recapitulation of Interest in Learning Mathematics Based on All Indicators

Number	Interval	Category	Class A	Class B	Percentage
1	$X > 68$	Very Interested	1	3	5,8
2	$56 < x \leq 68$	Interested	3	3	8,7
3	$44 < x \leq 56$	Quite Interested	20	17	53,6
4	$32 < x \leq 44$	Less Interested	4	7	15,9
5	$x \leq 32$	Not Interested	6	5	15,9
6					
	Amount		34	35	100

Table 1 shows that the average interest in learning mathematics in class XI of SMK Negeri 4 Ambon is 53.6 in the category of Quite Interested in learning mathematics, this shows that most students have interest but it is not strong yet, To prove the difference in the average interest in learning class XIA and class XIB, a t test was carried out with $\alpha = 5\%$, the test results showed that $t = 0.3364 < t_{\text{tab}} = 2.1318$ meaning that H_0 is accepted and H_a is rejected so it can be concluded that class XIA and class XIB have a relatively balanced distribution even though class XIB has slightly more students in the very interested category than class XIA, but class XIB has more students in the less interested category. To analyze the data on creative thinking in mathematics of class XI students majoring in Building at SMK Negeri 4 Ambon, initial observation data was used on the learning process carried out by the teacher, based on the results of the observation, students' creative thinking was found as follows:

Table 2. Summary of Students' Creative Thinking Based on Indicators

Creative Thinking Ability Level	Class XI.A	Class XI.B	Percentage
Creative	9	7	23,29
Lack of creativity	19	23	60,87
Not Creative	6	5	15,94
Amount	34	35	100

From table 2 above, it shows that the average creative thinking of students at SMK Negeri 4 Ambon is 23.29 consisting of the average creative thinking of class XIA (experimental) students of 26.47 with a class percentage of 26.5%. and class XIB (control) of 0.2 with a class percentage of 20%. F-test two sample for variance was conducted to test the homogeneity of the two classes above. The results of the analysis show that $F_{\text{hit}} = 1.4477797 < F_{\text{tab}} = 1.7720665$ meaning that both are homogeneous

Final Stage Data Analysis

The analysis of learning interest data using the same indicators after the PBjL learning process on spatial geometry material can be described as follows:

Table 3. Summary of Mathematics Learning Interest Based on All Indicators

Number	Interval	Category	Class A	Class B	Percentage
1	$X > 68$	Very Interested	12	17	42,03
	$56 < x \leq 68$	Interested	20	14	49,28

$44 < x \leq 56$	Quite Interested	1	3	5,80
$32 < x \leq 44$	Less Interested	1	1	2,89
$x \leq 32$	Not Interested	0	0	0
Amount		34	35	100

Table 3 above shows that the average interest in learning mathematics of class XI students of SMK Negeri 4 Ambon is 97.10, consisting of an average of class XIA students (experimental class) of 0.97 with a percentage of 97.06%, and an average interest in learning mathematics of class XIB (control class) of 0.94 or 94.29%, while to find out students' creative thinking can be assessed from the results of the product in the form of portfolio assessments and presentations of student work results. Based on the results of data analysis in the form of portfolios and percentages of student work results can be described as follows.

Table 4. Summary of Students' Creative Thinking Based on Indicators

Creative Thinking Ability Level	Class A	Class B	Percentage
Creative	31	29	82,86
Lack of creativity	2	4	12,86
Not Creative	1	2	4,29
Amount	34	35	

Table 4 above shows that the average creative thinking of class XI students at SMK Negeri 4 Ambon is 0.87 or 87% consisting of the average creative thinking for class XIA (experimental class) of 91.18% and class XIB (control class) of 82.86%. The following will explain the comparison of learning interest and creative thinking data at the initial and final stages as shown in the graph below

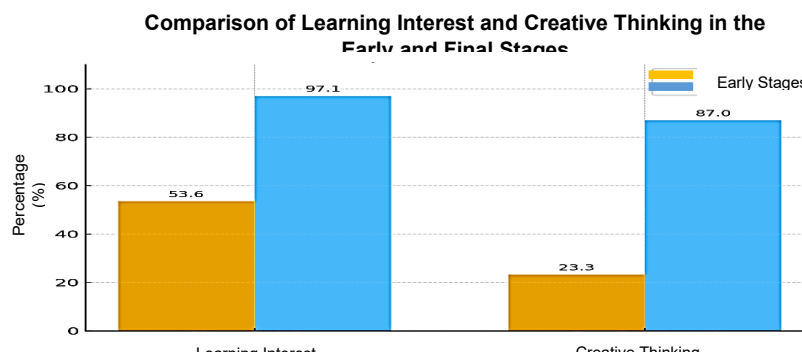


Fig 1. Comparison of Learning Interest and Creative Thinking in the Early and Final Stages

Based on the figure 1 above, both learning interest and creative thinking experienced a very significant increase from the initial to the final stage. This increase indicates that the learning approach or method used was effective in improving students' creative thinking skills.

a. Data from the Results of Analysis of the Implementation of Learning Spatial Structures in the Productive Sector

Analysis Data Determines Building Components

In order to obtain data related to the implementation of the mathematics learning process (space construction) in productive majors (building majors), the researcher provided a project in the form of a worksheet (LKS) to be worked on both independently and in groups, then analyzed the results of the students' work. Based on the results of the analysis of student work (Andyny, 2022), the overall picture can be described as follows:

Table 5 Questionnaire Results Summary

No	Sub Indicators	Average	Criteria
1	Determining building components	78,3	Good
2	Designing Building Component Costs	80,2	Good
3	Prepare a building work schedule	83,5	Good
4	Finalizing Building Component Costs	81,7	Good
5	Compile reports and presentations of estimated building component costs	85,4	Good
6	Evaluate the process and results of each building component cost	80,1	Good

Table 5 shows that each indicator in the building components can be determined or completed well by both class XIA and class XIB students.

i. Data Analysis Results Determine the Number of Construction Workers

To obtain data on the ideal number of construction workers in a construction project during the desired time, the researcher conducted tests on class XIA and class XIB students, the results of the students' work were analyzed and presented in the following table;

Table 6. Test Results Summary Determining the Number of Construction Workers for a Building Area of 9 meters x 12 meters

No	Month	Number of Workers	Class XI.A	Class XI.B	Criteria
1	I	33	80,7	75	Good
2	II	17	90	87.7	Very Good
3	III	11	95,8	89.5	Very Good
	Amount		34	35	

Table 6 shows that students of class XIA and class XIB in solving problems of spatial structures into the productive field in this case determining the number of construction workers, for the number of workers in completing a building in the first month of class XIA as many as 80.7% and class XIB as many as 75% of students can determine the number of construction workers in completing the project well. Meanwhile, in the second and third months both class XIA students (90% and 95.83%) and class XIB students (87.7% and 89.5%) can determine the number of construction workers very well.

B. Discussion

The learning process carried out by the researcher in two classes using different learning models, this is intended to determine the effectiveness of the PjBL learning model (Christian, 2021). The PjBL learning carried out by the researcher refers to the rules including creating basic questions that provide motivation for students, designing project plans in the form of LKS, compiling a project activity schedule including deadlines for each project stage, monitoring or monitoring the progress of project work, testing project results through group presentation results, evaluating project activities (according to the RPP). The PjBL learning carried out by the researcher is intended to provide a learning experience in spatial construction by confronting students of SMK Negeri 4 Ambon with building construction (productive fields) and determining the number of construction workers in real life for students who are solved in groups (Muljani & Bunawar, 2022). The project based learning model is an innovative learning model, which emphasizes contextual learning through complex activities. Through project-based learning, students are required to be able to choose topics and presentation/product projects, produce final products, and solve problems related to the real world, as well as involving various disciplines (Farihatun & Rusdarti, 2019).

Mathematics learning (spatial construction) for students majoring in construction or productive fields using the project-based PjBL learning model encourages and increases interest and creative thinking in identifying and solving mathematical problems in productive fields. Creative thinking skills expressed in identifying and solving problems are the hope for mathematics teachers in vocational schools. This thinking ability will lead students to be able to compete in the academic field and in real life in society (Analia & Bharata, 2022). The learning of spatial construction, carried out by researchers

using the PjBL learning model, is in principle divided into three stages, namely the pre-research stage, the research stage, and the post-research stage. In the pre-research stage, researchers conducted initial discussions with mathematics subject teachers and building construction teachers, this was done by researchers to collect initial information related to students' knowledge of mathematics and building construction that they had studied (Galuh & Riana, 2022) (Firdaus, 2023). At this stage, researchers also distributed questionnaires on learning interests and creative thinking to students in grades XIA and XIB and determined class XIA as the Experimental class and class XIB as the Control class (Musa'ad et al., 2024). For the second stage, the researcher carried out the learning process accompanied by the subject teacher as an observer. Meanwhile, in the third stage, the researcher distributed learning interest questionnaires and creative thinking questionnaires to the students (Slameto, 2015).

Based on the results of the analysis of the distribution of questionnaires on interest in learning mathematics to class XI students before being given treatment, it shows that the average interest in learning is 53.6% with a fairly interested category, likewise the ability to think creatively is still low, only reaching 23.3%. towards the mathematics learning process carried out by mathematics teachers. The mathematics learning process only focuses on the application of mathematical theory (space shapes) not at the level of application in the productive fields they are engaged in. so that students' thinking abilities are only at the level of mathematical understanding (Musa'ad et al., 2024) (Slameto, 2015). To prove the difference in the average interest in learning between class XIA and class XIB, a t-test was carried out with $\alpha = 5\%$, the test results showed that $t < t_{\text{tab}} = t_{0.3364} < t_{2.1318}$ meaning that accepting H_0 and rejecting H_a so it can be concluded that there is a difference in the average interest in learning between class XIA students and class XIB students (Slameto, 2015).

The results of this initial analysis serve as a reflection for researchers to plan a meaningful learning process oriented towards the productive field (construction department). The reflection results were discussed with subject teachers and colleagues (lecturers) to determine the model to be used in subsequent learning processes. Based on the discussion, the model to be used for subsequent learning was determined, namely learning using the PjBL learning model (Farihatun & Rusdarti, 2019).

The Project Based Learning learning process at SMK Negeri 4 Ambon is a student-centered learning approach using projects as a learning medium. Mathematics learning using Project Based Learning both in class XIA (experimental class) and class XIB (control class) with the same material, namely geometric shapes, the difference is that in class XIA (experimental class) students are given in-depth material on geometric shapes and are given projects in the form of calculating the area of a building, the required building components and the required (number) of workers for one month, two months, and three months. In the experimental class, the researcher explains the geometric shapes material accompanied by several examples of geometric shapes in the productive field, then divides them into small groups and is given a project in the form of a worksheet to work on (Sundari et al., 2024). The project (LKS) provided contains a building area as shown in the following image:



Fig 2. Rectangular building measuring 9 x 12 meters

In the Figure 2 above, it appears that the size of the house is rectangular with a size of 9 meters x 12 meters and a building height of 3.5 meters with 4 bedrooms, 2 bathrooms, living room, family room, kitchen, work room, car garage, terrace, garden space. Students are asked to calculate the number of house construction workers and building components needed with a specified time of 1 month, 2 months and 3 months while in class XIB, only focuses on understanding the material of spatial construction. To encourage students in the learning process, the researcher begins with a rhyme "Like a book my heart is open, understanding mathematics that opens the way, that's how we should be proud, fight and move forward to build a future home, then students are formed into small groups and given trigger questions to open students' thinking insights. Students are given projects according to productive fields, analyzing, working in teams, solving problems together, solving problems and presenting the results of their work.

Learning spatial structures by applying the Project Based Learning model shows students' active interest in learning the material being taught. From the results of distributing learning interest questionnaires after the spatial learning process was carried out by applying the Project Based Learning model, the average learning interest of class XI students of SMK Negeri 4 Ambon was 97.10%, consisting of an average of class XIA students (experimental class) of 0.97 with a percentage of 97.06%, and an average interest in learning mathematics for class XIB (control class) of 0.94 or 94.29%, students' thinking ability in analyzing and solving problems given both in the form of mathematical problems and productive field projects showed that the average student in creative thinking to analyze and solve spatial problems and projects was 0.87 or 87% consisting of an average of creative thinking for class XIA (experimental class) of 91.18% and class XIB (control class) of 82.86%.. while to determine building components, building component costs for both class XIA students (experimental class) and class XIB (control class) were in the good category, as well as determining the number of construction workers for both class XIA and Class XIB for 1 month is in the Good category, while for 2 months and 3 months both class XIA and class XIB are in the Very Good category. These results indicate that the PjBL learning process carried out by the researcher in class XIA (the experimental class) is more effective when compared to class XIB (the control class).

4. CONCLUSION

Based on the research results and discussion above, it can be concluded that the effectiveness of PjBL learning in class XI of the building department of SMK Negeri 4 Ambon is more effective when compared with other conventional learning. This is shown

1. The average learning interest of 11th-grade students was 97.10%, consisting of an average of 0.97 for 11th-grade students (the experimental class) or 97.06%, and an average of 0.94 for 11th-grade students (the control class).
2. Students' creative thinking skills in analyzing and solving problems, both in the form of math problems and productive field projects, showed that the average creative thinking ability for

analyzing and solving spatial problems and projects was 0.87, or 87%. This was comprised of an average of 91.18% for 11th-grade students (the experimental class) and 82.86% for 11th-grade students (the control class).

3. The ability to determine building components and building component costs for 11th-grade students (the experimental class) and 11th-grade students (the control class) was categorized as good.
4. The ability to determine the number of construction workers for both class XIA and class XIB is categorized as Good for the one-month period, while for the two-month and three-month periods, both classes fall into the Very Good category

REFERENCES

- Analia, A., & Bharata, H. (2022). Analisis Kemampuan Berpikir Kreatif Siswa SMK Pada Pembelajaran Jarak Jauh di Masa Pandemi Covid-19. *DELTA Jurnal Ilmiah Pendidikan Matematika*, 10(1), 1–12.
- Andyny, M. (2022). The Development of Ict-Based Interactive Student Worksheets (LKPD) Assisted By Software Construct 2 for Mts Students. *IJEMS: Indonesian Journal of Education and Mathematical Science*, 3(2), 2715–2985.
- Christian, Y. A. (2021). Meta Analisis Model Pembelajaran Project Based Learning terhadap Kreativitas dan Hasil Belajar Siswa di Sekolah Dasar. *Edukatif: Jurnal Ilmu Pendidikan*, 3(4), 2271–2278. <https://doi.org/10.31004/edukatif.v3i4.1207>
- Farihatun, S. M., & Rusdarti, R. (2019). Keefektifan Pembelajaran Project Based Learning (PjBL) Terhadap Peningkatan Kreativitas dan Hasil Belajar. *Economic Education Analysis Journal*, 8(2), 635–651. <https://doi.org/10.15294/eeaj.v8i2.31499>
- Firdaus, M. D. (2023). *Pengaruh Penerapan Pendekatan Science, Technology, Engineering, and Mathematics (STEM) terhadap Keterampilan Proses Sains Siswa Sekolah Dasar*. Universitas Pendidikan Indonesia.
- Galuh, B. P., & Riana, R. (2022). *Pengaruh vidio pembelajaran berbasis animasi terhadap minat dan hasil belajar pada anak di PAUD An-Nisa Kec*. E-Jurnal Pendidikan Mutiara.
- Giro, A., Hanifah, H., & Haji, S. (2024). Efektivitas model pembelajaran Project Based Learning terhadap kemampuan literasi matematis. *Jurnal Equation*, 7(1), 15–31.
- Implementasi Model Pembelajaran Berbasis Proyek Dalam Penanaman Konsep Matematika pada Siswa Sekolah Menengah. (n.d.). *Innovative: Journal Of Social Science Research*, 3(3), 8547–8557. Retrieved <https://j-innovative.org/index.php/Innovative/article/view/3155>
- Laili, D. N., Fithriyah, M., & Hadiyani, V. P. (2025). The Effect of the PJBL Learning Model on the Creativity of Fourth-Grade Students in SBDP Subjects at Madrasah Ibtidaiyah. *Pedagogik Journal of Islamic Elementary School*, 8(1), 157–168.
- Muljani, S., & Bunawar, B. (2022). Rancangan Pembelajaran Berkarakteristik Pembelajaran Inovatif Abad 21 Pada Materi Estimasi Biaya Konstruksi Dengan Model Pembelajaran Project Based Learning Di Smk Negeri 1 Adiwerna. *Cakrawala: Jurnal Pendidikan*, 9300, 182–197. <https://doi.org/10.24905/cakrawala.vi0.226>
- Murad, M., Sopamena, P., & Lessy, D. (2024). Mengenal Bentuk Bangun Ruang pada Sumur / Parigi Pusaka di Desa Lonthoir Banda Neira Maluku Tengah Understanding The Shapes of Three-Dimensional Structures in The Sacred Wells at a Lonthoir Village Banda Neira Central Maluku. *MATH LOCUS: Jurnal Riset Dan Inovasi Pendidikan Matematika*, 5(2), 128–136.
- Musa'ad, F., Ahmad, R. E., Sundari, S., & Hidayani, H. (2024). Pembelajaran berbasis proyek untuk meningkatkan kemampuan berpikir kritis siswa. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 8(2), 1481–1487. <https://doi.org/10.31004/cendekia.v8i2.3361>
- Oktaviastuti, B., Nurmallasari, R., & Damayanti, F. (2021a). Urgensi Technical Skill Bagi Tenaga Kerja Konstruksi Dalam Era Industri 4.0. *Rekayasa: Jurnal Teknik Sipil*, 5(2), 7. <https://doi.org/10.53712/rjrs.v5i2.1021>
- Oktaviastuti, B., Nurmallasari, R., & Damayanti, F. (2021b). Urgensi Technical Skill Bagi Tenaga Kerja Konstruksi Dalam Era Industri 4.0. *Rekayasa: Jurnal Teknik Sipil*, 5(2), 7. <https://doi.org/10.53712/rjrs.v5i2.1021>
- Qurohman, M. T., & Sungkar, M. S. (2018a). Integrasi Pembelajaran Matematika Problem Based Learning dengan Teknologi Informasi dan Komunikasi. *Desimal: Jurnal Matematika*, 1(3), 303–313. <https://doi.org/10.24042/djm.v1i3.2908>
- Qurohman, M. T., & Sungkar, M. S. (2018b). Integrasi Pembelajaran Matematika Problem Based Learning dengan Teknologi Informasi dan Komunikasi. *Desimal: Jurnal Matematika*, 1(3), 303–313. <https://doi.org/10.24042/djm.v1i3.2908>
- Rahim, A., Mangampang, E. T., Swandi, A., Rahmadhanningsih, S., & Salim, S. (2023). Analysis of learning difficulties of mathematics for vocational school students in the 3T area. *Indonesian Journal of Educational*

- Studies*, 26(1), 44–54. <https://doi.org/10.26858/ijes.v26i1.47104>
- Rifai, R., Turmudi, D., A., J., & Suhendra. (2025). Utilizing open-ended approach to enhance mathematical understanding among vocational high school students. *Jurnal Elemen*, 11(3), 564–576.
- Sabaruddin, S., Nur, M., Arifin, A. N., Rifai, R., Turmudi, D., A., J., Suhendra, Muljani, S., Bunawar, B., Murad, M., Sopamena, P., Lessy, D., Oktaviastuti, B., Nurmalasari, R., Damayanti, F., Gio, P. U., Sokop, R. M., Arsjad, T. T., Malingkas, G., ... Creswell, J. W. D. (2018). Rancangan Pembelajaran Berkarakteristik Pembelajaran Inovatif Abad 21 Pada Materi Estimasi Biaya Konstruksi Dengan Model Pembelajaran Project Based Learning Di Smk Negeri 1 Adiwarna. *Cakrawala: Jurnal Pendidikan*, 3(1), 303–313. <https://doi.org/10.24905/cakrawala.vi0.226>
- Slameto. (2015). *Belajar dan faktor-faktor yang mempengaruhinya*. Rineka Cipta.
- Sugiyono. (2022). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, R&D dan Penelitian Pendidikan*. Alfabeta.
- Sundari, S., Rusani, I., Musa'ad, F., Setyo, A. A., & Trisnawati, N. F. (2024). Pengaruh model Project Based Learning berbasis entrepreneurship terhadap keterampilan berpikir kreatif dalam pembelajaran geometri ruang. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 13(4), 1174–1188.

Enhancing Fifth Graders' Flat Geometry Problem-Solving Skills through STEM-Based 'Squid Math' Gamified Learning Media

Yuni Mariyati¹, Intan Dwi Hastuti², Reny Widya Astuti³

^{1,2,3}Department of Primary Education, Universitas Muhammadiyah Mataram, Pagesangan, Indonesia

¹rennywdyasti433@gmail.com, ²intandwihastuti88@ummat.ac.id, ³yunimariyati31@gmail.com.

ABSTRACT

This study investigates the effect of implementing the STEM-based "Squid Math" game on fifth-grade students' problem-solving abilities in plane geometry. The study employed a quasi-experimental method with a nonequivalent control group design. The participants consisted of 15 fifth-grade students from SDN 2 Jerowaru as the experimental group and 23 fifth-grade students from SDN 2 Pemas as the control group. The research instrument used was an essay test that had been tested for validity and reliability. The data were analyzed through normality testing, homogeneity testing, and hypothesis testing using the Independent Samples T-Test with the assistance of SPSS 25. The findings revealed that the experimental group's mean score increased from 53.67 in the pre-test to 84.17 in the post-test, whereas the control group's mean score improved from 41.26 to 61.85. The implementation of learning activities reached 100% in the experimental group and 95% in the control group. The hypothesis test produced a significance value of $p = 0.000 < 0.05$, indicating a significant difference between the two groups. Therefore, the STEM-based "Squid Math" game media significantly contributed to improving fifth-grade students' problem-solving skills in plane geometry. This learning media is recommended as an innovative alternative for mathematics instruction in elementary schools.

Keywords: STEM, Squid Math, Problem-Solving Skills, Flat Shapes.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Reny Widya Astuti,
Department of Primary Education,
Universitas Muhammadiyah Mataram,
Jalan KH Ahmad Dahlan No 1 Mataram, Indonesia.
rennywdyasti433@gmail.com

1. INTRODUCTION

Education constitutes an essential process in developing human potential in a holistic manner, including cognitive, affective, and psychomotor dimensions. Santoso and Widodo (2022) state that education is a structured and deliberate endeavor carried out by both individuals and institutions to establish a supportive learning environment, allowing students to acquire the knowledge, skills, and attitudes required for participation in social and professional contexts. This perspective is consistent with Rahmawati et al. (2022), who argue that education extends beyond the transmission of knowledge and serves as a process of value and character formation that prepares learners to respond effectively to global challenges in the digital era.

In the 21st-century context, education serves as a means of fostering a wide range of competencies, particularly critical thinking, creativity, collaboration, and communication (4Cs), which are vital for producing human resources that are both adaptive and innovative (Putra & Wulandari, 2021). In addition, Nurhasanah et al. (2022) explain that contemporary education should incorporate digital technology into the teaching and learning process in order to promote learning experiences that are interactive, individualized, and contextual, in accordance with current societal demands. Similarly, Wibowo and Lestari (2023) maintain that education functions as a process that supports the comprehensive development of students by implementing an inclusive and learner-centered approach. However, evidence from recent studies indicates that the use of digital technology in mathematics learning remains limited and has not been fully optimized to support students' higher-order thinking and problem-solving skills. Engelbrecht and Borba (2024) reported that digital technologies in mathematics classrooms are still predominantly utilized for content delivery and routine instructional activities, rather than for facilitating meaningful exploration, inquiry, and mathematical problem-solving. This condition suggests a discrepancy between the expectations of 21st-century education, which emphasize interactive and student-centered learning through technology integration, and the

reality of classroom practices. Consequently, there is a need for innovative digital learning media that not only incorporate technology but also actively engage students in solving mathematical problems through meaningful and contextual learning experiences.

From a theoretical standpoint, the learning process outlined above is strongly associated with constructivist theory, which views students as active participants who build their own understanding through interaction and experience. Likewise, experiential learning theory emphasizes that knowledge develops through direct involvement, reflection, and the application of concepts in meaningful situations. These ideas are further supported by the principles of game-based learning, where components such as challenges, immediate feedback, interaction, and rewards are able to increase learners' motivation and cognitive involvement. Taken together, these theoretical perspectives suggest that interactive, contextual, and student-centered learning environments can make a substantial contribution to the development of higher-order thinking skills, particularly problem-solving abilities.

Mathematics plays an essential role in developing students' ability to think logically, critically, and systematically. Therefore, mathematics learning needs to be organized in a contextual and meaningful way, enabling students to comprehend mathematical concepts more effectively and use them to solve problems found in everyday life (Hastuti, 2022). In elementary education, mathematics holds a particularly important position in strengthening students' cognitive development, especially in the field of geometry. One of the basic topics within this field is two-dimensional shapes, which are widely applied in various daily contexts.

Nevertheless, many students still encounter difficulties in comprehending these abstract concepts. A number of studies indicate that conventional teaching methods, which are generally passive in nature, are less effective in enhancing students' interest and understanding of the material (Rokhim, 2021). Mathematics instruction that is overly rigid and theoretical, without being connected to meaningful contexts, often results in only a superficial grasp of concepts. Therefore, instructional innovation is needed to help students develop a deeper understanding through learning experiences that are more active and enjoyable. One effective strategy is to relate mathematical concepts to real-life situations and students' own experiences, thereby making learning more meaningful and engaging (Hastuti et al., 2025).

The use of learning media grounded in Science, Technology, Engineering, and Mathematics (STEM) represents one strategy for addressing these challenges, as it connects mathematical concepts with technological elements and practical applications in real-life contexts. STEM-based learning emphasizes active student involvement and has been proven to improve learners' critical thinking and problem-solving skills (Anggraini et al., 2021). Furthermore, STEM-oriented educational games have been reported to enhance students' learning motivation and deepen their understanding of mathematical concepts (Nurussofa & Astuti, 2023). Educational games also help create a pleasant learning environment, motivating students to engage more actively while making it easier for them to understand the material being studied (Kero & Wewe, 2024).

Conceptually, the integration of STEM and game-based learning can be viewed as a process in which interdisciplinary STEM elements are incorporated into interactive instructional activities. This combination produces a gamified learning environment that promotes greater student engagement, motivation, and active participation. These conditions support a deeper understanding of concepts, which in turn leads to improved problem-solving abilities. Accordingly, the relationship among these components can be described as a sequential process: STEM integration → gamified learning experience → enhanced engagement and conceptual understanding → improved problem-solving skills.

Based on the researcher's observations conducted on April 3, 2025, students' problem-solving abilities, particularly in the topic of two-dimensional shapes, were still relatively low. Most students experienced difficulties in identifying and distinguishing various types of plane figures and were not yet able to relate mathematical concepts to their practical use in everyday situations. One of the primary factors contributing to this condition was the limited availability of interactive and contextual learning media. As a result, students tended to memorize formulas without fully understanding their underlying meaning and fundamental principles. Therefore, it is necessary to provide learning media that actively engage students in meaningful learning experiences.

Problem-solving ability is a cognitive competence that involves a sequence of strategic processes, including understanding the problem, devising a plan, carrying out the chosen strategy, and evaluating

the results, as proposed by Polya (Hartati & Adawiyah, 2023; Lestari et al., 2021). These stages demonstrate that effective instruction should offer students opportunities to participate actively in solving problems rather than merely receiving information passively.

A commonly applied solution in previous studies is the use of educational games to enhance students' understanding of mathematical concepts. Learning media such as *Squid Math*, which integrates game elements with STEM-based instruction, has been shown to increase student engagement and strengthen conceptual understanding in mathematics (Handican et al., 2023; Yuliani et al., 2023). Moreover, STEM-oriented games provide opportunities for students to solve authentic problems and develop critical thinking skills (Kedhi et al., 2024). The incorporation of game media inspired by popular culture can also create learning experiences that are more attractive and relevant to students' interests (Yustina & Yahfizham, 2023). In addition, video- and game-based instructional media have been found to improve both student motivation and learning outcomes (Prasetyo & Nugraheni, 2024).

Although previous research has confirmed the effectiveness of game-based learning and STEM approaches, most studies have examined these components independently or have not integrated them into a single, coherent instructional design. In addition, many studies have concentrated on overall mathematics achievement rather than specifically investigating problem-solving skills in geometry, particularly in the topic of two-dimensional shapes. Furthermore, research that combines STEM-based learning, gamification, and culturally relevant game adaptations such as *Squid Game* remains relatively limited.

Therefore, the novelty of this study lies in the simultaneous integration of three key elements: STEM-based learning, gamified instructional media, and the adaptation of culturally relevant game features through *Squid Math*. In contrast to previous studies that generally emphasize only one or two of these aspects, this research offers a more comprehensive approach by embedding STEM principles within a structured, contextual, and engaging game-based learning environment. In addition, the study specifically focuses on improving students' problem-solving abilities in plane geometry, an area that has received comparatively little attention in prior research.

Squid Math is a game-based mathematics learning concept that adapts elements from popular games to enhance students' engagement and mathematical abilities through challenges that incorporate cognitive, strategic, and social-emotional dimensions (Li, 2024; Yoon, 2023; Chonniah & Izzah, 2022). This game fosters an interactive, competitive, and collaborative learning environment that encourages students to think critically and solve problems effectively. In addition, it supports the development of character traits such as teamwork and emotional regulation (Park, 2024; Kim & Park, 2022). Nevertheless, its implementation must be carefully guided to ensure that educational objectives remain the primary focus and are not overshadowed by entertainment elements.

Several gaps identified in the existing literature include the limited number of studies that comprehensively investigate the influence of STEM-based game media on elementary students' understanding of two-dimensional shapes (Nurussofa & Astuti, 2023; Oliviana et al., 2024). Furthermore, research integrating culturally relevant game adaptations into STEM-based mathematics instruction is still relatively scarce. This indicates the need for further studies on innovative learning media that combine these components.

Accordingly, this study aims to examine the effect of STEM-based *Squid Math* game media on the problem-solving abilities of fifth-grade students in the topic of plane geometry. It is expected that this research will contribute both theoretically and practically by offering an innovative instructional approach that integrates STEM and game-based learning to improve students' understanding, engagement, and problem-solving skills in mathematics.

2. RESEARCH METHOD

a. Research Design

This research employed a quantitative approach using a quasi-experimental method with a nonequivalent control group design. This design was selected because random assignment of participants into different groups could not be conducted; therefore, the study made use of the existing classroom groups. The research involved two groups, namely an experimental group that received instruction through STEM-based *Squid Math* gamified learning media and a control group that was

taught through conventional instruction assisted by flashcard media. In this study, the STEM-based Squid Math learning media served as the independent variable, while students' problem-solving ability in plane geometry functioned as the dependent variable. The purpose of the study was to examine the effect of the STEM-based Squid Math game on students' problem-solving skills compared with conventional teaching methods.

b. Population and Sample

The population in this study included all fifth-grade students of SD Negeri 2 Jerowaru and SD Negeri 2 Paremas in the 2025/2026 academic year, totaling 38 students. Since the number of students was relatively limited, the entire population was used as the research sample. A purposive sampling technique was applied in this study by considering several factors, such as the implementation of the same curriculum, similar school characteristics, and comparable students' academic backgrounds. One class was assigned as the experimental group, while the other class served as the control group.

c. Research Procedure

The research was conducted through three main stages:

1. Pre-test Stage
At this stage, both the experimental and control groups were administered a pre-test to identify students' initial problem-solving abilities in the topic of flat geometry.
2. Treatment Stage
During the treatment stage, the experimental group participated in learning activities using STEM-based "Squid Math" gamified media, which combines mathematical concepts with interactive game elements. In contrast, the control group received instruction through conventional teaching approaches supported by flashcard media.
3. Post-test Stage
Following the treatment, both groups were given a post-test to determine the improvement in students' problem-solving abilities.
Through these stages, the researcher was able to compare the outcomes of the two groups and evaluate the effectiveness of the treatment.

d. Data Collection Technique

The data were obtained through a problem-solving ability test presented in the form of fill-in-the-blank questions with a structured essay format. The test was developed based on several indicators of problem-solving ability, namely:

1. identifying and understanding the problem,
2. formulating a solution strategy,
3. carrying out the selected solution, and
4. reviewing and evaluating the results.

To maintain measurement consistency, the same instrument was administered in both the pre-test and post-test.

e. Instrument Test

1) Validity Test

Validity testing was conducted by comparing the predetermined criteria with the actual conditions or situations observed. The validity of the instrument was examined after it had been administered to users or respondents by applying the Pearson Product Moment correlation. In testing external validity, SPSS version 25 for Windows was utilized, using the following formula:

$$r_{XY} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

Table 1. Interpretation of Validity Coefficients

Coefficient Interval	Category
0,00 – 0,199	Very Low
0,20 – 0,399	Low
0,40 – 0,599	Moderate
0,60 – 0,799	High
0,80 – 1,000	Very High

2) Reliability Test

Reliability testing was performed to determine the consistency of the instrument in measuring students' abilities. An instrument can be considered reliable when it produces stable and consistent results when administered repeatedly. In this study, the instrument reliability was analyzed using the Cronbach's Alpha method, which is suitable for essay-type tests. The reliability coefficient was calculated with the support of SPSS version 25. The criteria for interpreting the reliability coefficient are shown in Table 2.

Table 2. Criteria for the reliability coefficient of the instrument

Harga r	Description
0.00 – 0,20	Very Low
0,21 – 0,40	Low
0,41 – 0,60	Moderate
0,61 – 0,80	High
0,81 – 1,00	Very High

f. Data Analysis Technique

The data analysis process was carried out in several steps. The first step involved conducting prerequisite tests, which consisted of:

1. normality testing, to identify whether the data followed a normal distribution, and
2. homogeneity testing, to examine whether the variances between the groups were equal.

After these prerequisite assumptions were fulfilled, hypothesis testing was conducted using the Independent Sample T-Test to determine whether a significant difference existed between the experimental and control groups. All statistical analyses were completed with the assistance of SPSS version 25.

3. RESULTS AND DISCUSSION

The data in this study were analyzed using several statistical procedures, including normality testing, homogeneity testing, and hypothesis testing through the Independent Samples T-Test.

a. Prerequisite Test Analysis

1) Normality Test

The normality test was conducted to examine whether the data for each variable were normally distributed. In this study, the Shapiro-Wilk test was applied. The decision-making criterion was based on the significance value (Sig.); if the Sig. value was greater than 0.05, the data were considered to be normally distributed. The results of the Shapiro-Wilk test are presented in Table 3 below:

Table 3. Normality Test Results

Tests of Normality		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Kelas	Statistic	df	Sig.	Statistic	df	Sig.
Hasil Kemampuan	Pretest_Kontrol	,160	23	,130	,936	23	,146
Pemecahan Masalah	Posttest_Kontrol	,124	23	,200*	,929	23	,106
	Pretest_Eksperimen	,190	15	,150	,928	15	,256
	Posttes_Eksperimen	,133	15	,200*	,948	15	,496

Referring to the results shown in the table, the Shapiro-Wilk test was employed to assess the normality of data in both the experimental and control groups by considering the significance (Sig.) values. At a 5% significance level, the data were classified as normally distributed when the obtained significance value was greater than 0.05. Conversely, when the significance value was equal to or below 0.05, the data were interpreted as not normally distributed.

2) Homogeneity Test

The homogeneity test was performed to examine whether the data variances between the two research groups were equal. The data were categorized as homogeneous when the significance value was greater than 0.05, while a significance value lower than 0.05 indicated that the variances were not homogeneous. In this study, the homogeneity of variance was tested using Levene's Test (F test), with the analysis conducted through SPSS version 25 for Windows.

Table 4 Homogeneity Test Results
Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Hasil	Based on Mean	1,123	3	72	,346
Kemampuan	Based on Median	,997	3	72	,399
Pemecahan	Based on Median and	,997	3	59,434	,400
Masalah	with adjusted df				
	Based on trimmed	1,120	3	72	,347
	mean				

Table 4 indicates that the significance value for the experimental and control groups is 0.346. Since this value is greater than 0.05, it can be concluded that the variances of the two groups are homogeneous.

b. Hypothesis Testing

The prerequisite analysis results indicated that the problem-solving ability data obtained from both the experimental and control groups were normally distributed and showed homogeneous variances. Since these assumptions had been fulfilled, hypothesis testing was then carried out using SPSS version 25 for Windows. The hypothesis was tested using the Independent Samples T-Test as the statistical analysis technique.

Table 5. Hypothesis Test Results

Independent Samples Test		Levene's Test for Equality of Variances				t-test for Equality of Means		95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Upper
Hasil	Equal variances assumed	,483	,491	-7,066	36	,000	-22,31884	3,15871	-28,72500 -15,91268

Equal variances not assumed	- 7,014	29,335 ,000	-22,31884	3,18184	- 28,82321	- 15,81447
--------------------------------------	------------	-------------	-----------	---------	---------------	---------------

The hypothesis testing results showed a p-value of 0.000, which was below the 0.05 significance level, with a degree of freedom (df) of 36. This result indicates that the null hypothesis was rejected, while the alternative hypothesis was accepted.

Based on the descriptive statistical results, the use of STEM-based Squid Math game media was found to have a significant influence on enhancing students' problem-solving abilities. This can be seen from the increase in the experimental group's mean score, which improved from 53.67 in the pre-test to 84.17 in the post-test. This considerable increase indicates that the combination of STEM-based learning and educational game media offers a more meaningful learning experience and facilitates students' thinking processes in solving mathematical problems. In addition, the implementation of learning activities in the experimental group reached 100%, suggesting that the game media was implemented highly effectively and supported the learning process in accordance with the stages of the STEM approach.

The improvement may also reflect stronger cognitive engagement among students. Through interactive game-based learning activities, students were actively engaged in examining problems, trying out different strategies, and reflecting on the accuracy of their responses. This learning process aligns with the development of 21st-century 4C competencies, particularly critical thinking and creativity, since students were required to analyze mathematical problems, formulate possible solutions, and adjust their strategies within the game-based learning context.

In the control group, which received instruction through the Project Based Learning (PjBL) approach, students also showed progress in problem-solving abilities; however, the improvement was lower than that of the experimental group. The implementation of learning activities in the control class reached 95%. The average score of the control group increased from 42.17 in the pre-test to 60.00 in the post-test. Although the PjBL approach contributed positively to students' problem-solving skills, STEM-based learning assisted by Squid Math media resulted in a greater improvement. These results suggest that the intervention implemented in the experimental group was more effective than the instructional approach applied in the control group.

The results of the normality and homogeneity tests indicated that the data from both groups were normally distributed and had homogeneous variances. These findings confirm that the assumptions required for hypothesis testing using the Independent Samples T-Test had been met. The hypothesis test yielded a significance value of $p = 0.000$, which was less than 0.05. Thus, it can be concluded that a statistically significant difference existed between the problem-solving abilities of students taught using STEM-based Squid Math game media and those taught through the Project Based Learning (PjBL) approach.

In addition to the quantitative findings, qualitative data obtained through semi-structured interviews were analyzed to further explore students' mathematical problem-solving processes after participating in the STEM-based "Squid Math" learning activities. Representative interview responses are presented in Table 6.

Table 6. Representative Interview Findings

Problem-Solving Indicator	Interview Question	Representative Student Response
Understanding the problem	What information did you identify before solving the geometry problem?	"I first identified the known measurements and what the question asked before choosing the formula."
Planning the solution	How did you decide which strategy or formula to use?	"I selected the formula based on the shape shown in the game because each challenge reminded me of the correct concept."
Carrying out the solution	Can you explain how you solved the problem?	"I solved it step by step, starting from writing the formula, substituting the numbers, and calculating the answer."
Reviewing the solution	Did you check your answer after finishing? Why?	"Yes, I recalculated it because I wanted to make sure my answer was correct."
STEM learning experience	How did the STEM activities help you solve the geometry problems?	"The activities helped me understand how geometry is used in everyday situations, making the problems easier to solve."
Gamified learning experience	Which feature of Squid Math helped you the most?	"The levels, challenges, and instant feedback motivated me to keep trying until I solved the problems correctly."

The representative interview responses presented in Table 6 support the quantitative findings of this study. Overall, students in the experimental group demonstrated a better understanding of the mathematical problem-solving process after participating in the STEM-based "Squid Math" learning activities. Most students were able to identify relevant information from geometry problems, determine appropriate solution strategies, explain the solution procedures systematically, and verify the correctness of their answers. In addition, students reported that the STEM activities helped them relate geometric concepts to real-life situations, while the gamified features, such as challenges, levels, and immediate feedback, encouraged them to remain engaged and motivated throughout the learning process. These findings suggest that the integration of STEM principles with gamified learning not only improved students' mathematical problem-solving performance but also promoted reflective thinking and active participation during mathematics learning.

These findings suggest that the learning process supported students' reflective thinking, particularly in monitoring and evaluating their own problem-solving strategies. The interview responses indicated that many students consciously identified relevant information, selected appropriate solution strategies, and rechecked their answers before reaching a final solution. These learning behaviours reflect a more systematic approach to mathematical problem solving and reinforce the potential of STEM-based gamified learning to promote deeper engagement during mathematics instruction.

The findings of this study are consistent with those reported by Rosita Putri (2023), who found that STEM-based learning significantly improved students' mathematical problem-solving abilities. The increase in students' post-test scores observed in the present study further supports the argument that integrating STEM principles into mathematics instruction provides meaningful learning experiences that promote analytical thinking and systematic problem solving. These findings also extend previous evidence by showing that the integration of STEM principles with gamified learning activities can

support students in developing more organized reasoning processes while solving flat geometry problems.

The present findings are also consistent with the study conducted by Ajeng Fani Yustina and Yahfizham (2023), which demonstrated that game-based learning media can increase students' interest and engagement in mathematics learning. However, unlike the game-based media developed in their study, the STEM-based Squid Math media was intentionally designed by integrating Science, Technology, Engineering, and Mathematics principles into structured learning activities that actively guide students through mathematical problem-solving processes. This integration may explain the improvement in students' mathematical problem-solving performance observed in the present study. Rather than functioning solely as a motivational tool, the STEM-based Squid Math media provides opportunities for students to develop conceptual understanding and systematic reasoning while solving flat geometry problems.

Furthermore, compared with previous studies that primarily emphasized students' motivation and engagement, the present study provides additional evidence that integrating STEM components into game-based learning can facilitate more structured mathematical problem-solving processes. These findings suggest that the combination of STEM principles and gamified learning not only enhances students' learning engagement but also supports systematic reasoning throughout the problem-solving process.

The findings of the present study further support recent evidence that integrating game-based learning within a STEM framework can enhance students' motivation, focus, and active participation during mathematics learning. In this study, students became more engaged in solving flat geometry problems through contextual challenges and interactive game activities embedded in the STEM-based "Squid Math" media. These findings are consistent with recent studies indicating that gamification in STEM learning positively influences students' engagement, learning outcomes, and mathematical understanding when supported by appropriate instructional design (Arlinwibowo et al., 2025; Mursalin et al., 2024; Wang et al., 2022). Furthermore, the interview findings revealed that students actively explored mathematical concepts, discussed solution strategies, and connected geometry concepts with contextual situations presented in the game. These learning experiences are aligned with the characteristics of STEM education, which emphasize problem solving, creativity, collaboration, and meaningful learning. Therefore, the integration of STEM principles with gamified learning, as implemented in the Squid Math media, provides meaningful opportunities for elementary school students to construct mathematical understanding while strengthening their problem-solving skills.

Moreover, the collaborative elements embedded in the game provide opportunities for students to communicate, exchange ideas, and work together while solving mathematical problems. Such learning experiences are consistent with the characteristics of 21st-century learning, which emphasize communication, collaboration, critical thinking, and problem-solving as essential competencies for students.

This study provides empirical evidence that the STEM-based Squid Math media is an effective instructional approach for improving elementary school students' mathematical problem-solving skills. Through contextual, interactive, and challenging learning activities, students demonstrated a better understanding of flat geometry concepts and were able to apply more systematic problem-solving strategies when solving mathematical tasks. These findings are consistent with recent studies reporting that integrating STEM education with gamified learning enhances students' conceptual understanding, engagement, and mathematical problem-solving performance (Arlinwibowo et al., 2025; Wang et al., 2022). Therefore, the STEM-based Squid Math media can be considered an innovative instructional alternative for supporting meaningful mathematics learning at the elementary school level.

However, several limitations of this study should be acknowledged. First, the sample size was relatively limited, involving only 38 students, which may reduce the generalizability of the findings. Second, the intervention was implemented over a relatively short period, making it difficult to evaluate the long-term effects of the STEM-based Squid Math media. Third, although teacher readiness and competence in implementing STEM-based game learning were not specifically examined in this study, these factors may influence the effectiveness of classroom implementation and should therefore be considered when interpreting the findings.

Therefore, future research is recommended to involve larger and more diverse samples, implement the intervention over longer periods, and examine the effectiveness of the STEM-based Squid Math media across different mathematical topics, grade levels, and educational settings. Future studies may also investigate the long-term retention of students' mathematical problem-solving skills to provide more comprehensive evidence regarding the effectiveness and broader applicability of STEM-based gamified learning.

4. CONCLUSION

Based on the results of data analysis, instrument validation, and statistical testing, it can be concluded that the STEM-based Squid Math game media significantly influenced fifth-grade students' problem-solving abilities. The improvement in the experimental group was shown by the increase in the mean score from 53.67 in the pre-test to 84.17 in the post-test. Meanwhile, the control group, which received instruction through the Project-Based Learning (PjBL) approach, experienced a lower improvement, with the average score rising from 41.26 to 61.85. In addition to the test scores, the level of learning implementation also supported the effectiveness of the instruction. The experimental group obtained an implementation score of 100%, while the control group achieved 95%. This difference shows that STEM-based learning using the Squid Math game was carried out more optimally than the PjBL approach.

The Independent Samples T-Test results produced a significance value of $p = 0.000$, which was below 0.05. This result shows that there was a statistically significant difference in students' problem-solving abilities between the two groups. Therefore, the alternative hypothesis (H_a) was accepted, indicating that the Squid Math game media was effective in enhancing students' problem-solving skills. Overall, the integration of STEM-based learning with educational game media created a learning experience that was more interactive, attractive, and meaningful. Thus, Squid Math media can be recommended as an innovative alternative for improving the quality of mathematics learning in elementary schools.

From a practical point of view, the findings of this study have important implications for teachers in applying innovative learning strategies. Squid Math media can be adjusted to various classroom contexts by modifying the difficulty level, learning objectives, and students' characteristics. Teachers can integrate this game into regular classroom learning, group discussions, or problem-solving activities to increase students' engagement and participation. Furthermore, the use of this media encourages teachers to implement more interactive and student-centered learning approaches, which are important elements of 21st-century education.

From a theoretical point of view, this study contributes to the development of STEM-based game learning models, particularly in elementary geometry learning. The integration of STEM components with gamified learning through culturally relevant media such as Squid Math provides a more comprehensive framework for improving students' problem-solving abilities. In addition, this study strengthens the idea that combining interdisciplinary learning with game elements can enhance both the cognitive and affective dimensions of the learning process.

Suggestions for Further Research

This study has several limitations; therefore, further research is recommended to extend and enrich the present findings. Future studies may investigate the application of the STEM-based *Squid Math* media to other mathematical topics, such as algebra, fractions, or measurement, in order to evaluate its effectiveness across a wider range of content areas. In addition, subsequent research could develop digital or technology-enhanced versions of the game to improve its accessibility and scalability in diverse educational contexts. It is also advisable to involve larger and more heterogeneous samples and to implement the intervention over a longer period so that more comprehensive and generalizable findings can be obtained.

REFERENCES

- Anggraini, N., Rahman, F., & Sari, D. (2021). Pengaruh pendekatan STEM terhadap keterampilan berpikir kritis dan pemecahan masalah siswa. *Jurnal Pendidikan Matematika*, 9(2), 123-134.
- Arlinwibowo, J., Ishartono, N., Linguistika, Y., Purwoko, D., & Suprpto. (2023). Gamification in the Stem Domain Subject: the Prospective Method To Strengthen Teaching and Learning. *Jurnal Pendidikan IPA Indonesia*, 12(4), 562–572. <https://doi.org/10.15294/jpii.v12i4.48388>
- Chonnia, N., & Izzah, N. (2022). Gamification in mathematics learning: A study of Squid Math's effectiveness. *International Journal of STEM Education*, 9(2), 150-162.
- Chonnia, N. S., & Izzah, M. (2022). Pengembangan media permainan berbasis STEM untuk pembelajaran matematika. *Jurnal Pendidikan Matematika*, 10(1), 45-57.
- Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. *ZDM - Mathematics Education*, 56(2), 281–292. <https://doi.org/10.1007/s11858-023-01530-2>
- Handican, R., Suryanto, D., & Puspitasari, D. (2023). Pengembangan media permainan edukatif berbasis STEM untuk meningkatkan motivasi dan pemahaman matematika siswa. *Jurnal Pendidikan Matematika Terapan*, 8(1), 50-62.
- Hartati, S., & Adawiyah, R. (2023). Polya dan kemampuan pemecahan masalah dalam pembelajaran matematika. *Jurnal Pendidikan Matematika*, 12(1), 67-78.
- Hastuti, I. D., Ahyansyah, Mawaddah, S., & Sutarto. (2022). Development of student books characterized by Indonesian realistic mathematics education to support mathematics problem solving ability. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 6(1), 36–50. <https://doi.org/10.31764/jtam.v6i1.5338>
- Hastuti, I. D., Mariyati, Y., & Sutarto. (2025). The learning trajectory on integer operations material using Sagele dance. *Mathematics Education Journal*, 19(2), 365–388. <https://doi.org/10.22342/mej.v19i2.pp365-388>
- Kedhi, S., Prasetyo, R., & Nugraheni, D. (2024). Efektivitas permainan edukatif berbasis STEM dalam pembelajaran matematika. *Jurnal Teknologi Pendidikan*, 10(1), 89-101.
- Kero, L., & Wewe, A. (2024). Media permainan edukatif sebagai sarana pembelajaran interaktif. *Jurnal Media dan Pembelajaran*, 8(2), 55-64.
- Kim, S., & Park, J. (2022). Pengaruh media permainan terhadap pengembangan karakter siswa. *Jurnal Pendidikan Karakter*, 8(2), 77-89.
- Lestari, D., Oktavianti, M., & Sari, R. (2021). Pemecahan masalah dalam matematika: Indikator dan strategi pembelajaran. *Jurnal Pendidikan dan Pembelajaran Matematika*, 6(1), 34-42.
- Li, X. (2024). Using game-based learning to enhance problem-solving skills: Case of Squid Math. *Journal of Educational Innovation*, 15(1), 33-44.
- Mursalin, M., Fonna, M., Saputra, E., Ali, M., & ... (2024). Gamification in STEM Education: A Systematic Literature Review. *International Journal of Trends in Mathematics Education Research*, 7(3), 182–188. <http://ijtmr.saintispub.com/ijtmr/article/view/368>
- Nurhasanah, D., Prasetyo, A., & Wibowo, T. (2022). Integrasi teknologi digital dalam pembelajaran abad ke-21. *Jurnal Teknologi Pendidikan*, 15(2), 123-136. <https://doi.org/10.1234/jtp.2022.152123>
- Nurussofa, N., & Astuti, D. (2023). Pengaruh media permainan edukatif berbasis STEM terhadap pemahaman matematika siswa. *Jurnal Pendidikan Dasar*, 14(1), 23-35.
- Nurussofa, R., & Astuti, T. (2023). Game-based learning sebagai media pembelajaran matematika di sekolah dasar. *Jurnal Pendidikan Dasar*, 11(2), 15-30.
- Oliviana, A., Wahyuni, S., & Putri, R. (2024). Media pembelajaran berbasis STEM pada materi geometri. *Jurnal Penelitian Pendidikan Matematika*, 11(2), 123–134.
- Park, J. (2024). Developing social and emotional learning through educational games: Case study of Squid Math. *International Journal of Educational Technology*, 17(2), 113–129.
- Prasetyo, A., & Nugraheni, R. (2024). Video dan permainan digital sebagai media motivasi belajar matematika. *Jurnal Edukasi Matematika*, 9(1), 40-53.
- Putra, M. A., & Wulandari, S. (2021). Pengembangan kompetensi 4C dalam pendidikan dasar. *Jurnal Pendidikan dan Pembelajaran*, 10(1), 45-60. <https://doi.org/10.5678/jpp.v10i1.567>
- Rahmawati, F., & Juandi, D. (2022). Pembelajaran matematika dengan pendekatan STEM: Systematic literature review. *Teorema: Teori dan Riset Matematika*, 7(1), 15-28. <https://doi.org/10.25157/teorema.v7i1.6914>
- Rokhim, S. (2021). Pengaruh metode pembelajaran terhadap pemahaman konsep bangun datar siswa SD. *Jurnal Pendidikan dan Pembelajaran*, 9(3), 78-85.
- Rosita Putri. (2023). Pengaruh pembelajaran berbasis STEM terhadap kemampuan pemecahan masalah matematis siswa kelas IV pada materi bangun datar. *Jurnal Pendidikan Matematika*, 11(2), 140-155.
- Santoso, E., & Widodo, H. (2022). Menciptakan lingkungan belajar yang kondusif untuk pengembangan potensi siswa. *Jurnal Pendidikan Holistik*, 12(3), 210-225. <https://doi.org/10.2345/jph.2022.123210>
- Sugiyono. (2017). Statistik untuk penelitian. Bandung: Alfabeta.

- Sugiyono. (2019). Metode penelitian kuantitatif, kualitatif, dan R&D. Bandung: Alfabeta.
- Wang, L. H., Chen, B., Hwang, G. J., Guan, J. Q., & Wang, Y. Q. (2022). Effects of digital game-based STEM education on students' learning achievement: a meta-analysis. *International Journal of STEM Education*, 9(1). <https://doi.org/10.1186/s40594-022-00344-0>
- Wibowo, R., & Lestari, M. (2023). Pendidikan inklusif dan pembelajaran berpusat pada siswa. *Jurnal Kajian Pendidikan*, 18(2), 85-98. <https://doi.org/10.3456/jkp.2023.18285>
- Yoon, S. (2023). The effect of Squid Math game on students' problem-solving skills. *Asian Journal of Educational Technology*, 8(2), 99-112.
- Yuliani, T., Handican, H., & Nugroho, A. (2023). Squid Math: Media permainan edukatif berbasis STEM dalam pembelajaran matematika. *Jurnal Inovasi Pendidikan*, 15(2), 180-193.
- Yustina, A. F., & Yahfizham, Y. (2023). Game Based Learning Matematika dengan Metode Squid game dan Among us. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 7(1), 615–630. <https://doi.org/10.31004/cendekia.v7i1.1946>
- Yustina, D., & Yahfizham, M. (2023). Cultural adaptation of educational games: Squid Game in math learning context. *Journal of Indonesian Education*, 8(1), 90-102.
- Yustina, D., & Yahfizham, M. (2023). Implementasi media permainan berbasis STEM di sekolah dasar. *Jurnal Teknologi Pendidikan*, 11(1), 23-36.
- Yustina, D., & Yahfizham, M. (2023). Implementasi pembelajaran berbasis game dalam konteks pendidikan dasar. *Jurnal Pendidikan Dasar*, 11(3), 55-68.

Wordwall Based Learning Media in Earth Structure Learning: Student Engagement and Students' Perceptions

Dede Saepuloh¹, Lina Mafazi Salsabila², Hurriyah Salasih F³, Hairunisa Lestari⁴,
Herman Supriyatna⁵, Yuli Nurmawati Dewi⁶, Supardi U.S⁷, Dasmo⁸

^{1,6}Department of Mathematics and Science Education, Universitas Indraprasta PGRI, Jakarta, Indonesia

¹dedesaepuloh82@gmail.com, ²linasalsabila2804@gmail.com, ³Hurriyahsalasihf@gmail.com, ⁴hairunisa900@gmail.com,

⁵hermansupriyatna919@gmail.com, ⁶yulinurmawatidewi78@gmail.com, ⁷yulinurmawatidewi78@gmail.com,

⁷supardius77@gmail.com, ⁸amo0903unindra@gmail.com.

ABSTRACT

This study examines the implementation of Wordwall-based learning media in teaching Earth Structure, focusing on student engagement and perceptions in science learning. The study was motivated by the need for interactive learning media to address students' difficulties in understanding abstract science concepts and to increase their learning motivation. A descriptive qualitative design was employed, involving eighth-grade students and science teachers as participants. Data were collected through document analysis, interviews, and questionnaires and analyzed through data reduction, data display, and conclusion drawing. The findings indicate that Wordwall created a more interactive and engaging learning environment. Students demonstrated increased enthusiasm, motivation, and participation during learning activities. Questionnaire results showed that 83.7% of students responded positively to the use of Wordwall. Students perceived game-based activities and interactive quizzes as enjoyable and helpful in understanding Earth Structure concepts. Teachers also reported that Wordwall facilitated assessment and encouraged greater student participation. Despite challenges related to internet connectivity and limited access to digital devices, Wordwall was found to be a useful digital learning medium for supporting student engagement and positive learning experiences in science education.

Keywords: Wordwall, Science Education, Student Engagement, Students' Perceptions, Digital Learning Media.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Dede Saepuloh,

Department of Mathematics and Science Education,

Universitas Indraprasta PGRI (UNINDRA),

Jl. ranco indah No.39, Tj. Barat, Kec. Jagakarsa, Kota Jakarta Selatan 12530, Indonesia

Email: dedesaepuloh82@gmail.com

1. INTRODUCTION

Science education is widely recognized as a cornerstone for developing students' scientific literacy, critical thinking, and problem-solving competencies. Nevertheless, several science topics at the junior secondary level remain challenging because they involve concepts that are abstract and cannot be directly experienced by learners. One such topic is Earth Structure, which requires students to comprehend the Earth's internal layers, tectonic activities, and geological processes through models, diagrams, and scientific representations. In many schools, however, instruction on this topic continues to rely heavily on teacher-centered approaches supported by textbooks and static visual materials. Such practices often provide limited opportunities for students to actively construct knowledge, resulting in lower engagement, declining learning motivation, and persistent misconceptions regarding Earth Structure (Syaiifi & Murwitaningsih, 2022).

Student engagement has increasingly been recognized as a critical component of effective learning because it encompasses students' cognitive, emotional, and behavioral participation throughout the instructional process. A substantial body of educational research suggests that learners who are actively involved in classroom activities generally demonstrate stronger conceptual understanding, greater motivation, and higher academic achievement than those who participate passively (Rahmasari et al., 2022). Consequently, designing learning environments that encourage meaningful interaction has become an essential objective in science education, where conceptual mastery depends not only on receiving information but also on actively exploring scientific ideas through various learning experiences (Li et al., 2023).

The continuing advancement of educational technology has created new opportunities for teachers to redesign conventional instruction into more interactive and learner-centered experiences. Among the many digital learning applications currently available, Wordwall has gained considerable attention as a

gamification-based instructional platform capable of increasing student participation. The platform allows teachers to create a wide range of interactive learning activities, including quizzes, matching tasks, crossword puzzles, maze chase games, random wheels, word searches, and collaborative competitions (Dehghanzadeh et al., 2024). Rather than functioning merely as a medium for content delivery, Wordwall integrates multimedia resources with immediate feedback, scoring mechanisms, leaderboards, countdown timers, and repeated practice opportunities within a single learning environment. These integrated features encourage sustained attention, stimulate students' curiosity, foster healthy competition, and strengthen learning motivation throughout classroom instruction (Li et al., 2024).

Compared with other widely used gamified learning platforms, such as Kahoot and Quizizz, Wordwall provides greater instructional versatility. A single set of learning materials can be transformed into multiple interactive activity formats without requiring teachers to recreate the instructional content (Zeng et al., 2024). Furthermore, the platform supports both online implementation and printable offline activities, enabling teachers to adjust learning according to classroom conditions and the availability of technological resources. Such flexibility is particularly valuable in science education, especially when teaching abstract concepts like Earth Structure that require repeated visualization, continuous reinforcement, and active student interaction to facilitate meaningful conceptual understanding (Ramírez Ruiz et al., 2024).

Growing empirical evidence has confirmed the educational value of Wordwall across different learning contexts. Bariyah and Prastowo (2024) reported that integrating Wordwall into classroom instruction significantly enhanced students' engagement and learning motivation. Saputri and Sukmawati (2024) observed increased classroom participation and a more enjoyable learning environment following the implementation of Wordwall in science lessons. Oktavia et al. (2025) also demonstrated that Wordwall effectively promoted student engagement in geography learning, while Wulandari et al. (2026) found that interactive Wordwall activities improved students' understanding of science concepts. Similar findings were reported by Muh. Al Amin et al. (2026) who showed that gamification-based learning through Wordwall strengthened students' motivation and participation during classroom activities. In addition, Rachmawati et al. (2026) concluded that Wordwall contributes to more enjoyable learning experiences while simultaneously improving students' learning outcomes. Taken together, these studies consistently suggest that Wordwall is an effective digital learning medium capable of supporting active learning across various educational settings.

Although previous investigations have provided convincing evidence regarding the effectiveness of Wordwall, the majority have concentrated primarily on learning achievement, motivation, or academic performance using quantitative and quasi-experimental research designs. Comparatively little attention has been directed toward examining students' engagement during the learning process or exploring how learners perceive the implementation of Wordwall, particularly within science topics that involve abstract concepts such as Earth Structure. Moreover, limited research has examined the practical experiences and challenges encountered by both teachers and students when implementing Wordwall in authentic classroom environments through qualitative inquiry. As a result, there remains insufficient understanding of students' actual learning experiences, engagement patterns, and perceptions during the use of Wordwall-based learning media (Nur et al., 2025).

Considering these research gaps, the present study investigates students' engagement and perceptions regarding the implementation of Wordwall-based learning media in teaching Earth Structure at the junior secondary school level. Using a qualitative descriptive approach, this study aims to provide a more comprehensive understanding of how Wordwall supports active participation, facilitates meaningful science learning, and addresses practical issues that emerge during classroom implementation (Rachmawati et al., 2026). The findings are expected to contribute to the growing body of literature on gamification-based science education while offering practical recommendations for teachers seeking to integrate interactive digital media into science learning environments (Rohmani & Adelia, 2025).

2. RESEARCH AND METHOD

This methodology section details the step-by-step timeline of our project—mapping out the overarching framework, field deployment steps, informational tools, and analytical workflows.

Ultimately, a qualitative descriptive design was chosen to carefully unpack student involvement and personal attitudes surrounding the integration of Wordwall tools during Earth Structure lessons (Rachmawati et al., 2026). Opting for a qualitative lens allows researchers to truly grasp the lived experiences, reactions, and viewpoints of participants right within an authentic classroom environment. As highlighted by Sugiyono (2022:9), this line of inquiry seeks a comprehensive grasp of social dynamics, whereas descriptive studies strive to offer an orderly, fact-based portrait of current realities (Nazir, 2023:43). To this end, our inquiry sets out to map active student involvement, personal viewpoints toward the Wordwall platform, alongside the practical roadblocks encountered during field deployment (Sugiyono, 2022).

SMP Harapan Baru hosted this fieldwork during the second term of the 2025/2026 academic year. The actual participants involved a single science educator alongside thirty eighth grade students who actively interacted with the tailored Earth Structure modules via Wordwall. Architecturally, the project unfolds across three core phases namely planning, execution, and final assessment. During the initial planning phase, the team carried out exploratory background observations, cross examined relevant literature, crafted specialized diagnostic instruments, and structured the gamified activities. Following this, the execution stage focused heavily on embedding the digital platform into daily science classes while gathering evidence via student surveys, interviews, and documentation (Widhiatama & Brameswari, 2024). Lastly, the final assessment phase centered on sorting, streamlining, studying, and thoroughly decoding all gathered evidence before forming conclusions regarding student involvement and views on Wordwall-based media (Sugiyono, 2022).

Information gathering relied on three core tools: surveys, semi structured dialogue, and documentation records. The primary questionnaire aimed to capture student involvement specifically mapping out focus, enthusiasm, peer collaboration, and classroom interaction alongside personal attitudes regarding how practical, appealing, and user-friendly Wordwall actually felt (Saputro & Winanto, 2024). To unlock richer perspectives, we held semi-structured dialogues with the science instructor and five student representatives diving into their real classroom experiences, views on the Wordwall rollout, and operational hurdles. Meanwhile, physical documentation captured real-time class dynamics, student participation, and tangible evidence of platform integration as lessons progressed. The specific blueprint detailing these measurement tools is outlined in Table 1 (Oktavia et al., 2025).

Table 1. Data Acquisition Instrument Grid Matrix

No	Instrument Type	Data Source Subject	Indicators Measured	Format Reference
1	Questionnaire (Likert Scale)	30 Grade VIII Students	Student engagement (participation, enthusiasm, attention, and interaction) and students' perceptions regarding the usefulness, attractiveness, and ease of use of Wordwall	Kemendikbudristek (2022:88)
2	Semi-Structured Interview Guidelines	Science Teacher & 5 Student Informants	Learning experiences, perceptions of Wordwall implementation, classroom participation, and implementation challenges	Oktavia (2025:45)
3	Field Documentation Sheet	Classroom Activities and Wordwall Platform	Learning activities, student participation, learning interactions, and evidence of Wordwall implementation	Rachamawati (2026:172)

We processed the evidence through Miles and Huberman's interactive framework a cycle spanning data collection, reduction, thematic display, and conclusion drawing (Sugiyono, 2022). Our analytical efforts centered on uncovering distinct patterns concerning active student involvement and personal viewpoints regarding Wordwall's rollout during Earth Structure lessons. To safeguard the overall rigor and credibility of our conclusions, the study relied on two layers of verification source and technique triangulation. The former viewpoint compared between the science teachers and their students, whereas

the latter cross examined the evidence across surveys, personal dialogues, and written records. We applied these thorough cross checks to solidify the validity of our insights, offering a well-rounded perspective on how Wordwall shapes connection and perception in science learning (Panis et al., 2025).

3. RESULTS AND DISCUSSION

Here, we unpack the empirical evidence concerning active student involvement and personal viewpoints regarding the rollout of Wordwall-based media within Earth Structure lessons. Organized tables and illustrative charts layout the core data step-by-step, paired with an in depth analytical breakdown to shed much needed light on the multi faceted journey of our learners throughout this field deployment. Specifically, our evaluation closely traces how this gamified medium shifted classroom focus, curiosity, intrinsic drive, and peer communication while simultaneously capturing how practically and visually engaging the tool felt for everyday science schoolwork (Widhiatama & Brameswari, 2024).

Results

Wordwall-Based Learning Media for the Earth Structure Topic

The learning activities in this study employed a Wordwall-based instructional medium specifically designed for the Earth Structure topic in junior high school science. The platform was selected because it supports interactive, game-oriented learning experiences that encourage students to participate more actively during classroom instruction. Rather than functioning solely as a quiz application, the media integrates conceptual explanations, visual representations, and formative assessments within a single learning environment.(Maulana & Widodo, 2025).

The instructional content was organized according to the learning objectives of the Earth Structure topic. Students were introduced to the main layers of the Earth through illustrated diagrams accompanied by concise explanations describing the characteristics and functions of each layer. Following the presentation of the material, learners completed several interactive activities, including drag-and-drop labeling tasks and multiple-choice questions. These activities enabled students to reinforce their understanding while receiving immediate feedback regarding their responses (Widhiatama & Brameswari, 2024).

In addition to the learning content, the media incorporates several supporting features that improve usability and learner engagement. The interface provides navigation menus, learning instructions, progress indicators, score displays, countdown timers, and game elements that allow students to monitor their achievement throughout the lesson (Rachmawati et al., 2026). The combination of visual illustrations, interactive exercises, and instant feedback creates a more engaging learning experience and encourages students to remain actively involved during science instruction Figure 1 illustrates the interface of the Wordwall-based learning media implemented in this research.



Fig 1. Percentage Graph of Student Response Indicator Achievement for Wordwall Media

As illustrated in Figure 3, the developed Wordwall media integrates learning materials and interactive activities within a single digital platform. The interface presents visual explanations of the Earth's structure, accompanied by labeling tasks and formative quizzes that reinforce students' conceptual understanding. Interactive features, including score displays, timers, and progress indicators, provide immediate feedback and support active participation throughout the learning process. These

features were designed to create a more engaging learning experience before evaluating students' engagement and perceptions toward the use of the Wordwall-based learning media (Ramírez Ruiz et al., 2024).

Student Engagement and Students' Perceptions toward Wordwall-Based Learning Media

Survey responses gathered from thirty eighth grade learners at SMP Harapan Baru indicate that deploying Wordwall-driven platforms noticeably boosted active involvement throughout the science lessons. This heightened connection manifests across multiple core markers namely classroom focus, excitement, hands-on participation, and peer communication as gamified challenges and interactive quizzes naturally encourage students to take a much more active role in classroom work. The numerical breakdown tracking these specific engagement levels alongside individual student attitudes is comprehensively detailed in Table 2 ((Nur et al., 2025).

Table 2. Student Engagement and Students' Perceptions toward Wordwall-Based Learning Media

No	Assessment Indicators	Percentage of Achievement	Interpretation Category
1	Student Interest	86%	Very good
2	Learning Motivation	83%	Very good
3	Ease of Use	82%	Very good
4	Understanding of Material	80%	Good
5	Classroom Participation	78%	Good

According to the outcomes recorded in Table 2, learners shared highly favorable feedback regarding Wordwall's integration. Peak scores emerged within learner curiosity markers (86%), followed closely by academic drive (83%) and overall platform accessibility (82%) proving that this tool successfully fostered a highly dynamic, participatory classroom atmosphere. Furthermore, strong metrics in conceptual comprehension (80%) along with daily class engagement (78%) demonstrate that students remained thoroughly invested and gained immensely rewarding educational experiences throughout this study. These identical trends are visually mapped out for comparison in Figure 2.

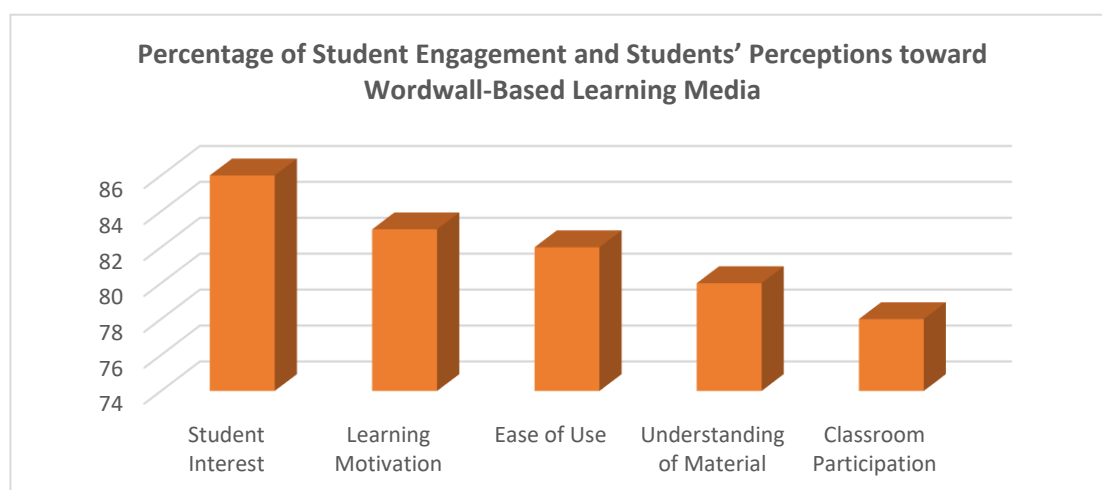


Fig 2. Percentage Graph of Student Response Indicator Achievement for Wordwall Media

As visually depicted, learner curiosity emerged as the peak metric, outstripping scholastic drive and seamless usability. Taken together, every marker fell into the favorable or highly favorable tier proving that children viewed Wordwall positively while showcasing active immersion during the Earth Structure sessions.

Qualitative feedback from discussions further cemented these favorable attitudes regarding this application's role in science education. A majority of learners noted that the system rendered classes

were much more fun, dynamic, and free from traditional monotony. They highly valued the gamified setup since it sparked deep fascination and drove hands on participation during school hours. Additionally, the children felt that Wordwall's rich visual and responsive features actively anchored their educational journey. Integrating diagrams, short tests, and real time feedback helped them unpack the Earth Structure topics much more easily. Finally, they noted that leaderboard dynamics spiked their competitive drive, encouraging everyone to join in. (Dehghanzadeh et al., 2024).

Complementing student views, the science educator highlighted favorable outcomes from the Wordwall rollout noting that the tool streamlined classroom coordination, eased the grading workload, and spurred active learner involvement. Such insights demonstrate that the platform acts as a dual-purpose asset: fostering student connections while offering tangible, timesaving benefits for teachers navigating daily lessons (Syaifi & Murwitaningsih, 2022).

Challenges in Implementing Wordwall-Based Learning Media

While the implementation yielded encouraging results, it was not without its hurdles. The foremost obstacle was unstable internet access, which occasionally disrupted students' engagement with learning materials. Furthermore, a shortage of digital devices restricted the involvement of several participants. These issues were heavily reflected in the interviews with both educators and students, underscoring that technical readiness remains a pivotal factor in executing digital learning initiatives (Rohmani & Adelia, 2025).

To mitigate these issues, students were advised to collaborate in pairs or small teams whenever there was a shortage of devices. Additionally, targeted time management practices were introduced to guarantee that learning goals were met efficiently, all while keeping the students actively invested throughout the entire session (Oktavia et al., 2025).

Discussion

The findings demonstrate that the implementation of Wordwall positively influenced students' learning experiences during the Earth Structure lesson. This conclusion is supported by the high percentage of students reporting increased learning interest (86%), motivation (83%), and active classroom participation (78%). These results suggest that integrating interactive and game-based learning activities encouraged students to engage more actively with the instructional content rather than acting as passive recipients of information. The combination of visual representations, immediate feedback, and interactive challenges appears to have contributed to maintaining students' attention and participation throughout the learning process (Li et al., 2023).

The positive responses observed in this study are consistent with previous research indicating that gamification can enhance learners' motivation and classroom engagement (Sari, 2023; Wardani, 2022). In particular, the Earth Structure topic contains several abstract concepts that are difficult to observe directly. The use of visual diagrams and interactive exercises within the Wordwall platform may therefore assist students in connecting theoretical explanations with visual representations, making scientific concepts easier to understand. Furthermore, the immediate feedback provided after each activity allows students to identify incorrect responses and improve their understanding during the lesson (Nur et al., 2025).

Another noteworthy finding is that 82% of students considered the platform easy to use. This percentage indicates that most participants experienced little difficulty navigating the learning activities despite interacting with a digital platform. Ease of use is an important factor because students can devote more attention to understanding the lesson content instead of learning how to operate the application. Consequently, the instructional process becomes more efficient and creates opportunities for greater classroom interaction between teachers and students (Zeng et al., 2024).

Despite these encouraging findings, several limitations should also be considered. The effectiveness of Wordwall depends on the availability of reliable internet access and appropriate digital devices. In classrooms where connectivity is unstable or devices are limited, students may experience interruptions that reduce the continuity of learning activities. Moreover, because Wordwall incorporates competitive game elements, some learners may focus more on achieving high scores than on developing a comprehensive understanding of the concepts being studied. If not accompanied by teacher guidance and reflective discussion, this tendency may reduce opportunities for deeper conceptual learning.

Nevertheless, the limitations identified in this study did not appear to substantially influence the overall learning outcomes. The consistently high percentages of students' interest (86%), motivation

(83%), classroom participation (78%), and perceived ease of use (82%) indicate that the educational benefits of the platform outweighed the technical challenges encountered during implementation. These findings suggest that Wordwall can serve as an effective instructional medium when supported by adequate technological infrastructure and appropriate pedagogical strategies (Rahmasari et al., 2022).

Overall, this study highlights that Wordwall extends beyond its function as an interactive assessment platform. By integrating gamification, visual learning resources, and continuous feedback, the platform creates a more engaging learning environment that supports students' conceptual understanding and active participation in science learning. Future research is recommended to examine its long-term effects on learning achievement, knowledge retention, and higher-order thinking skills across different science topics and educational settings.

4. CONCLUSION

This research investigated how the integration of Wordwall as a digital learning tool influences student engagement and attitudes during lessons on Earth Structure. The empirical data indicates a strong positive impact on science education, highlighted by high metrics across several indicators: student interest reached 86%, learning motivation stood at 83%, user-friendliness was rated at 82%, conceptual comprehension hit 80%, and active classroom participation sat at 78%. These outcomes are largely driven by Wordwall's gamified and interactive elements, which successfully stimulated student involvement and established a dynamic classroom atmosphere. Ultimately, learners regarded the platform as an appealing and efficient resource that enriched their educational journey and clarified complex geological concepts. While infrastructure hurdles like erratic internet connections and limited device access were noted, they did not overshadow Wordwall's utility in driving engagement and positive feedback. Moving forward, subsequent studies should explore the platform's efficacy across diverse scientific subjects and varied educational settings.

REFERENCES

- Syaifi, A. C., & Murwitaningsih, M. (2022). Analysis of the influence of Problem Based Learning models assisted by Wordwall media on science learning outcomes of primary school students. *Jurnal Paedagogy*, 9(4), 754–762. <https://doi.org/10.33394/jp.v9i4.5430>.
- Rahmasari, R., Murdiono, M., & Sunarso, S. (2022). The influence of utilizing gamification media Wordwall on the improvement of Pancasila Education learning outcomes. *QALAMUNA: Jurnal Pendidikan, Sosial, dan Agama*, 14(2), 671–684. <https://doi.org/10.37680/qalamuna.v14i2.3400>.
- Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis. *Frontiers in Psychology*, 14, 1253549. <https://doi.org/10.3389/fpsyg.2023.1253549>.
- Dehghanzadeh, S., et al. (2024). Using gamification to support learning in K–12 education: A systematic literature review. *British Journal of Educational Technology*, 55, 34–70. <https://doi.org/10.1111/bjet.13335>.
- Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research and Development*, 72, 765–796. <https://doi.org/10.1007/s11423-023-10337-7>.
- Zeng, J., Sun, D., Looi, C. K., & Fan, A. C. W. (2024). Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from the year 2008 to 2023. *British Journal of Educational Technology*, 55(6), 2478–2502. <https://doi.org/10.1111/bjet.13471>.
- Ramírez Ruiz, J. J., Vargas Sanchez, A. D., & Boude Figueredo, O. R. (2024). Impact of gamification on school engagement: A systematic review. *Frontiers in Education*, 9, 1466926. <https://doi.org/10.3389/educ.2024.1466926>.
- Bariyah, C., & Prastowo, A. (2024). Application of Word Wall to foster students' understanding of science learning. *Jurnal Sustainable*, 7(1), 88–97. <https://doi.org/10.32923/kjmp.v7i1.4467>.
- Saputri, A. L., & Sukmawati, W. (2024). The influence of the Team Games Tournament (TGT) learning model assisted by Wordwall on the scientific literacy of class V elementary school students. *Jurnal Penelitian Pendidikan IPA*, 10(7), 3787–3798. <https://doi.org/10.29303/jppipa.v10i7.7992>.
- Dwi Oktavia, A., Ansarullah, M., Tabbu, S., & Simpuluh, I. (2025). Learning with Wordwall: A Quasi-Experimental Study on Student Engagement and Learning Outcomes in Geography at a Senior High School in Indonesia. In *Information Technology Education Journal* (Vol. 4, Number 2).
- Fitria Irza Wulandari Paus Iskarni. (2026). Enhancing Student Engagement and Learning Outcomes in Social Studies Through Wordwall: A Quasi-Experimental Study in a Junior High School Context. *Indonesian Research Journal in Education (IRJE)*, (1), 416–431. <https://doi.org/10.22437/irje>

- Muh. Al Amin, Syam, N., & Juniarti Iryani. (2026). Effectiveness of Wordwall as a Gamification-Based Learning Medium in Enhancing Student Motivation and Learning Activities. *Jurnal Pendidikan Fisika*, 14(1), 53–68. <https://doi.org/10.26618/bw3a7f07>
- Nur, A. M., Fitriany, N., & Syahra, P. A. (2025). Effectiveness Of Using Gamification-Based Wordwalls To Improve Understanding Of Food Chain Concepts In Science Learning. In *Jurnal Keguruan dan Ilmu Pendidikan(JKIP) FKIP Unismuh Makassar* (Vol. 12, Number 1).
- Rachmawati, S., Rahmawati, I., Ernawati, S., & Wahjusaputri, S. (2026). Evaluation of Wordwall Based Joyful learning to Improve Learning Outcomes. *Jurnal Penelitian Ilmu Pendidikan*, 19(1), 13–22. <https://doi.org/10.21831/jpip.v19i1.94567>
- Rohmani, S., & Adelia, B. (2025). *The Role of Wordwall Gamification Mechanics in Developing Higher Level Thinking Skills (HOTS) of Elementary Science Students: A Systematic Review*. 3, 6076–6086. <https://journal.institiercom-edu.org/index.php/multiple>
- Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Edisi ke-2). Alfabeta.
- Widhiatama, D. A., & Brameswari, C. (2024). *International Journal of Linguistics, Literature and Translation The Effectiveness of Wordwall in Enhancing Students' Engagement and Motivation in Literature Classes*. <https://doi.org/10.32996/ijllt>
- Saputro, A. P. S., & Winanto, A. (2024). The use of Word Wall learning media in science learning to improve the learning outcomes of class 5C students of elementary school Kutowinangun 01 Salatiga. *Tunas: Jurnal Pendidikan Guru Sekolah Dasar*, 10(1). <https://doi.org/10.33084/tunas.v10i1.9098>.
- Oktavia, A. D., Ansarullah, M., Tabbu, S., & Simpuruh, I. (2025). Learning with Wordwall: A quasi-experimental study on student engagement and learning outcomes in geography at a senior high school in Indonesia. *Information Technology Education Journal*, 4(2), 94–105. <https://doi.org/10.59562/intec.v4i2.7779>.
- Panis, I. C., Ki'i, O. A., Juan, F. A., & Freitas, M. L. F. (2025). Gamification-based learning design with Wordwall to improve learning activities and learning outcomes. *Journal of Education Technology*, 9(2). <https://doi.org/10.23887/jet.v9i2.96543>.
- Maulana, A., & Widodo, S. T. (2025). Effectiveness of the Deep Learning approach assisted by Wordwall media on Civics Education learning outcomes. *Jurnal Prima Edukasia*, 13(3). <https://doi.org/10.21831/jpe.v13i3.85058>

MaturiKKA-SD: Development of a Maturity Assessment Instrument for CAI Implementation in Indonesian Primary Schools

Mirlanda Diaz Pratiwi¹, Cepi Riyana², Riche Cynthia Johan³

^{1,2,3}Department of Curriculum Development, Universitas Pendidikan Indonesia, Bandung, Indonesia

¹mirlanda.diaz.2410@upi.edu, ²cepi@upi.edu, ³riche@upi.edu

ABSTRACT

The rollout of the Coding and Artificial Intelligence (CAI) subject in Indonesian primary schools in 2025/2026 exposed a systemic gap because no validated instrument was existed to diagnose schools' institutional readiness for this curricular innovation. This study developed and validated MaturiKKA-SD, a multi-respondent maturity assessment instrument grounded in eight institutional dimensions derived from global maturity frameworks and Indonesian curriculum policy. A Design and Development Research (DDR) approach was applied through four stages consisting of content analysis of forty-one academic and policy sources, instrument design refinement through a mini focus group discussion with four experts, content validation by three independent experts using the Content Validity Index, and limited usability testing at one primary school in Bandung. Content analysis produced eight core dimensions and twenty-seven operational constructs underpinning a five-level maturity continuum. The focus group discussion confirmed the conceptual architecture while refining descriptors and terminology, yielding fifty-nine items across three respondent forms. Expert judgment returned an Item-level CVI of 1.00 for all items and a Scale-level CVI Average of 1.00. Usability testing confirmed independent readability and practical applicability while surfacing minor refinements. MaturiKKA-SD provides a theoretically grounded and policy-aligned diagnostic tool for Indonesian primary schools implementing the CAI subject.

Keywords: AI literacy, Coding and AI subject, Content validity, Maturity assessment, Primary education



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Mirlanda Diaz Pratiwi,
Department of Curriculum Development,
Universitas Pendidikan Indonesia,
Jalan Dr. Setiabudhi No. 229, Bandung 40154, Jawa Barat, Indonesia.
mirlanda.diaz.2410@upi.edu

1. INTRODUCTION

Artificial intelligence, big data, the Internet of Things, and algorithmic automation have reshaped much of how societies, economies, and education systems work in the twenty-first century (OECD, 2025). The World Economic Forum (2023) projects that most of the jobs today's primary school children will eventually hold do not yet exist, and that many of those jobs will demand advanced digital competencies. Education systems around the world have responded by shifting away from knowledge transmission toward computational thinking, algorithmic reasoning, and AI literacy (Rodrigues et al., 2024; Selwyn, 2022). UNESCO (2023) argues that digital competence is no longer something reserved for higher education. It must be cultivated from the earliest years of schooling.

Indonesia responded to this push with *Permendikdasmen Nomor 13 Tahun 2025*. The regulation introduced the Coding and Artificial Intelligence (CAI) subject as an elective in primary and secondary curricula starting in the 2025/2026 academic year. Grade V was given seventy-two annual lesson hours and Grade VI sixty-four. Both grades work toward four learning outcomes, namely computational thinking, digital literacy, AI literacy and ethics, and AI utilization and development (Kemendikdasmen, 2025). With this regulation, CAI moved from being an optional enrichment to a formal curriculum component with defined outcomes, time allocations, and evaluation mechanisms.

In practice, however, Indonesian primary schools are far from uniform in their readiness for this curricular change (Pratiwi et al., 2026; Susanto et al., 2026; Zhang et al., 2026). Independent survey evidence corroborates this disparity. A nationwide mixed-methods survey of 468 Indonesian elementary school teachers found that 85.3% had received no relevant training or professional development for AI integration, and that urban teachers reported significantly higher competence than their rural counterparts across all five AI-related knowledge domains (Zhang et al., 2026). Similarly, a survey of 296 elementary teachers across three provinces recorded only a "fairly good" level of digital literacy, with pronounced deficiencies in the cognitive dimension, underscoring the uneven distribution of professional development across regions (Zulaikha et al., 2025). Further studies show that school-level

enactment of the coding and AI policy is incremental and contingent on leadership and individual "champion" teachers (Susanto et al., 2026), that even early-adopting, well-resourced schools confront second-order barriers such as teacher beliefs (Sulasmono et al., 2026), and that teachers who use technology confidently in their personal lives rarely integrate it into instruction (Fitri, 2025). Findings from our preliminary interviews with six teachers from public and private schools in early 2025 reflected these patterns. Public-school teachers reported a lack of specific training in CAI pedagogy, whereas private-school teachers described access to internal curricula, better facilities, and regular professional development. Without a systematic way to diagnose each school's capacity, the national CAI policy could end up widening the gap between resource-rich and under-resourced schools (Pratiwi et al., 2026; Zhang et al., 2026).

Five structural challenges shape how the CAI subject can be implemented at the primary level, and they tend to reinforce one another. The most pressing is teacher competence. Unlike secondary schools that employ subject-specific informatics teachers, primary schools rely on class teachers whose digital competencies are usually elementary (Abidin et al., 2023; Isrokatun et al., 2023; Warsina, 2025). The second is uneven digital infrastructure, which limits technology-based learning particularly outside major urban centers (Zuhri et al., 2024). The third is that school-level operational curricula have not yet integrated CAI into Operational School Curricula, Learning Trajectories, or teaching modules adapted to Grade V and VI learners (Awaluddin & Hadi, 2025). The fourth is school leadership. Institutional assessments treat it as a primary variable because the commitment and managerial capacity of the principal directly shapes whether curricular innovation can happen at all (Normén-Smith et al., 2024). The fifth is cultural resistance and low psychological safety for teachers to try new pedagogical approaches. These barriers often go unnoticed but quietly determine outcomes (Selwyn, 2022). These five problems do not operate in isolation. Competent teachers cannot do much without leadership support. Infrastructure goes unused when teachers lack pedagogical capacity. Curriculum documents stall in schools that resist change.

The heterogeneity of Indonesian primary schools makes the picture more complex. *Permendikdasmen Nomor 13 Tahun 2025* implicitly acknowledges this through clauses like "in accordance with available resources" and the permission to adopt unplugged approaches. That flexibility is sensible. But without a way to measure where each school stands, implementation has no clear map. Fullan et al. (2023) argues that failed curricular innovation rarely comes from poor policy. More often it comes from undetected institutional incapacity, which allows generic interventions to bypass the specific position of each school. Suharyo et al. (2024) make a similar point. Without precise diagnostic tools, capacity-building interventions tend to miss their targets and waste resources.

Maturity assessment fits this problem well. Baihaqy & Subriadi (2023) show that maturity instruments let educational organizations map their position gradually and across multiple dimensions, producing diagnostic information that can guide interventions proportionate to each institution's situation. Begicevic Redjep et al. (2024) make a complementary point. In the absence of systematic maturity assessment, digital transformation can degenerate into sporadic, unsustainable technology adoption. This is roughly the pattern Indonesian primary schools have followed in ICT adoption over the past twenty years. Maturity models also suit the heterogeneity of Indonesian schools because they treat readiness as a developmental continuum rather than a yes/no condition.

A systematic review of the international literature, together with consultation with national policy authorities, indicates a clear research gap. The maturity instruments currently available, such as the Framework for Digital Maturity in Schools (Begicevic Redjep et al., 2021), the ICT in School Education Maturity Model (Valdivia-Salazar et al., 2025), and the Maturity Model for Educational Organizations (Baihaqy & Subriadi, 2023), were developed for general schooling or for higher education in European contexts. None of them treat CAI as a formal subject with specific learning outcomes. Arriola-Mendoza and Ureña (2024) and Casal-Otero et al. (2023) report through systematic literature reviews the near absence of AI-readiness instruments aimed at primary schools. Hartikainen et al. (2023) and Yim and Su (2024) add that dimensions of ethics and human-centered AI are still missing from existing instruments. Consultation with curriculum specialists at the National Centre for Curriculum and Learning of the Indonesian Ministry confirmed that no systematic instrument or guide currently exists for diagnosing CAI implementation maturity at the primary level. The result is a kind of layered

blindness. Schools do not know where they stand. Education offices cannot map progress. Policymakers lack the evidence to evaluate national implementation.

The present study tries to address that gap by developing MaturiKKA-SD, a multi-respondent maturity assessment instrument for CAI implementation in Indonesian primary schools. We developed it through a Design and Development Research approach (Richey & Klein, 2014). The approach combines content analysis, focus group discussion, expert judgment, and limited usability testing within a single iterative cycle. The instrument has eight institutional dimensions and twenty-seven operational constructs, organized on a five-level maturity continuum from Initial to Optimized. The contribution is novel in three ways. On the theoretical side, it offers an original synthesis of global maturity frameworks and Indonesian national curriculum policy. On the methodological side, it operationalizes DDR for instrument development through a sequence of qualitative and psychometric procedures. On the practical side, it gives schools, supervisors, and policymakers a diagnostic tool they can actually use. The study asks what configuration of dimensions, constructs, and maturity descriptors produces a valid, reliable, and applicable instrument for diagnosing the implementation of CAI at the primary education level.

2. RESEARCH METHOD

This study used the Design and Development Research (DDR) approach formulated by Richey and Klein (2014). DDR is a systematic investigation into the design, development, and evaluation of educational products. Its aim is to produce both the product itself and generalizable knowledge about how the product was developed and where it works. The choice of DDR was deliberate. We considered other options and rejected them. The ADDIE model (Branch, 2009) focuses on instructional program development rather than instrument construction. The Research and Development model of Creswell and Creswell (2023) requires extensive field cycles that are disproportionate to the present focus. Design-Based Research (Tinoca, 2022) is better suited to iterative classroom interventions than to building institutional measurement tools. DDR was the only approach that explicitly treats instrument development as a legitimate object of research that yields both a product and methodological knowledge.

Operationally, the study moved through four stages aligned with Richey and Klein's framework. Stage 1 was needed analysis through content analysis. Stage 2 was instrument design development through a mini focus group discussion. Stage 3 was design validation through expert judgment. Stage 4 was instrument evaluation and revision through limited usability testing. Each stage produced an output that fed into the next, which kept continuity across the development cycle. Figure 1 shows the four-stage procedure with the corresponding inputs, methods, and outputs.

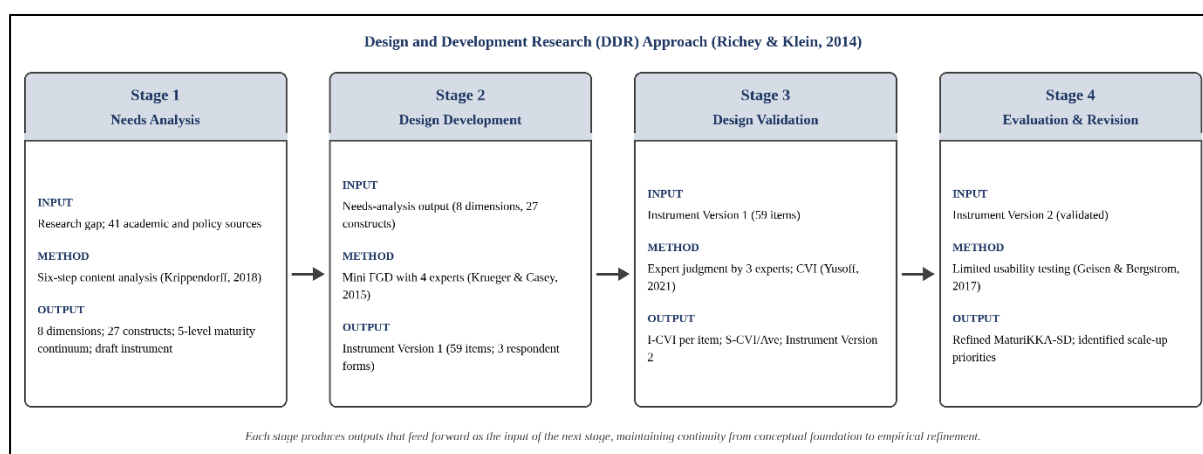


Fig 1. The Four-Stage DDR Procedure of MaturiKKA-SD

The first stage applied content analysis using the six-step procedure of Krippendorff (2018). The six steps are unitising, sampling, recording and coding, reducing, inferring, and narrating. We selected forty-one sources through purposive document sampling using three inclusion criteria. The first criterion covered peer-reviewed journal articles indexed in Scopus or SINTA on digital transformation maturity, ICT maturity models, AI-in-education maturity models, and educational assessment. The

second covered global policy documents issued by UNESCO and the OECD on digital education and AI. The third covered Indonesian national policy instruments, namely the Academic Manuscript on CAI, *Permendikdasmen Nomor 13 Tahun 2025*, the CAI Implementation Guide, and related curriculum documents. We excluded sources that fell outside substantive relevance or relied on a single theoretical frame. Trustworthiness was secured through source triangulation and through documentation of each analytic decision in an audit trail.

The second stage used a mini focus group discussion with four expert participants. We chose them through purposive sampling on the basis of competence, experience, and direct relevance to the constructs we were developing. The four participants represented four complementary perspectives. These perspectives covered curriculum development and educational technology, computer science and artificial intelligence, evaluation and assessment of primary education, and field practice in CAI-implementing primary schools. The discussion followed the small-group format recommended by Krueger and Casey (2015). In their formulation, four to six expert participants generally produce deeper analytical engagement than larger groups when the topic is technical. Each session lasted about ninety minutes. We audio-recorded with consent and supported the recordings with field notes and verbatim transcripts. The moderator used a guide of twenty questions organized into five discussion phases. The questions covered the eight dimensions, the items per respondent, the descriptors of each maturity level, and the scoring guide. The aim was not to construct items from scratch. The aim was to validate and refine the draft instrument produced in Stage 1 through expert judgment.

The third stage involved content validation by three independent experts whose qualifications differed from those of the focus group participants. The validators represented three complementary fields. These fields were curriculum development and educational technology, evaluation and educational measurement, and substantive CAI and digital education policy. The choice of three validators followed the Content Validity Index procedure of Yusoff (2021), which sets $I-CVI \geq 0.78$ and $S-CVI/Average \geq 0.90$ as adequacy thresholds for a three-validator panel. Each expert evaluated all items on a four-point scale. A score of 1 meant not relevant, 2 meant somewhat relevant, 3 meant quite relevant, and 4 meant highly relevant. There was no neutral midpoint. We followed Guraya et al. (2022) in dropping the midpoint to improve decisional clarity and reduce ambivalence bias. Items receiving scores of 3 or 4 from two or more experts were considered content-valid. The validators worked independently and did not consult one another. This preserved the objectivity that the CVI procedure requires.

The fourth stage examined the usability and practical applicability of the validated instrument through a limited usability test at one primary school in Bandung. Three respondents took part. They represented the actual user roles for which the instrument was designed, namely the principal, the deputy head for curriculum, and a Grade V class teacher. The small sample was deliberate. As Swacha et al. (2025) note, three to five participants in usability testing surface most usability problems for a well-defined instrument. Data collection followed the survey usability testing framework of Geisen and Bergstrom (2017) and used three integrated techniques. The think-aloud protocol asked participants to verbalize their reasoning while reading instructions, interpreting terminology, and selecting maturity levels. Structured observation captured non-verbalized indicators of difficulty. These included hesitations longer than ten seconds, expressions of confusion, and changes of answer. A semi-structured probing interview at the end of each session gathered retrospective reflections on clarity, level differentiation, and recommendations for improvement. We analyzed findings thematically and translated them into targeted refinements of the instrument.

3. RESULTS AND DISCUSSION

A. Needs Analysis and the Formation of the MaturiKKA-SD Conceptual Model

The content analysis of forty-one sources produced four cumulative findings. First, the unitizing phase classified the sources into six thematic clusters. Eleven sources dealt with digital and AI maturity models. Four covered digital and AI competency frameworks for educators. Nine dealt with digital, AI, and coding literacy. Three came from the psychology of technology adoption and scale development. Seven were international policy documents, and another seven were national policy documents. The geographic spread of these sources was wide. It covered the United States, the European Union, UNESCO, the People's Republic of China, Hong Kong, and Indonesia. The sources also combined

maturity models, competency frameworks, literacy literatures, and policy instruments. This range meant the conceptual model rested on broad rather than narrow evidence.

Second, the recording and coding phase yielded four hundred and twenty-nine analytical units distributed across seven thematic sections. Three sections carried most of the weight. Curriculum and CAI Pedagogy had eighty-two units. The Concept of Maturity and Levels had seventy-eight units. Teacher Competencies had seventy-one units. These numbers tell us something simple. The themes appear often in the global and national literature. The reducing phase then consolidated more than eighty candidate dimensions into eight core dimensions and twenty-seven operational constructs, as Table 1 shows.

Table 1. The Eight Core Dimensions and Twenty-Seven Constructs of the MaturiKKA-SD Model

Code	Dimension	Constructs	n
D1	Leadership and Governance	Strategic Vision; Resource Commitment; Change Management; Program M&E	4
D2	Curriculum and CAI Pedagogy	Curriculum Planning; Computational Thinking; Digital Literacy; AI Literacy and Ethics; AI Utilization and Development; Pedagogical Strategies	6
D3	Teacher Competence and Professional Development	AI and Coding Content Knowledge; CAI Pedagogical Competence; Continuing Professional Development; Self-Efficacy and Innovation Orientation	4
D4	Infrastructure and Technology Resources	Hardware and Software Availability; Internet Connectivity; Learning Platforms; Data Security and Privacy	4
D5	AI Literacy and Ethics in CAI Use	Conceptual AI Understanding; Critical Use and Application; Ethical and Responsible AI Use	3
D6	Assessment and Quality Assurance	Student Assessment in CAI; Implementation M&E; Data Use for Improvement	3
D7	Digital Culture and School Climate	Innovation Climate; Digital Collaboration; Digital Citizenship	3
D8	Partnership and External Support	External Partnerships; Government and Regulatory Support; Sustainable Financing	3

Third, the inferring phase produced a systematic comparison between the international literature and the Indonesian policy context. The comparison surfaced twenty-nine critical gaps distributed across the eight dimensions. The Curriculum and CAI Pedagogy dimension (D2) and the Teacher Competence dimension (D3) received the highest priority because they sit at the centre of the substantive and human-resource pillars of implementation. The gaps were not only quantitative. They fell into six typological categories, namely instrument gap, standard gap, content gap, depth gap, contextual gap, and ecosystem gap. Each category implied a different demand on the instrument under development. The instrument gap in particular established the rationale for this study, since no operational instrument existed for diagnosing CAI implementation maturity in Indonesian primary schools.

Fourth, the narrating phase synthesized these outputs into the MaturiKKA-SD conceptual model. The model uses a five-level maturity continuum that runs from Initial (Level 1) through Developing (Level 2), Structured (Level 3), and Integrated (Level 4) to Optimized (Level 5). This five-level structure is consistent with the convergent architecture observed across the global maturity frameworks we reviewed. It also accommodates the heterogeneity of Indonesian primary schools by treating readiness as a developmental continuum rather than a binary condition. The architecture therefore brings together three sources of authority. These are the global maturity literature, the Indonesian national policy framework (specifically the four learning outcomes of CAI in Phase C and the five

essential conditions stipulated by Kemendikdasmen, (2025), and the principles of human-centred AI from Hartikainen et al. (2023). Table 2 presents the conceptual configuration.

Table 2. The Conceptual Architecture of the MaturiKKA-SD Model

Functional Tier	Dimensions
Systemic Prerequisites	D1 — Leadership and Governance D4 — Infrastructure and Technology Resources
Substantive Core	D2 — Curriculum and CAI Pedagogy D3 — Teacher Competence D5 — AI Literacy and Ethics
Feedback Circuit	D6 — Assessment and Quality Assurance
Socio-Cultural Foundation	D7 — Digital Culture and School Climate
Supporting Ecosystem	D8 — Partnership and External Support

Note: The five functional tiers represent the role each dimension plays in the ecological model of CAI implementation; the tiers are interdependent rather than hierarchical.

Two principles guided the instrument design that followed. The principle of multidimensionality came first. CAI implementation is too complex to be captured by a single composite score, so measurement had to be differentiated at the dimension level. The principle of multi-respondent observation came second. No single school actor can see across all eight dimensions. Principals are best placed to observe leadership, infrastructure, climate, and partnerships. Deputy heads for curriculum can observe planning and quality assurance. Class teachers can observe classroom-level pedagogy, teacher competence, and AI literacy in practice. We distributed the items across the three respondent forms accordingly, as Table 3 shows.

Table 3. Distribution of MaturiKKA-SD Version 1 Items by Respondent

Respondent	Dimensions Covered	Items
Principal	D1, D4, D6 (program M&E), D7, D8	22
Deputy Head for Curriculum	D2 (planning), D6 (implementation M&E), D7 (collaboration), D1 (governance)	12
Grade V/VI Class Teacher	D2 (instruction), D3, D5, D6 (student assessment), D7 (classroom climate)	25
Total	8 dimensions across the three forms	59

B. Instrument Design Refinement through the Mini Focus Group Discussion

The mini focus group discussion engaged four experts who looked at the draft instrument from distinct but complementary angles. The discussion ran in two online sessions on 29 April 2026, working through the draft option by option with the instrument shared on screen. Their combined judgment confirmed the conceptual architecture of MaturiKKA-SD. The eight dimensions, twenty-seven constructs, five-level maturity continuum, and multi-respondent design were all retained without structural changes. What the discussion contributed was at the operational layer, namely descriptors, terminology, and supporting components such as the scoring guide and the glossary. Fig 2 reproduces three sample items in their pre- and post-discussion form and shows what this operational refinement looked like in practice.

Version Prototype (pre-FGD draft)	Version 1 (post-FGD)
Exemplar 1 - Principal Instrument, Item 1, Option a - Construct KTG-1 (CAI Vision & Strategic Policy) Revision category: splitting of compound descriptors - Source: Expert N1	
"Sekolah belum memiliki dokumen visi atau kebijakan yang menyebut KKA; kepala sekolah baru mengetahui keberadaan regulasi KKA." <i>"The school has no vision or policy document mentioning CAI; the principal has only just become aware that CAI regulation exists."</i>	"Sekolah belum memiliki dokumen visi atau kebijakan yang mencantumkan program KKA dalam bentuk apa pun." <i>"The school has no vision or policy document that includes a CAI programme in any form."</i>
Revision action. Two distinct indicators were bundled in a single response option: the existence of a policy document and the principal's personal awareness of the regulation. The revised option retains one observable institutional behaviour; individual awareness was relocated to the construct where it belongs.	
Exemplar 2 - Principal Instrument, Item 10, Option d - Construct IST-2 (Connectivity) Revision category: terminological accessibility - Source: Expert N1	
"Konektivitas internet bersifat sporadis dan tidak dapat diakses secara real time." <i>"Internet connectivity is sporadic and cannot be accessed in real time."</i>	"Konektivitas internet tidak menentu dan tidak dapat diakses secara langsung." <i>"Internet connectivity is unreliable and cannot be accessed directly."</i>
Revision action. English loan terms (sporadic, real time) were replaced with Indonesian operational equivalents, and both were subsequently entered into the operational glossary appended to the form.	
Exemplar 3 - Teacher Instrument, new Item 8 (drafted as 7a) - Construct KPK-6 (Pedagogical Strategies) Revision category: addition of a new item / plugged-unplugged equivalence - Source: Expert N4	
— Tidak terdapat butir yang mengukur penggunaan media unplugged ketika perangkat teknologi tidak tersedia. — <i>No item measured the use of unplugged media when technological devices are unavailable.</i>	"Saya menggunakan media unplugged (kartu, papan tulis, atau permainan) untuk mengajarkan konsep KKA ketika perangkat teknologi tidak tersedia." <i>"I use unplugged media (cards, a whiteboard, or games) to teach CAI concepts when technological devices are unavailable."</i>
Revision action. A measurement gap was identified: schools teaching CAI without devices had no route to demonstrate pedagogical maturity. The new item makes the unplugged approach scorable at every maturity level, removing a systematic bias against low-resource schools.	

Fig 2. Sample MaturiKKA-SD Items Before and After the Mini Focus Group Discussion

Six categories of revision emerged from the thematic analysis. First, forty-three compound descriptors that bundled two or more indicators within a single response option were split into single-indicator options. Exemplar 1 in Fig 2 is representative: one option asked simultaneously about the existence of a policy document and about the principal's personal awareness of the regulation, so a respondent whose school satisfied one condition but not the other had no accurate answer available. This sharpened the relevance of each item to the construct it was meant to measure. Second, technical terminology drawn from the global literature (terms like computational thinking, AI literacy, and AI ethics) was complemented by Indonesian-language operational glosses to accommodate the linguistic and conceptual diversity of the respondent pool, as Exemplar 2 illustrates. We appended three operational glossaries to the three respondent forms, containing thirty-four terms in total. Third, the equivalence between plugged and unplugged implementation approaches (which *Permendikdasmen Nomor 13 Tahun 2025* explicitly permits) was made visible in the descriptors of Dimension 2 and Dimension 4. This eliminated a systematic bias that would otherwise have penalized schools using unplugged approaches in their initial implementation phase. Fourth, two new items were added to fill measurement gaps identified during the discussion. One was within construct APM-2 (Implementation M&E). The other was within construct KPK-6 (Pedagogical Strategies), shown as Exemplar 3 in Fig 2. Fifth, three clarification tables explaining the logic of the two-options-per-level structure were added to help respondents navigate the five-level continuum. Sixth, the instructions for completing the instrument were rewritten with attention to register, sequencing, and clarity for users without prior familiarity with maturity assessments.

The configuration of these revisions is informative. All revisions operated at the descriptor, terminology, and supporting-element layers. The eight dimensions and twenty-seven constructs of the conceptual architecture stayed the same. This is the pattern one expects from a DDR-anchored development cycle. Stage 1 establishes the conceptual foundation. Stage 2 confirms and refines its operationalization. Stage 3 subjects the refined instrument to psychometric testing (Richey & Klein, 2014). The outcome of Stage 2 was Instrument Version 1 (Fig 3). It comprised fifty-nine items distributed across three respondent forms, four scoring guides, and three operational glossaries. Conceptually it was unchanged from the Stage 1 model. Operationally it had been sharpened.

(a)

G. BUTIR INSTRUMEN — GURU KELAS 5 DAN/ATAU 6 (GURU TIK JIKA ADA)

Petunjuk: Untuk setiap pernyataan, pilihlah SATU opsi (a sampai j) yang paling tepat menggambarkan kondisi sekolah Saudara saat ini. Beri tanda silang (X) atau lingkaran pada huruf pilihan Saudara.

D2 – KURIKULUM & PEDAGOGI KKA (Implementasi di Kelas)

Konstruk: KPK-2: Implementasi Berpikir Komputasional

1. Saya mengimplementasikan pembelajaran Berpikir Komputasional (dekomposisi, pengenalan pola, abstraksi, perancangan algoritma) secara terencana dan sistematis dalam kegiatan KKA di kelas.

Pilihlah SATU pernyataan di bawah ini yang paling tepat menggambarkan kondisi sekolah Saudara saat ini:

☐	a.	Saya belum pernah mengajarkan berpikir komputasional dalam pembelajaran apapun.
☐	b.	Saya tidak memahami konsep berpikir komputasional yang perlu diajarkan kepada murid.
☐	c.	Saya mencoba mengajarkan satu elemen berpikir komputasional (misalnya algoritma) secara tidak menentu.
☐	d.	Saya mengajarkan 1–2 elemen berpikir komputasional, belum sistematis dan tidak rutin.
☐	e.	Saya mengajarkan 2–3 elemen berpikir komputasional secara terencana dalam beberapa pertemuan KKA.
☐	f.	Saya mengajarkan keempat elemen berpikir komputasional secara terencana dengan modul ajar yang memuatnya.
☐	g.	Seluruh elemen berpikir komputasional saya ajarkan secara sistematis dan progresif dengan murid aktif terlibat.
☐	h.	Pembelajaran berpikir komputasional saya integrasikan dengan konteks kehidupan nyata murid secara bermakna.
☐	i.	Saya mengembangkan aktivitas berpikir komputasional inovatif dan aktif berbagi praktik baik dengan rekan guru.
☐	j.	Saya menjadi guru ahli berpikir komputasional yang menjadi mentor bagi rekan guru dan terus mengembangkan pendekatan.

(b)

LANGKAH 2 — Hitung DSD per Dimensi Menggunakan Rumus Pembobotan

Gunakan rumus berikut untuk masing-masing dimensi:

D1	$DSD-D1 = (0,78 \times RSD-Kasek-D1) + (0,22 \times RSD-Waka-D1)$
D2	$DSD-D2 = (0,57 \times RSD-Guru-D2) + (0,43 \times RSD-Waka-D2)$
D3	$DSD-D3 = 1,00 \times RSD-Guru-D3$
D4	$DSD-D4 = 1,00 \times RSD-Kasek-D4$
D5	$DSD-D5 = 1,00 \times RSD-Guru-D5$
D6	$DSD-D6 = (0,375 \times RSD-Kasek-D6) + (0,25 \times RSD-Waka-D6) + (0,375 \times RSD-Guru-D6)$
D7	$DSD-D7 = (0,50 \times RSD-Kasek-D7) + (0,50 \times RSD-Guru-D7)$
D8	$DSD-D8 = (0,67 \times RSD-Kasek-D8) + (0,33 \times RSD-Waka-D8)$

LANGKAH 3 — Hitung IMKS (Indeks Maturitas KKA Sekolah)

IMKS adalah rata-rata aritmatika dari delapan nilai DSD:

Indeks Maturitas KKA Sekolah (IMKS):

$$IMKS = (DSD_D1 + DSD_D2 + DSD_D3 + DSD_D4 + DSD_D5 + DSD_D6 + DSD_D7 + DSD_D8) \div 8$$

Fig 3. Samples of MaturiKKA-SD Instrument Version 1: (a) teacher-form item with five-level, two-options-per-level maturity descriptors (Construct KPK-2, Dimension 2); (b) cross-respondent weighting and IMKS computation in the integrated school-level scoring guide. The instrument is administered in Indonesian.

C. Content Validation through Expert Judgment

Instrument Version 1 was submitted for content validation to three independent experts representing curriculum development and educational technology, evaluation and educational measurement, and substantive CAI and digital education policy. The three experts were independent of the focus group panel and worked without consultation with one another, satisfying the independence requirement of the Content Validity Index procedure (Yusoff, 2021). Each expert evaluated all fifty-nine items on the four-point relevance scale recommended by Guraya et al. (2022).

The validation produced converging results. All three experts assigned a score of 4 (highly relevant) to every item across the three respondent forms. The Item-level Content Validity Index, computed as the proportion of experts rating each item at 3 or 4, was therefore 1.00 for all fifty-nine items, exceeding the 0.78 threshold of acceptability. The Scale-level Content Validity Index, computed by averaging the I-CVI values across items (S-CVI/Average), was 1.00 for each of the three respondent forms, exceeding the 0.90 threshold defining excellent scale validity. Table 4 summarizes these results.

Table 4. Content Validity Index Results for MaturiKKA-SD Version 1

Respondent Form	Items	Experts at Score ≥ 3	I-CVI	S-CVI/Ave	Decision
MaturiKKA-Principal	22	22 / 22 / 22	1.00	1.00	Content-valid
MaturiKKA-Deputy Head	12	12 / 12 / 12	1.00	1.00	Content-valid
MaturiKKA-Teacher	25	25 / 25 / 25	1.00	1.00	Content-valid
Whole instrument	59	59 / 59 / 59	1.00	1.00	Content-valid

These results need careful interpretation. A perfect S-CVI/Average of 1.00 is uncommon in instrument development and likely reflects one or a combination of three factors. First, the conceptual architecture of MaturiKKA-SD entered Stage 3 already refined through forty-one literature sources in Stage 1 and through expert refinement in Stage 2. That left little for the validators to dispute on substantive relevance grounds. Stefana et al. (2025), in a recent guide to scale development, argue that high CVI scores do not arise automatically from a well-designed instrument. They arise from instruments that have passed through a systematic exploratory phase before validation, so that the items present validators with clear judgmental tasks. Second, the four-point scale without a neutral midpoint, recommended by Guraya et al. (2022) to reduce ambivalence bias, may have encouraged validators to commit to a clear decision in either direction. Third, the panel was made up of senior experts deeply familiar with both the substantive content of CAI and the methodological logic of maturity instruments. Their judgments may therefore have been grounded in the suitability of the conceptual frame rather than in line-by-line scrutiny of phrasing.

A qualitative reading of the validation comments adds context to the quantitative results. Two of the three validators recommended the instrument for use without revision. The third endorsed the instrument with minor revisions. The third validator noted that completing the full instrument on a single sheet might cause respondent fatigue and recommended the development of a digital application to ease access. This comment does not affect content validity in the psychometric sense, but it surfaces a practicality concern that the subsequent usability testing was designed to examine. Yusoff (2021) and Zamanzadeh (2022) explicitly note that the CVI procedure addresses relevance and clarity but does not exhaust the dimensions of instrument quality. Usability and respondent burden are separately testable concerns.

Two minor refinements were made before the instrument moved to Stage 4. The phrasing of two items was lightly adjusted on the basis of the validator's narrative commentary, and the design of a future digital version of the instrument was documented as a development trajectory for subsequent

research. The resulting Instrument Version 2 retained the fifty-nine-item structure and the eight-dimension architecture intact.

D. Limited Usability Testing

The fourth stage tested the usability of Instrument Version 2 in an authentic primary school setting. The site was selected through purposive sampling and corresponded to a public primary school in Bandung that had begun implementing the CAI subject in the 2025/2026 academic year. Three respondents participated: the principal, the deputy head for curriculum, and a Grade V class teacher. The sample size, although small, is consistent with Swacha’s et al. (2025) finding that three to five participants in usability testing surface roughly eighty percent of usability problems for a well-defined instrument, and with the limited-pilot orientation of Stage 4 in the DDR framework.

The findings from the three sessions converged on four themes. The first concerns instructional clarity. All three respondents completed the instrument without external guidance. They reported that the introductory instructions, the explanation of the five maturity levels, and the two-options-per-level logic enabled them to read and respond independently. The clarification tables added at Stage 2 were spontaneously cited by two respondents as helpful for navigating the response architecture. The second theme concerns terminology. The operational glossary cleared up most ambiguities surrounding technical terms such as computational thinking and AI literacy. The deputy head for curriculum did report brief hesitations on three terms (data governance, learning analytics, and AI ethics applied to primary school contexts). These point to opportunities for further glossing.

The third theme concerns differentiation between adjacent maturity levels. Respondents found the contrast between Level 1 (Initial) and Level 5 (Optimized) unambiguous. The contrast between Level 2 (Developing) and Level 3 (Structured) required more deliberation. This finding aligns with the broader literature on maturity-rubric design, which observes that adjacent intermediate levels carry the highest interpretive load on respondents (Ten Hove et al., 2024). We therefore refined two descriptors at the Level 2–3 boundary to sharpen their distinctive markers. The fourth theme concerns completion time and respondent burden. The principal completed the twenty-two-item form in about thirty-five minutes. The deputy head completed the twelve-item form in about twenty minutes. The class teacher completed the twenty-five-item form in about forty-five minutes. All three respondents reported that the completion time was acceptable for an instrument intended for institutional reflection rather than routine survey use. They did recommend that future digital implementations include navigation aids and progress indicators. Table 5 summarizes the usability findings and the corresponding refinements.

Table 5. Summary of Limited Usability Testing Findings and Refinements

Theme	Finding from Limited Usability Test	Refinement
Instructional clarity	Independent completion achieved; clarification tables endorsed by participants	Retained, with minor sequencing improvements
Terminology	Glossary effective for most terms; three terms (data governance, learning analytics, AI ethics at primary level) flagged	Glossary expanded; three additional operational definitions added
Level differentiation	Level 2–3 boundary required deliberation	Descriptors of two items refined for sharper contrast
Completion time and burden	Acceptable in the institutional reflection context; digital aids recommended	Future digital version planned; navigation aids identified for development

The limited usability test produced no findings that called into question the dimensional architecture, the constructs, or the maturity continuum of MaturiKKA-SD. All refinements operated at

the descriptor and supporting-element layers, consistent with the pattern observed in Stage 2. The convergence of refinements across Stages 2 and 4 indicates that the conceptual architecture of the instrument is sufficiently robust to be tested at larger scale in subsequent research. The instrument that emerged from Stage 4, MaturiKKA-SD Version 3, is therefore proposed as ready for broader empirical examination, including reliability testing across multiple schools and the computation of inter-respondent agreement using the Intraclass Correlation Coefficient (Ten Hove et al., 2024).

E. Theoretical and Practical Implications of the MaturiKKA-SD Framework

Three implications follow from the development cycle reported in this article. On the theoretical side, MaturiKKA-SD contributes an original synthesis that brings together global maturity frameworks (Baihaqy & Subriadi, 2023; Begicevic Redjep et al., 2021; Valdivia-Salazar et al., 2025), competency frameworks (Jiménez Sierra et al., 2023; UNESCO, 2023), AI literacy literatures (Ng et al., 2021), and the Indonesian national policy framework into a coherent eight-dimension architecture grounded in twenty-seven operational constructs. The model is original in three senses. It places the four learning outcomes of CAI Phase C as substantive anchors within Dimension 2. It explicitly accommodates the plugged-unplugged duality permitted by national policy. It adopts a multi-respondent design that triangulates the differential observational capacities of school actors (Arias Valencia, 2022).

On the methodological side, the study illustrates how DDR can be operationalized through a sequence of qualitative and psychometric procedures that produce both an instrument and methodological knowledge. The four-stage cycle (content analysis, focus group discussion, expert judgment, and limited usability testing) demonstrates a replicable pathway for developing context-sensitive maturity instruments in other curriculum innovations. The combination of an S-CVI/Average above the conventional threshold and convergent qualitative findings across stages suggests that the early-stage investment in content analysis pays off in later-stage validation.

On the practical side, MaturiKKA-SD addresses a concrete and pressing need. Indonesian primary schools, supervisors, and policymakers have so far lacked an operational instrument to map CAI implementation. MaturiKKA-SD gives them a school-level diagnostic tool that produces a profile across eight dimensions and locates the school on a five-level maturity continuum. The instrument is designed for independent completion by schools without external facilitation. Its output, the School Maturity Profile, supports evidence-informed planning at the school level, program design at the district level, and policy monitoring at the national level.

Several limitations should be acknowledged. The CVI results, while above conventional thresholds, were based on a three-expert panel. Future research could extend validation to a larger panel to refine the reliability of the CVI estimates. The limited usability testing involved three respondents at one school in Bandung. That produced depth but not breadth. Subsequent research should extend usability testing across schools varying in geography, status, and CAI adoption level. The Intraclass Correlation Coefficient for inter-respondent reliability was not estimated in this cycle and remains a priority for the next development cycle. Finally, the digital version of the instrument recommended by one validator and supported by the usability findings is an important next step for widening the practical reach of MaturiKKA-SD.

4. CONCLUSION

This study developed and validated MaturiKKA-SD, a multi-respondent maturity assessment instrument for the implementation of the Coding and Artificial Intelligence subject in Indonesian primary schools. It addresses a gap left open by the rollout of *Permendikdasmen Nomor 13 Tahun 2025*. Through a Design and Development Research cycle of four stages, the study produced an instrument with eight dimensions, twenty-seven operational constructs, fifty-nine items distributed across three respondent forms, and a five-level maturity continuum from Initial to Optimized. Content validation by three independent experts produced an Item-level CVI of 1.00 and a Scale-level CVI Average of 1.00, exceeding the conventional thresholds of acceptability. Limited usability testing at one primary school confirmed that the instrument could be completed independently by its intended users. The test also yielded targeted refinements at the descriptor layer. The resulting MaturiKKA-SD framework integrates global maturity scholarship with the substantive anchors of Indonesian national policy. It offers schools, supervisors, and policymakers a diagnostic tool for evidence-informed planning of CAI implementation at the primary level. Future research should extend reliability testing across multiple schools, estimate

inter-respondent agreement using the Intraclass Correlation Coefficient, and develop a digital implementation of the instrument to broaden its practical reach.

REFERENCES

- Abidin, Y., Herman, T., & Wahyudin, D. (2023). Teachers' understanding of computational thinking and its integration in elementary education: A case study in Indonesia. *Journal of Educational Technology and Curriculum Innovation*, 15(2), 45–58. <https://doi.org/10.31004/jetci.v15i2.425>
- Arias Valencia, M. M. (2022). Principles, scope, and limitations of the methodological triangulation. *Investigación y Educación en Enfermería*, 40(2), e03. <https://doi.org/10.17533/udea.ice.v40n2e03>
- Arriola-Mendoza, J., & Ureña, G. V. (2024). Essential elements for implementing AI tools in elementary school: A systematic literature review. *Education Sciences*, 14(12), 1292. <https://doi.org/10.3390/educsci14121292>
- Awaluddin, A., & Hadi, M. S. (2025). Integrasi pembelajaran coding dan kecerdasan buatan di sekolah dasar: Tantangan dan peluang. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 10(1), 1081–1086. <https://doi.org/10.23969/jp.v10i01.21753>
- Baihaqy, A., & Subriadi, A. P. (2023). Development of digital transformation maturity measurements: A systematic literature review. *Journal of Nonformal Education*, 9(1), 58–68. <https://doi.org/10.15294/jne.v9i1.41985>
- Begicevic Redjep, N., Balaban, I., & Žugec, B. (2021). Assessing digital maturity of schools: Framework and instrument. *Technology, Pedagogy and Education*, 30(5), 643–658. <https://doi.org/10.1080/1475939X.2021.1944291>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10, Article 29. <https://doi.org/10.1186/s40594-023-00418-7>
- Creswell, J.W. and Creswell, J.D. (2023) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Sage Publications Ltd.
- Fitri, S. (2025). The (technology) implementation gap among Indonesian teachers: Understanding the disconnect between personal and professional technology use. *Cambridge Educational Research e-Journal*, 12, 211–225. <https://doi.org/10.17863/CAM.123119>
- Fullan, M., Azorín, C., Harris, A., & Jones, M. (2023). Artificial intelligence and school leadership: Challenges, opportunities and implications. *School Leadership & Management*, 44(4), 339–346. <https://doi.org/10.1080/13632434.2023.2246856>
- Geisen, E., & Bergstrom, J. R. (2017). *Usability testing for survey research*. Morgan Kaufmann. <https://doi.org/10.1016/C2014-0-03467-4>
- Guraya, S. S., Yusoff, M. S. B., Nor, M. Z. M., Fredericks, S., Rashid-Doubell, F., Harkin, D. W., & Guraya, S. Y. (2022). Validating the Medical Education e-Professionalism Framework Using the Content Validity Index. *Education in Medicine Journal*, 14(3)f, 61–74. <https://doi.org/10.21315/eimj2022.14.3.5>
- Hartikainen, M., Väänänen, K., Lehtiö, A., Ala-Luopa, S., & Olsson, T. (2023). Human-centered AI design in reality: A study of developer companies' practices. *International Journal of Human-Computer Studies*, 178, 103077. <https://doi.org/10.1016/j.ijhcs.2023.103077>
- Isrokatun, I., Hanifah, N., Maulana, M., Sujana, A., Nurulhaq, D., & Pratiwi, V. (2023). Teachers' digital competence in the implementation of the Independent Curriculum. *Journal of Education and Learning*, 17(3), 423–432. <https://doi.org/10.11591/edulearn.v17i3.20844>
- Jiménez Sierra, Á. A., Ortega Iglesias, J. M., Cabero-Almenara, J., & Palacios-Rodríguez, A. (2023). Development of the teacher's technological pedagogical content knowledge (TPACK) from the Lesson Study: A systematic review. *Frontiers in Education*, 8, Article 1078913. <https://doi.org/10.3389/educ.2023.1078913>
- Kemendikdasmen. (2025). *Permendikdasmen Nomor 13 Tahun 2025 tentang perubahan atas Permendikbudristek Nomor 12 Tahun 2024 tentang kurikulum pendidikan dasar dan menengah*. Kementerian Pendidikan Dasar dan Menengah Republik Indonesia.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE Publications.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>

- Normén-Smith, T., Hultén, M., & Stigmar, M. (2024). Principals' leadership for digitalisation: A systematic review. *Educational Management Administration & Leadership*, 52(3), 511–530. <https://doi.org/10.1177/17411432231184290>
- OECD. (2025). *Trends shaping education 2025*. OECD Publishing. https://doi.org/10.1787/trends_edu-2025-en
- Pratiwi, D. E., Agustin, N., Suprihatien, S., & Subayani, N. W. (2026). Digital literacy, reading culture, and elementary teachers' readiness for AI and coding: A comparative study of Indonesia and Malaysia. *EduStream: Jurnal Pendidikan Dasar*, 10(1), 19–30. <https://doi.org/10.26740/eds.v10n1.p19-30>
- Richey, R. C., & Klein, J. D. (2014). *Design and development research: Methods, strategies, and issues*. Routledge. <https://doi.org/10.4324/9780203826034>
- Rodrigues, L. F., Oliveira, A., & Rodrigues, H. (2024). Artificial intelligence in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 7, 100296. <https://doi.org/10.1016/j.caeai.2024.100296>
- Selwyn, N. (2022). *Education and technology: Key issues and debates* (3rd ed.). Bloomsbury Academic.
- Stefana, E., Marciano, F., Cocca, P., Rossi, D., & Tomasoni, G. (2025). Scale development in instrument research: A contemporary methodological guide. *Educational and Psychological Measurement*, 85(2), 287–311. <https://doi.org/10.1177/00131644241246712>
- Suharyo, S., Subyantoro, S., & Pristiwati, R. (2024). Artificial intelligence in the context of the Independent Curriculum at primary and secondary levels: Building skills toward Golden Indonesia 2045. *Humanika*, 30(2), 208–217. <https://doi.org/10.14710/humanika.v30i2.60563>
- Sulasmono, B., Friska, Y., Sajiman, S. U., Hasbullah, H., & Purnama, S. (2026). School readiness for coding and AI education: A case study of policy implementation and pedagogical adaptation in Indonesia. *AL-ISHLAH: Jurnal Pendidikan*, 18(1). <https://doi.org/10.35445/alishlah.v18i1.8910>
- Susanto, E., Sugiyanti, W., Arifin, M., Fathonah, S., & Nurkolis, N. (2026). Enacting coding and artificial intelligence education policy in elementary schools. *Journal of Innovation and Research in Primary Education*, 5(1), 1616–1628. <https://doi.org/10.56916/jirpe.v5i1.3135>
- Swacha, J., Radliński, L., Muszyńska, K., Marx, S., & Queirós, R. (2025). Investigating the Effect of Sample Size and Respondent Characteristics on Usability Measurement: The Case of ChatGPT. <https://doi.org/10.62036/ISD.2025.107>
- Ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2024). Interrater reliability for multilevel data: A generalisability theory approach. *Psychological Methods*, 29(2), 354–369. <https://doi.org/10.1037/met0000391>
- Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-based research in the educational field: A systematic literature review. *Education Sciences*, 12(6), Article 410. <https://doi.org/10.3390/educsci12060410>
- UNESCO. (2023). *Technology in education: A tool on whose terms? Global education monitoring report 2023*. UNESCO. <https://doi.org/10.54676/UZQV8501>
- Valdivia-Salazar, C., Serquén, O., Guevara, L., Bravo-Jaico, J., Alarcón, R., Germán, N., ... & Moreno Heredia, A. (2025). Development and validation of an instrument to assess the maturity of digital transformation in higher education institutions. In *Frontiers in Education* (Vol. 10, p. 1691178). Frontiers Media SA. <https://doi.org/10.3389/feduc.2025.1691178>
- Warsina, J. (2025). Measuring the digital competence of primary school teachers in Indonesia. *Jurnal Teknodik*, 29(1), 121–132. <https://doi.org/10.32550/8f0mr443>
- World Economic Forum. (2023). *Future of jobs report 2023*. World Economic Forum.
- Yim, I. H. Y., & Su, J. (2024). A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. *Computers in Human Behavior: Artificial Humans*, 2(2), 100065. <https://doi.org/10.1016/j.chbah.2024.100065>
- Yusoff, M. S. B. (2021). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>
- Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2022). Details of content validity and objectivity of content in instrument development. *Nursing Practice Today*, 9(2), 163–171. <https://doi.org/10.18502/npt.v9i2.9228>
- Zhang, W., Mita, R., & Darko, E. N. K. O. (2026). Preparing teachers for the age of artificial intelligence: Understanding the challenges and needs of AI-TPACK in Indonesian elementary education. *Frontiers in Education*, 11, Article 1769204. <https://doi.org/10.3389/feduc.2026.1769204>
- Zuhri, M., Setiawan, D., & Rahayu, N. (2024). Information communication technologies education in elementary school: A systematic literature review. *EduLearn: Jurnal Pendidikan dan Pembelajaran*, 14(3). <https://edulearn.intelektual.org/index.php/EduLearn/article/download/21435/10485>

Zulaikha, S., Fadholi, M., Sururi, S., Syahril, S., Jamil, S. N., & Ariyanti, P. N. (2025). "Bridging the digital divide": Assessing and advancing teachers' digital literacy across Indonesian provinces. *Journal of Educational Management and Instruction*, 5(1), 195–212. <https://doi.org/10.22515/jemin.v5i1.11586>

Evaluating the Accuracy of ARIMA vs. Holt's Double Exponential Smoothing in Predicting Indonesia's Gold Price Trends

Septi Ika Apriliani¹, Laila Fitriana^{2*}

^{1,2}Department of Mathematics Education, Universitas Sebelas Maret, Surakarta, Indonesia

lailafitriana_fkip@staff.uns.ac.id

ABSTRACT

This study investigates the performance of two forecasting techniques — the Autoregressive Integrated Moving Average (ARIMA) model and Holt's Double Exponential Smoothing — in predicting gold prices in Indonesia. The study uses monthly gold prices for PT Aneka Tambang Tbk from January 2021 to December 2025, yielding 60 data points. The analysis in this study employs a time-series framework that incorporates both ARIMA and Holt's Double Exponential Smoothing. The performance of each model is evaluated using Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The ARIMA (1,1,6) model outperformed Holt's model in historical prediction accuracy, yielding lower error rates. However, the actual recorded gold prices for January and April 2026 showed that Holt's Double Exponential Smoothing achieved better forecasting accuracy. This finding indicates that the ARIMA model is excellent for predicting historical accuracy; however, for recent data shifts, Holt's model is more advantageous for short-term forecasts. The findings of this study indicated that, to achieve forecasting goals, the choice of forecasting model should be tailored to the specific characteristics of the available data.

Keywords: Forecasting, ARIMA, Holt's Double Exponential Smoothing, Gold Prices, Time Series



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Laila Fitriana,
Department of Mathematics Education,
Universitas Sebelas Maret Surakarta,
Jalan Ir. Sutami No. 36A, Surakarta, 57126, Indonesia.
lailafitriana_fkip@staff.uns.ac.id

1. INTRODUCTION

Due to its significance and its history as currency and an investment, the metric for gold is important for assessing the economy. Gold is an asset known for its value during an economic crisis (Lastri, 2021). This is also supported by Su (2024), who indicated that gold is resilient and steady during economic downturns. Gold is also liquid, low risk, and a good long-term investment (Sagala et al., 2023). In the long run, gold's value, even during high inflation, is important for assessing its worth, and that is why the public invests in gold (Mulia, 2024).

Investments in gold, particularly in gold bars sold by PT Aneka Tambang Tbk (Antam), are becoming more widespread in Indonesia. Due to their high quality, certified authenticity, and strong market credibility, Antam gold products have become increasingly popular and accepted as a long-term investment option. Kristhy et al. (2022) explain that the high level of interest in gold investment within the population stems from the convenience of conducting buying and selling transactions — either in person or digitally. Additionally, gold is sought after as a long-term investment because it is stable, as noted by Mulia (2024).

The world economy has been all over the place in the last few years, especially since the COVID-19 outbreak began. Because of this, many people looking to invest are considering gold. Gold is an investment because it tends to hold its value and even appreciate when the economy is bad, and prices are rising (Mulia, 2024). Other people have also said that gold is an investment because its value does not change much, it is easy to buy and sell, and it helps protect against inflation, which is when prices rise too fast. So people like to invest in gold for a time (Sagala et al., 2023). Methods exist for forecasting future economic conditions, including the Autoregressive Integrated Moving Average, abbreviated as ARIMA. These methods are good at analyzing past events and using that analysis to predict future outcomes, like the consumer price index ((Djami & Nanlohy, 2022). Other people have also said that when inflation rises and currencies are not worth much, gold prices tend to go up (Lastri, 2021). Gold is also good because it can handle short-term changes in the economy (Su, 2024).

People are becoming more aware of how important it is to invest. So investment trends in Indonesia are also getting bigger. A survey conducted by Jakpat (2024) found that people are still very interested in investing, even when the economy is not doing well. The survey shows that most people already have investment instruments, and gold and gold-related products are very popular. This is because people think gold is safer and more stable than other investments.

On the one hand, the price of gold in Indonesia, especially gold made by PT Aneka Tambang Tbk, goes up and down over time. If we look at data from logammulia.com, we can see that gold prices rose significantly from 2021 to 2025. There were also some short-term changes due to developments in global markets, exchange rates, and economic policies. There is a graph of gold prices from January 2021 to December 2025 that shows how gold prices changed.



source: www.logammulia.com

Fig 1. Indonesian Gold Purchase Prices for the 2021–2025 Period

The fluctuations in gold prices exhibit complex behavior in the market, which do not depend solely on local factors but are also affected by the international economy. As such, it is essential to employ a rigorous analytical method that can effectively capture trends in gold prices to forecast outcomes that guide decision-making. Although many studies have been conducted on forecasting gold prices using time-series methods, their results show inconsistent precision rates across the techniques used. The ARIMA model is commonly used in scholarly studies due to its effectiveness in capturing autocorrelation and the correlation between past events in the dataset (Chasanah, 2021; Wibowo et al., 2025). Conversely, the method of exponential smoothing, especially Holt's Double Exponential Smoothing, is considered more appropriate for analyzing data with trends because of its straightforwardness and adaptability (Mandati et al., 2025; Yenikar et al., 2025).

Nevertheless, past studies indicate that no single approach guarantees the highest accuracy in forecasting. Numerous studies have found that the ARIMA approach is effective for predicting gold prices, as it captures patterns in autocorrelation and correlations within historical data. As illustrated by Wang (2025), the ARIMA method can produce optimal forecasts with relatively low error rates for gold prices. Additionally, some researchers report that the ARIMA method yields the best results among other approaches when applied to certain gold price predictions (Bing & Nor, 2024). At the same time, various studies show that the exponential smoothing method is advantageous for dynamic data with trends. According to Bhattacharya's (2025) research, it was easier to forecast changes in gold prices in an unstable situation using exponential smoothing due to better adaptability, which results in lower rates of errors compared to those created by the ARIMA approach. Although these studies demonstrate the effectiveness of ARIMA and exponential smoothing methods, the reported findings remain inconsistent because the best-performing model varies depending on the characteristics of the dataset and the evaluation approach used.

There is a research gap, and it is getting wider as more people are interested in buying gold, especially gold bullion from PT Aneka Tambang Tbk. The way people invest is changing, and the market is becoming more complex, so we need a way to analyze it. We need to conduct research using the information and compare different methods to determine the best way to predict gold prices.

The results are different. This shows that the performance of forecasting models depends on the type of data used, such as trends and seasonal patterns. Most studies have problems, such as using global

or outdated data that do not reflect what is happening in the Indonesian gold market. They also do not use data showing what is happening after the pandemic. We need to compare ARIMA and Holt's Double Exponential Smoothing using recent Indonesian gold price data, as studies specifically focusing on monthly PT Aneka Tambang Tbk gold prices and validating the forecasting results with actual future observations remain limited. Furthermore, previous studies generally selected the best forecasting model based only on statistical error measures, such as MAPE, MAE, and RMSE, while limited attention has been given to validating whether the model with the best historical performance also produces the most accurate forecasts when compared with actual future market observations.

Unlike previous studies, this study not only evaluates forecasting performance using statistical error measures such as MAPE, MAE, and RMSE, but also validates the forecasting results using the actual monthly gold prices of PT Aneka Tambang Tbk from January to April 2026. This additional validation provides a more comprehensive assessment of forecasting performance under real market conditions and offers practical insights into selecting an appropriate forecasting model for the Indonesian gold market.

To address these gaps, this study uses monthly gold price data from PT Aneka Tambang Tbk for the 2021–2025 period to compare the forecasting performance of ARIMA and Holt's Double Exponential Smoothing. The forecasting performance is evaluated using MAPE, MAE, and RMSE, followed by validation with actual monthly gold prices from January to April 2026 to provide a forecasting model that is more relevant to current market conditions.

In the situation described, this research aims to compare the forecasting performance of ARIMA and Holt's Double Exponential Smoothing using monthly gold prices of PT Aneka Tambang Tbk from 2021 to 2025. The forecasting performance is evaluated using MAPE, MAE, and RMSE and further validated against the actual monthly gold prices from January to April 2026. The findings are expected to provide a forecasting model that better reflects current market conditions and supports investment decision-making. In time series forecasting, the ARIMA model is considered an efficient technique due to its ability to model relationships between past and current values using autoregressive (AR), integrated (I), and moving-average (MA) components. The ARIMA model starts by identifying the model, estimating parameters, and testing diagnostics before determining the optimal model (Chasanah, 2021; Wibowo et al., 2025). Moreover, the strength of this model lies in its ability to handle non-stationary data sets by differentiating; therefore, it is commonly used for forecasting economic and financial data.

Conversely, Holt's Double Exponential Smoothing is used to account for trends in the dataset. This method requires parameters for both the level (α) and the trend (β), thus allowing the model to be flexible with any changes occurring within the dataset. The advantages of this method include its simplicity and effectiveness in providing accurate predictions for datasets with linear trends (Mandati et al., 2025; Yenkikar et al., 2025). The ultimate objective of this study is to formulate a predictive model that aligns well with current market conditions and provides valuable insights for investment decisions.

2. RESEARCH METHOD

A. Type and Research Approach

This study uses a quantitative, time-series approach. The choice of the quantitative approach is based on the nature of the data, which is numerical and is subjected to statistical analysis to develop a forecast model. The process makes it easy to objectively and quantitatively analyze the past data trends and predict the future values (Wibowo et al., 2025; Yenkikar et al., 2025). The use of the time series approach is justified, as the data used in the research consist of numerical gold prices arranged in sequential order, particularly from 2021 to 2025. Time series analysis focuses on recognizing distinct patterns within the data, such as trends and variations, which serve as a foundation for making predictions (Hyndman & Athanasopoulos, 2021).

The techniques used in this research are the ARIMA and Holt's double-exponential smoothing models. They have been selected based on their relevance to analyzing time series and the different approaches used to model the information. While ARIMA focuses on the statistical relationships in past data, the second approach concerns data smoothing and trend adjustment (Wibowo et al., 2025; Yenkikar et al., 2025). This research is quantitative, and its main goal is to develop a forecasting model

to accurately predict gold prices in Indonesia using data from PT Aneka Tambang Tbk from 2021 to 2025.

B. Data Sources and Types

The information utilized in this research is secondary, as evidenced by the gold bar prices sourced from PT Aneka Tambang Tbk through the Logam Mulia site. The dataset comprises monthly gold prices ranging from January 2021 to December 2025, featuring a total of 60 observations. The information employed in this study is quantitative, consisting of time-series data organized in sequential order to identify any changes in the trend of gold prices (Chatfield, 2022; Montgomery et al., 2021).

C. Data Collection Techniques

The data collection procedure used in this study is the documentation method, which involves collecting, recording, and managing information that already exists, either in writing or in archival form, from certain sources (Creswell & Creswell, 2022; Sugiyono, 2022). The documentation method is chosen because the information used in this study is in secondary form and has been officially made available to everyone. Data collected includes gold bar prices from PT Aneka Tambang Tbk, through the Logam Mulia website. The procedure that was taken to collect data includes accessing the price data history of gold bars on the website and recording them based on the time period in this study, that is, January 2021 until December 2025. The data used in this study include monthly data collected on the 10th day of each month (Hyndman & Athanasopoulos, 2021).

D. Data Analysis Methods

The methodology used in this study involves adopting a time-series forecasting approach using two models: the Autoregressive Integrated Moving Average (ARIMA) and Holt's Double Exponential Smoothing. The purpose is to determine which of the two methods produces the most accurate forecasts of gold prices in Indonesia. The procedure adopted in this study was to use EViews 12 to analyze the data, including model identification, parameter estimation, diagnostic testing, and forecasting. In this study, time-series analysis was conducted using monthly gold prices in Indonesia from January 2021 to December 2025. Thereafter, a descriptive analysis was conducted to examine the data characteristics.

1) Autoregressive Integrated Moving Average (ARIMA) Method

The Autoregressive Integrated Moving Average (ARIMA) model is one of the most commonly used approaches for predicting time-series data, thanks to its ability to detect autocorrelation patterns in historical data and handle non-stationary data through differencing (Bhattacharya, 2025; Wibowo et al., 2025). An ARIMA model is composed of three elements: autoregressive (AR), integrated (I), and moving average (MA). The concept of autoregressive (AR) suggests that the values from earlier periods have an effect on the present value of a variable. Concurrently, moving average (MA) indicates that the present value of a variable is affected by the errors from the preceding period. The ARIMA model is written as ARIMA(p,d,q), where:

- p is the AR order
- d is the difference order
- q is the MA order

For time series that are not stationary, the ARIMA model takes an additional step of differencing to make the series stationary (Chatfield, 2022). In general, the ARIMA model may be expressed as shown below (Hyndman & Athanasopoulos, 2021):

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t$$

First-order differencing can be described by the following process (Hyndman & Athanasopoulos, 2021):

$$\Delta Y_t = Y_t - Y_{t-1}$$

Once first-order differencing is completed, the development of the ARMA model can take place into an ARIMA model. The ARIMA model can be described using the following formula for first-order differencing:

$$\Delta Y_t = \phi + \sum_{i=1}^p \theta_i \Delta Y_{t-i} + \varepsilon_t - \sum_{i=1}^q \lambda_i \varepsilon_{t-i}$$

The approach is used to transform the non-stationary time series data into stationary data through the process of differencing.

From the model stated above, the steps involved in carrying out the ARIMA analysis in this study include:

- a. **Series Stationarity Test**
To determine whether the data on gold prices are stationary, a series stationarity test is conducted. This refers to data which has consistent variance and average at all times. The procedure uses the Augmented Dickey-Fuller (ADF) test, where the data will be considered stationary if the p-value is less than 0.05 (Hyndman & Athanasopoulos, 2021).
- b. **Differencing**
If the data is not stable over time, a differencing procedure is applied to eliminate the trend. This is done by taking the current data point and subtracting the value from the earlier period, specifically: $\Delta Y_t = Y_t - Y_{t-1}$
This process can be performed more than once until the data becomes stationary (Hyndman & Athanasopoulos, 2021).
- c. **Model Identification (ACF and PACF)**
Once the data was verified to be stable, the following action was to establish the model by employing Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) diagrams. The ACF diagram helped in recognizing the order of the moving average (q), while the PACF diagram was utilized to ascertain the order of autoregression (p). The *cut-off* and *tailing-off* patterns in the plots were used as the basis for selecting the model (Chatfield, 2022).
- d. **Determination of the ARIMA (p,d,q) Model**
According to the ACF and PACF evaluations, multiple potential ARIMA models, represented as ARIMA(p,d,q), were identified. The model chosen is the one that most accurately aligns with the established data pattern (Montgomery et al., 2021).
- e. **Parameter Estimation**
Parameter estimation was performed to obtain the optimal parameters for the ARIMA model using EViews 12. When the probability value of a parameter, also referred to as the p-value, is less than 0.05, it means that the parameter is statistically significant. When a model has significant parameters, it implies that the variable does impact the model, enabling the model to statistically explain the relationship between the data variables (Hyndman & Athanasopoulos, 2021).
Additionally, selecting the best model depends on the values of the Akaike Information Criteria (AIC) and the Schwarz Criteria (SC). The best model will be the one having the lowest values of both the AIC and SC (Montgomery et al., 2021).
- f. **Diagnostic Tests**
Diagnostic tests were conducted to assess whether the model residuals are white noise and uncorrelated. These assessments were conducted utilizing the residual correlogram (Q-statistics) or the Ljung-Box test, which are formulated to identify the existence of autocorrelation within the residuals. The hypotheses used are as follows:
- H0: the residuals have no autocorrelation (*white noise*)
- H1: The residuals exhibit autocorrelation
A determination is reached according to the probability value (p-value), specifically:
- p-value > 0.05 → the residuals are not *autocorrelated* (the model is suitable for use)
- p-value ≤ 0.05 → the residuals are *autocorrelated* (the model needs to be improved)
Additionally, model stability is assessed via *inverse root* analysis, where the model is deemed stable if all values fall within the permissible range (Montgomery et al., 2021). Should the model fail to satisfy these criteria, the selection process for models will be conducted once again until the model is identified that most appropriately aligns with the characteristics of the data.

2) **Holt's Double Exponential Smoothing Method**

The Double Exponential Smoothing technique proposed by Holt is utilized in this research as a benchmark method against ARIMA for forecasting gold prices in Indonesia. This method is an enhancement of Single Exponential Smoothing, which is capable of capturing trend patterns within

data. Consequently, it is deemed suitable for predicting gold price data that exhibits a trend of rising prices and variable patterns. There have been several studies that demonstrate the effectiveness of this method not only in predicting gold prices but also in other goods that exhibit linear trends (Ariffin et al., 2024; Mandati et al., 2025; Zheng, 2025).

This method uses two elements, which are the level element and the trend element. The level element represents the basic value of data during a certain period of time, while the trend element indicates the direction of changes in data for a certain period of time. The combination of both elements enables this method to provide more accurate predictions on data that show linear trends. According to Kahraman (2022), the exponential smoothing method is efficient for predicting the prices of metal commodities due to its adaptive nature to trend changes.

Mathematically, *Holt's Double Exponential Smoothing* model can be written as follows. (Hyndman & Athanasopoulos, 2021):

A. *Level Equation*

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

B. *Trend Equation*

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

C. *Forecasting Equation*

$$\hat{Y}_{t+m} = L_t + mT_t$$

where:

- L_t = The *level* component in period t
- T_t = *trend* component in period t
- Y_t = actual data for period t
- α = *level smoothing* parameter ($0 < \alpha \leq 1$)
- β = *trend smoothing* parameter ($0 < \beta \leq 1$)
- $\hat{Y}_{t+m}$ = forecast values for the next m periods

The values of the parameters α and β are crucial for forecasting results, as they control the degree of data smoothing. Optimal parameter values will result in a lower forecasting error rate. According to this model, the steps for analysis employing Holt's Double Exponential Smoothing technique in this research consist of the following:

a. Finding patterns in the data

The first step is to use time-series plots to identify patterns in gold price data. The data demonstrated distinct trend patterns, which made the Holt method appropriate for use (Su, 2024).

b. Calculating the smoothing parameters (α and β)

An ideal model is obtained by determining the smoothing parameters α and β , where α smooths the level component and β smooths the trend component. Parameter determination is performed using *EViews 12* software, which automatically searches for the best parameter combination based on the smallest forecasting error. Determining optimal parameters is crucial because it directly affects forecast accuracy (Zheng, 2025).

c. Determining the trend and level components

The Holt's method is applied to calculate the level and trend components once the parameters have been estimated. The trend component indicates the historical change in the data, while the level component reveals the forecasted base value of the data. These calculations are iteratively repeated for each period to allow the model to adapt to changes in data patterns and produce more accurate predictions (Mandati et al., 2025).

d. *Forecasting Process*

Forecasting is done for the coming period based on the level and trend obtained from the equations. Predictions are made about the price of gold in the near future based on the forecasts. The Holt approach can be used effectively for data showing a linear trend since it is flexible enough to capture any change in patterns (Bhattacharya, 2025).

e. Selection of the Best Model

The model that is chosen should have the minimum forecasting error. The model that uses the Holt's approach with the optimal values of α and β that yield the lowest forecasting error is considered the best model.

The most effective method for predicting the gold price in Indonesia is determined by comparing the forecasting results from the Holt's Double Exponential Smoothing and ARIMA approaches using measures such as MAPE, MAE, and RMSE (Pirmanto, 2024).

E. Measures of Accuracy

The evaluation of forecast inaccuracies takes place after acquiring the best models by utilizing the Autoregressive Integrated Moving Average (ARIMA) and the Double Exponential Smoothing techniques. The purpose of this measurement is to ascertain which approach yields the most precise forecasts. Several error measures (forecast error) are used to compare the forecast values with the actual data in order to do the evaluation. Because they can quantitatively show the model's accuracy, the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are frequently used in forecasting research (Ariffin et al., 2024; Pirmanto, 2024).

a. Mean Absolute Percentage Error (MAPE)

The difference in percentages between actual data and forecasted values is measured using the MAPE. The MAPE number makes it simpler to understand by expressing the size of the model's mistake as a percentage. The following is the MAPE formula. (Hyndman & Athanasopoulos, 2021):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%$$

b. Mean Absolute Error (MAE)

The average absolute value of the discrepancy between actual data and anticipated outcomes is measured using MAE. Unlike MSE, which penalizes extreme errors more severely, this indicator accurately depicts the size of prediction errors. MAE is frequently used to compare the accuracy of various forecasting models because it is easier to interpret, as it uses the same units as the original data (Montgomery et al., 2021).

The following is the MAE formula. (Montgomery et al., 2021):

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

The smaller the MAE value, the better the forecasting model is considered to be, as it indicates a lower average prediction error.

c. RMSE, or root mean squared error

The level of prediction inaccuracy, expressed in the same units as the original data, is assessed using RMSE, which is derived from the square root of MSE. Since RMSE is presented in the same units as the actual measurements, it is easier to interpret compared to MSE. The RMSE formula is as follows (Chatfield, 2022):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

A lower RMSE indicates a lower prediction error and better accuracy.

d. Determining the Best Model

The best model in this study was determined based on the smallest MAPE, MAE, and RMSE values. To find out which approach resulted in the lowest forecasting error rate, the ARIMA and Holt's Double Exponential Smoothing models were compared. Because it was thought to have more accurate prediction abilities for predicting gold prices in Indonesia, the model with the lowest error value was chosen as the best model. These results will subsequently form the basis for the research conclusions.

F. Research Flowchart

In order to understand the various stages of research conducted, a research flow chart was formulated. This flow chart shows how the research process was carried out systematically, including gathering

data, formulating models using ARIMA and Holt's Double Exponential Smoothing approaches, and choosing the best model through forecasting accuracy.

The following is the flow chart that was used for this research:

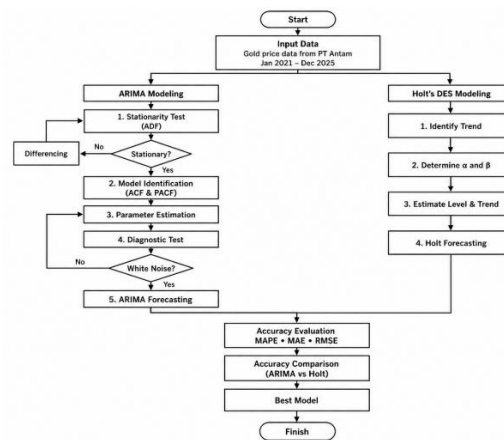


Fig 2. Research Flowchart

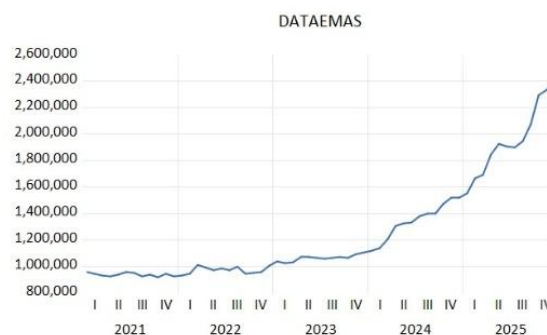
3. RESULTS AND DISCUSSION

A. Description of Research Data

The price of gold from PT Aneka Tambang Tbk (Antam) collected using Logam Mulia from the company's official website serves as the source of the data in this paper. Specifically, the price of gold bars per gram collected monthly, from January 2021 to December 2025, will be analyzed in this research. Monthly data is chosen because it allows for better analysis of trends in gold prices' medium and long-term variations, thus making it more suitable for time series forecasting purposes. The purpose of collecting the data on a specific day each month is to mitigate biases caused by extreme fluctuations in gold prices each day.

To provide an overview of the dataset, descriptive statistics were calculated for the gold price data covering the period from 2021 to 2025. The results show that the mean gold price was **1,255,500**, while the median was **1,063,000**, indicating that the average price was influenced by relatively higher gold prices during the observation period. The gold prices ranged from a minimum value of **914,000** to a maximum value of **2,416,000**, reflecting substantial price variation over the five-year period. Furthermore, the standard deviation of **406,790.1** indicates considerable variability in monthly gold prices, suggesting that the data experienced noticeable fluctuations around the mean.

Based on the results obtained in the time series plot, there are price variations in gold prices that have an upward tendency sometimes. Fluctuations in gold prices are affected by factors such as global interest rates, inflation, rupiah exchange rate, and economic uncertainty (Ariffin et al., 2024; Zheng, 2025). Uncertain economies are likely to increase demand for gold since it acts as a safe haven in case of uncertainties. ARIMA method and Holt's double exponential smoothing are chosen for this research due to the characteristics exhibited by the data, which include pattern and variation.



source : www.logammulia.com

Fig 3. PT Antam Tbk Gold Price Chart for the 2021–2025 Period

B. ARIMA Modeling Results

1) Stationarity Test

In order to check whether the gold prices data fits in the stationarity assumption for the model, the initial step in the ARIMA Modeling is performing a stationarity test. Since ARIMA modeling assumes that the mean and the variance remain constant in time series data, it becomes a basic requirement. For conducting the stationarity test, the Augmented Dickey Fuller (ADF) Test has been performed through the use of the software package, EViews 12. The following theories were proposed:

H0: Non-stationary data

H1: The information is stationary

The probability value (p-value) is used to make the test decision; if the p-value is less than 0.05, the data are deemed stationary.

Table 1. ADF Test Results at the *Level*

	t-statistic	Prob.
<i>Augmented Dickey-Fuller test statistic</i>	5.023249	1.0000
Test critical values:		
1% level	-3.546099	
5% level	-2.911730	
10% level	-2.593551	

The ADF test findings at the specified level yielded a probability value of 1.0000, which is greater than 0.05, according to Table 1. This shows that the gold price data still exhibit a unit root and are not stationary at the level. To attain stationarity, a differencing procedure is necessary, as the null hypothesis (H0) is accepted. First-order differencing, or first difference, is used since the data is not stationary at the level. The data is then deemed stationary at the initial difference when the ADF test is performed once again and the results indicate that the probability value is less than 0.05.

Table 2. ADF Test at the *First Difference*

	t-statistic	Prob.
<i>Augmented Dickey-Fuller test statistic</i>	-5.008320	0.0001
Test critical values:		
1% level	-3.548208	
5% level	-2.912631	
10% level	-2.594027	

Thus, based on the results in Table 2, the ARIMA model used has a *differencing* value of one, or $d = 1$.

2) ARIMA Model Identification (ACF and PACF)

Once the data has been established as stationary following the first difference, the subsequent phase involves determining the ARIMA model by utilizing graphical assessments of ACF and PACF.

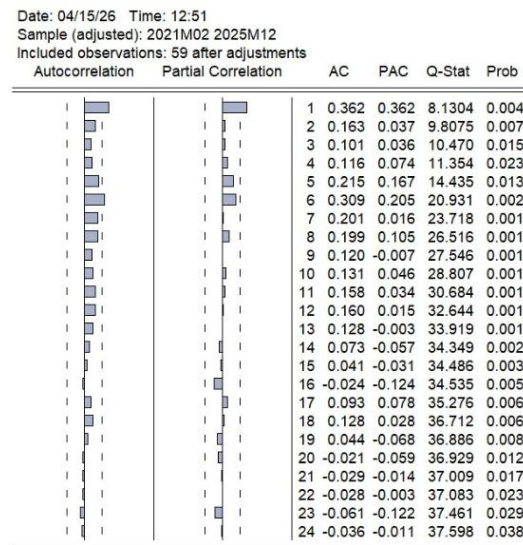


Fig 4. Results of ACF and PACF Analysis

Additionally, the findings of the ACF and PACF studies in Figure 4 demonstrate that the ACF exhibits prominent values at lags 1 and 6, whereas the PACF has a prominent value at lag 1. Thus, ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(0,1,6), ARIMA(1,1,0), ARIMA(1,1,1), and ARIMA(1,1,6) are the ARIMA model candidates examined in this study. By comparing parameter significance, Akaike Information Criterion (AIC) values, Schwarz Criterion (SC) values, and the outcomes of residual diagnostic tests, the optimal model was chosen. The optimal ARIMA model was determined to have significant parameters, the lowest AIC and SC values, and residuals that met the white noise condition.

3) ARIMA Model Parameter Estimation

The next stage is to estimate the parameters to find the model that best fits the features of the gold price data after identifying a number of candidate ARIMA models. The probability value (p-value) for each parameter is used to test each model. If a parameter's probability value is less than 0.05, it is deemed significant. Furthermore, the model that fits the data better is thought to have lower AIC and SC values.

Table 3. ARIMA Model Parameter Estimation Results

Model	AIC	SC
ARIMA(0,1,0)	24,37068	24,40589
ARIMA(0,1,1)	24,31524	24,42088
ARIMA(0,1,6)	24,27486	24,41571
ARIMA(1,1,0)	24,2950	24,40064
ARIMA(1,1,1)	24,27387	24,41472
ARIMA(1,1,6)	24,19138	24,36744

The ARIMA(1,1,6) model has the lowest Akaike Information Criterion (AIC) value of 24.19138 and Schwarz Criterion (SC) value of 24.36744 among the candidate models, according to Table 3's estimation results. A better match between the model and the data is shown by smaller AIC and SC values. As a result, the ARIMA(1,1,6) model is chosen as the best option for the time being and will proceed to the residual diagnostic test and parameter significance test.

4) Parameter Significance Test

After the ARIMA(1,1,6) model was selected as the best candidate, the next step was to conduct a parameter significance test to determine whether the model's coefficients had a statistically significant effect. The significance test was conducted by examining the probability value (*p-value*) for each parameter. A parameter was considered significant if the probability value was less than 0.05.

Table 4. ARIMA Model Parameter Estimation Results

Variable	Coefficient	Standard Error	Prob.	Result
----------	-------------	----------------	-------	--------

C	35376,52	50192,46	0,4840	Not Significant
AR(1)	0,984319	0,041694	0,0000	Significant
MA(1)	-1,060710	0,130107	0,0000	Significant
MA(6)	0,250120	0,163318	0,1315	Not Significant

Table 4 shows that the calculated values for the AR(1) and MA(1) parameters have a probability value of 0.0000, making them significant. In contrast, the MA(6) parameter has a p-value of 0.1315, which is deemed inconsequential because it exceeds 0.05. Furthermore, with a p-value of 0.4840, the model constant is likewise not significant. The ARIMA(1,1,6) model is nevertheless considered despite having one irrelevant parameter because it has more significant parameters than the other candidate models and the lowest AIC and SC values.

5) Residual Diagnostic Tests

To make sure that the model's residuals are random (white noise) and do not show autocorrelation, a residual diagnostic test is then carried out. The Correlogram Q-Statistics (LjungBox Test) is used for this test. If the Q-Statistic probability value is higher than 0.05, the model is considered to meet the white noise assumption.

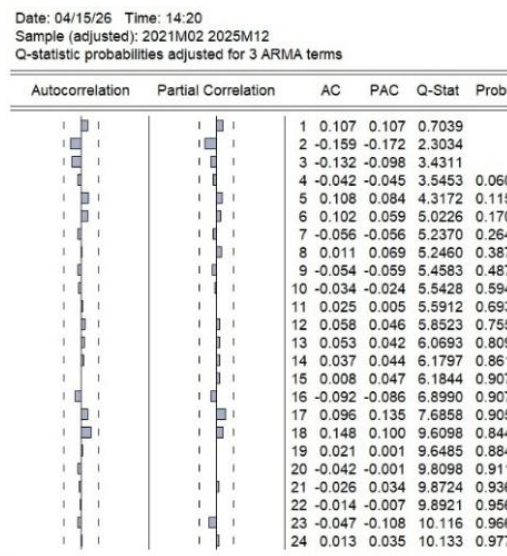


Fig 5. Results of the ARIMA (1,1,6) Residual Diagnostic Test

The ARIMA(1,1,6) model displays a probability value of 0.06, which is higher than 0.05, according to Figure 5. These findings show that the residuals of the model satisfy the white noise assumption and do not show autocorrelation.

6) ARIMA Model Forecasting

The gold price forecasting procedure is carried out after the ARIMA(1,1,6) model was determined to be the best model based on AIC and SC values, parameter significance tests, and residual diagnostic tests.

Table 5. Gold Price Forecasting Results Using ARIMA

Date	Price(Rp/gram)
10/1/2026	2.459.447
10/2/2026	2.397.312
10/3/2026	2.552.297
10/4/2026	2.642.508

The ARIMA(1,1,6) model shows good performance in following the patterns of gold price movement, according to the data in Table 5. However, these forecasting results must be compared with Holt's Double Exponential Smoothing utilizing accuracy metrics like MAPE, MAE, and RMSE in order to identify the more accurate approach.

C. Results of Holt's Double Exponential Smoothing Modeling

The next stage is to model using Holt's Double Exponential Smoothing as a comparative approach to predict gold prices in Indonesia, after obtaining the forecasting results from the ARIMA method. This approach is applied to time series data with a trend pattern since the gold price data shows a fluctuating pattern with an increasing tendency. The analysis steps for this method include identifying the data pattern, determining the *smoothing* parameters α (α) and β (β), calculating *the level* and *trend*, and the *forecasting* process. From the time series graph in Subsection 3.1, it can be observed that the prices of gold have an oscillating pattern with an increasing trend from 2021 to 2025. Hence, it can be inferred that the data is sufficiently trending; thus, the use of Holt's Double Exponential Smoothing method is suitable.

The subsequent phase involves determining the values for the two smoothing factors, α (α) and β (β). The α factor is utilized to adjust the level aspect, whereas the β factor is employed to refine the trend aspect. A better forecasting model requires optimal values for the smoothing constants. In this study, the best results from the Holt model estimation are used to select the α and β parameters. The Holt-Winters No Seasonal model, which is similar to Holt's Double Exponential Smoothing because it considers only the level and trend components without seasonality, was used to determine the smoothing parameters for Holt's Double Exponential Smoothing in EViews 12.

The α (α) and β (β) parameters were found to have a value of 1.0000 and 0.1500, respectively, according to the results of the estimation process. Although the small value of the β parameter implies that the trend factor will be smoothed continuously, the α parameter of value one means that the model responds sensitively to variations in recent data. The technique used to forecast gold prices for January to April 2026 is Holt's Double Exponential Smoothing.

Table 6. Gold Price Forecasting Results Using Holt's Double Exponential Smoothing

Date	Price(Rp/gram)
10/1/2026	2.490.342
10/2/2026	2.564.685
10/3/2026	2.639.027
10/4/2026	2.713.369

D. Comparison of the Accuracy of the ARIMA Model and Holt's Double Exponential Smoothing

To identify the most effective model for predicting gold prices in Indonesia, a comparison was made between the forecast accuracy of the ARIMA(1,1,6) model and Holt's Double Exponential Smoothing model. Forecasting accuracy comparisons were conducted using forecast error measures: MAPE, MAE, and RMSE. The three measures were used to evaluate the accuracy of the forecasts generated by each model. Those models that had relatively smaller values of MAPE, MAE, and RMSE were regarded as having high accuracy.

Table 7. Comparing the ARIMA and Holt's Double Exponential Smoothing Models' Accuracy

Method	MAPE	MAE	RMSE
ARIMA(1,1,6)	1,98	26,227	37.926,51
Holt DES	2,90	57,438	41.367,74

According to Table 7, the ARIMA(1,1,6) model shows greater accuracy in relation to the Holt's Double Exponential Smoothing approach due to having lesser MAPE, MAE, and RMSE values. The superior performance of the ARIMA(1,1,6) model can be explained by the characteristics of the historical gold price series used in this study. As shown in Figure 1, during the 2021–2025 period, the monthly gold prices of PT Aneka Tambang Tbk generally exhibited an increasing trend, although several temporary fluctuations occurred. After the differencing process, the data became stationary, allowing the ARIMA model to effectively capture the autocorrelation structure and temporal dependence within the historical observations. These characteristics are consistent with the strengths of ARIMA in modelling non-stationary time series after differencing, which contributed to its lower

forecasting errors. This finding is consistent with Wang (2025) and Bing and Nor (2024), who reported that ARIMA performs well when historical time series exhibit stable temporal dependence and can be transformed into stationary data through differencing. Therefore, it is clear that the ARIMA model is more accurate, as evidenced by its historical performance. Accuracy of forecasting in this study was verified based on the comparison of forecast values to actual gold price values for the period from January to April 2026, along with the use of accuracy measurements such as MAPE, MAE, and RMSE. As gold prices for January through April 2026 were known, verification was needed to determine which forecasts were most accurate. Forecast values made using both the ARIMA and Holt approaches were compared to actual gold price values from PT Antam Tbk. The one which had the least difference to the actual value was considered the most accurate one for making forecasts about gold prices.

Table 8. Comparison of Forecasts and Actual Data for 2026

Month	Actual	ARIMA Forecast	Holt Forecast	Nearest Neighbor Method
January	2.602.000	2.459.447	2.490.342	Holt
February	2.954.000	2.397.312	2.564.685	Holt
March	3.047.000	2.552.297	2.639.027	Holt
April	2.857.000	2.642.508	2.713.369	Holt

As seen in Table 8, the prediction results of the double exponential smoothing Holt approach provide prediction values that are closer to the real value compared to the ARIMA model during the entire observation period. For example, in January 2026, the Holt prediction of Rp2,490,342 is closer to the actual value of Rp2,602,000 than the ARIMA prediction of Rp2,459,447. The same trend also holds for February, March, and April 2026, where the Holt double exponential smoothing predictions show lower differences than the actual values. It is evident that, even though the ARIMA model produces better results in terms of MAPE, MAE, and RMSE on historical data, it is less accurate in validation using recent real values. Holt is more accurate.

The reason for this difference lies in the differences between the two models' data processing methods. ARIMA puts a lot of importance on the analysis of historical trends and autocorrelations within previous data. Hence, the ARIMA algorithm is likely to generate steady predictions but is not sensitive to recent shifts within trends. At the same time, Holt's Double Exponential Smoothing is able to react to the changes in patterns quickly because it assigns more significance to recent data using level (α) and trend (β) parameters. In this research project, the proximity of the α parameter to one suggests that the Holt model can respond to changes within data patterns quickly. Under such circumstances, the changes in gold prices within the most recent period can play a leading role in shaping the forecast's outcome. Therefore, when gold prices rose sharply in early 2026, the Holt approach could detect this change more quickly than the ARIMA model, which was still influenced by historical trends. Overall, the Holt technique produces more precise predictions when data trends shift frequently, whereas ARIMA performs best with historical trends.

4. CONCLUSION

Based on the research, both Holt's Double Exponential Smoothing and ARIMA models can be employed to estimate future gold prices, owing to their ability to identify patterns and fluctuations in historical gold price data. However, the accuracy levels provided by the two methods differ. Forecasting errors are lower when compared to those of Holt's method, which makes the ARIMA(1,1,6) the optimal model, given that the historical performance of the two is used for comparison purposes. However, when validation of predictions using real figures for January–April 2026 is considered, the results indicate that Holt's double exponential smoothing technique provides realistic estimations.

These disparities reveal that the ARIMA forecasting technique is a better tool in estimating past trends than Holt's technique, which is sensitive to recent trend changes and hence suitable for short-term forecasting. Hence, in forecasting, it is advisable to employ techniques based on the aims of the forecasting. Accordingly, this research shows that the ARIMA (1,1,6) technique is the most appropriate when considering past trends, while Double Exponential Smoothing by Holt performs better in short-term forecasting compared with the current actual figures.

ACKNOWLEDGEMENTS

The authors are very grateful to The Institute for Research and Community Service (LPPM) Universitas Sebelas Maret Surakarta under grant Penguatan Kapasitas Grup Riset (PKGR-UNS) B Nomor: 462/UN27.22/PT.01.03/2026.

REFERENCES

- Ariffin, A. F., Haiqal, M. S., Azriani, N., Nor, M., Hamid, M. H., & Samian, M. H. (2024). *A Comparative Analysis Of Holt's, Arima, And Prophet Methods In Forecasting Kijang Gold Prices*. 7 (25), 325–338. <https://doi.org/10.35631/IJEMP.725027>
- Bhattacharya, R. (2025). *Linear versus Exponential Forecasting Techniques: A Comparative Study of ARIMA and Holt–Winters Triple Exponential Smoothing Models for Indian Retail Gold Prices—Evidence from daily retail gold price data in India, 2014–2025*. 6(10), 1–19.
- Bing, K. K., & Nor, M. E. (2024). *Gold Price Forecasting Using Disaggregation of Time Series Data*. 4(2), 317–328.
- Chasanah, U. (2021). *Analysis of the ARIMA Model in Forecasting Gold Prices During a Period of Global Economic Uncertainty*.
- Chatfield, C. (2022). *The Analysis of Time Series: An Introduction (7th ed.)*. CRC Press.
- Creswell, J. W., & Creswell, J. D. (2022). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications. <https://books.google.co.id/books?id=Rkh4EAAAQBAJ>
- Djami, R. J., & Nanlohy, Y. W. A. (2022). *Consumer Price Index Forecasting in Ambon City Using Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing*. 4, 1–14.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed.)*. OTexts. <https://otexts.com/fpp3/>
- Kahraman, E., & Akay, O. (2022). Comparison of exponential smoothing methods in forecasting global prices of major metals. *Mineral Economics*, 36 , 427–435. <https://doi.org/10.1007/s13563-022-00354-y>
- Kristhy, M. E., Oktalita, S., Yonathan, S., Susanto, T., & Wiji, B. (2022). *The Benefits Of Investing In Antam Gold During The Covid-19 Pandemic*. 5, 388–399.
- Lastri. (2021). *Analysis Of Gold Prices In Indonesia (An Empirical Study from 1996 to 2020)*. 21(1), 129– 136.
- Mandati, S. A., Poniman, & Novianti, T. (2025). *Forecasting the Gold Price of LM Antam using the Double Exponential Smoothing Approach*. 08(1), 54–59.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2021). *Introduction to Time Series Analysis and Forecasting (3rd ed.)*. Wiley.
- Mulia, D. (2024). *Increasing Sales of Gold as an Investment Instrument*. 3 (8), 44–48. <https://doi.org/10.56472/25835238/IRJEMS-V3I8P106>
- Pirmanto, D. (2024). *Double Exponential Smoothing Forecasting of Food Crop Yields Using Geographic Information Systems*. 13, 179–185.
- Rady, E. L. H. A., Fawzy, H., Mohamed, A., & Fattah, A. (2021). *Time Series Forecasting Using Tree-Based Methods*. 244(1), 229–244.
- Report, J. S. (2024). *INDONESIA 2024 Jakpat Survey Report*.
- Sagala, J. P., Dewi, E., Si, T. S., & Si, M. (2023). *Analysis of Antam Gold Price Forecasting Using ARIMA*. 2(1), 77– 84.
- Su, Z. (2024). *Gold Price Forecast Based on ARIMA and ETS Models*. 0 , 185–189. <https://doi.org/10.54254/2754-1169/143/2024.GA18972>
- Sugiyono. (2022). *Quantitative, Qualitative, and R&D Research Methods*. Alfabeta.
- Wang, Z. (2025). *The Effectiveness of Forecasting Gold Prices Using ARIMA Models*. 0 , 232–241. <https://doi.org/10.54254/2754-1169/2025.LH23823>
- Wibowo, A. Z. A., Pratama, B. P., Aulia, M., & Kartiasih, F. (2025). *Comparison of ARIMA and Holt's Double Exponential Smoothing Models for Forecasting Indonesian Gold Prices in 2025*. 7 (2), 68–84. <https://doi.org/10.31605/jomta.v7i2.4934>
- Yenkikar, A., Mishra, V. P., Bali, M., & Ara, T. (2025). *Explainable forecasting of air quality index using a hybrid random forest and ARIMA model*. 15 (July).
- Zheng, H. (2025). *Analysis of International Gold Prices Using ARIMA, SVR with Linear Regression*. 13 (1), 15–21. <https://doi.org/10.18178/joams.1>

Mathematics in the *Nyadran* Tradition and How Students Solve Permutation Problems Based on It

Al Ainaa Almardhiyyah¹, Fitria Sulistyowati², Pardimin Pardimin³, Wikan Budi Utami⁴, Luc Bergman⁵

^{1,2,3}Department of Mathematics Education, Universitas Sarjanawiyata Tamansiswa, Yogyakarta, Indonesia

⁴Department of Mathematics Education, Universitas PGRI Kanjuruhan Malang, Malang, Indonesia

⁵University of Sydney, Sydney, Australia

¹alainaaal02@gmail.com ²fitria.sulistyowati@ustjogja.ac.id ³pardimin@ustjogja.ac.id

⁴wikanbudiutami@unikama.ac.id ⁵lucbergman3@gmail.com

ABSTRACT

This qualitative research aims to reveal the relationship between local culture and the concept of cyclic permutation through an ethnomathematical approach to food arrangement in the *Nyadran* tradition in Banyon Village, Klaten, Central Java, Indonesia. The *Nyadran* tradition in Banyon is an activity of paying respect to ancestors, and the important thing is the arrangement of food in *Tenong*, which is done alternately every year. Data were collected through tests, documentation, observations, in-depth interviews, and literature reviews, and analyzed using the stages of data reduction, data presentation, and conclusion. Interviews were conducted with several traditional authorities and citizens who participated in the *Nyadran* tradition in Banyon. The test subject was one of the students who participated in the *Nyadran* tradition in Banyon. The results of the study show that the arrangement of food in *Tenong* follows an annual rotation pattern that reflects cyclical permutations. This finding proves that the concept of cyclic permutation has been implicitly applied in people's lives through the *Nyadran* tradition. Test results indicate that students can solve cyclic permutation problems in the context of the *Nyadran* tradition. This research is expected to be a reference in developing culturally relevant mathematics learning for students.

Keyword : Ethnomathematics, Cyclic Permutations, *Nyadran* Tradition, *Tenong*.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Fitria Sulistyowati,
Department of Mathematics Education,
Universitas Sarjanawiyata Tamansiswa,
Jl. Batikan UH-III/1043 Umbulharjo, 55167, Yogyakarta, Indonesia.
fitria.sulistyowati@ustjogja.ac.id

1. INTRODUCTION

Indonesia is known as a country rich in diverse cultures and local traditions. Each region possesses its own local wisdom that has been passed down through generations and is continuously preserved by the local community (Brondizio et al., 2021). These traditions not only carry spiritual and social values but also contain patterns of thought that reflect certain order and systematization (Muliadi, 2019; Yanich et al., 2019). One way to understand the relationship between cultural patterns and scientific knowledge is through the ethnomathematical approach (Rodríguez-Nieto et al., 2023), which explores how mathematical concepts naturally emerge within the cultural practices of a society (Prahmana & D'Ambrosio, 2020; Yogasmara Binsar Putra Wimansyah et al., 2025). Through ethnomathematics, mathematics is not merely seen as a collection of formulas and numbers (Abasi, 2025), but as an integral part of real life, embedded in traditions, symbols, and everyday activities.

One example of the application of ethnomathematical concepts in community life can be found in Banyon Village, Klaten Regency, Central Java, Indonesia. The people in this village continue to preserve the *Nyadran* tradition, a communal prayer ritual and feast held as a form of respect for their ancestors (Annisa Ul Aliyah, 2022; Arumugam, 2022). In its practice, they use *Tenong*, a traditional round container made from woven bamboo, to carry and arrange food (Chen et al., 2023; Dzarna et al., 2022). Interestingly, the arrangement of food within the *Tenong*, as well as the circular seating formation of the participants around it, reveals a regular pattern that can be examined from a mathematical perspective, specifically through the concept of cyclic permutations (Borga, 2021; Mauleto, 2025). This pattern is not merely a hereditary custom but reflects an arrangement that carries cultural significance while also embodying fundamental mathematical principles naturally embedded in community practices.

The *Tenong* used in the *Nyadran* tradition is not merely a practical container for carrying food, but also holds rich symbolism and geometric structure embedded with cultural and mathematical values.

Made from lightweight yet sturdy woven bamboo, its round shape reflects equality and togetherness (Handoko, 2019). In real-life practice, community members sit in a circle regardless of social status, forming a symmetrical pattern that not only symbolizes social harmony but can also be analyzed as a representation of a circle in geometry (Gropp et al., 2020; Mubarok, 2024). This circular formation creates a space for equal social interaction and strengthens bonds among community members (Sunarno & Rahmawati Zakiyah, 2024). Philosophically, the shape of the *Tenong* symbolizes the continuity of life, intergenerational connections, and the life cycle values aligned with the spirit of honoring ancestors within the *Nyadran* tradition (S. D. Putri et al., 2025). Thus, this cultural practice embodies not only social and spiritual meaning but also a tangible manifestation of how mathematical principles and cultural values are interwoven in community life.

In addition, from a functional perspective, the circular shape of the *Tenong* facilitates arrangement in open spaces, allowing it to be accessed from all sides and creating a neat and orderly layout. This circular arrangement reflects the concept of circular patterns in mathematics learning, specifically the regular positioning and grouping of objects around a central point (Jainuri et al., 2025; Payadnya et al., 2023). This concept can be taught to students through simple activities such as arranging objects in a circle, determining adjacent or opposite positions, and recognizing the regularity of repeating arrangements (Triyono et al., 2025). These activities can strengthen students' understanding of order, spatial relationships, and the connection between shape and arrangement (Muhammad Yusuf et al., 2023). Although the community does not explicitly refer to these arrangements as mathematical concepts, their practice demonstrates a contextual and natural application of mathematical ideas (Meliasari & Wardana, 2023). This makes the *Tenong* a concrete example of ethnomathematics, a fusion of local culture and mathematical concepts that can be effectively utilized in more meaningful learning experiences.

Interestingly, the arrangement of food in this tradition is not done randomly. Community members arrange the *Tenong* and the food around it in a circular pattern, both during the communal prayer and when the food is opened and shared together (Apisari et al., 2024). This circular arrangement is not coincidental but is a conscious choice that follows the round physical shape of the *Tenong*, creating visual harmony between the container and the layout (Iswandi et al., 2025). This circular formation also reflects the social philosophy of Javanese society, which highly values equality, togetherness, and harmony (Atu et al., 2023; Lu, 2022). In a circle, there is no highest or lowest position; everyone sits on the same level, sharing space and roles equally (L. C. Silalahi et al., 2020). This structure subtly mirrors a social system in which relationships among community members are built on the principles of mutual cooperation and unity.

Mathematically, the patterns found in the *Nyadran* tradition can be further analyzed to reveal elements such as circular shapes, arrangement patterns, and methods of organizing space in an orderly manner. For example, the round shape of the *Tenong* can be introduced as a type of two-dimensional figure, and the circular arrangement of food can be explained as a form of repeating pattern or grouping (Astuti & Madawistama, 2025). With such an approach, cultural traditions can be utilized as concrete and relatable teaching materials in mathematics, making it easier for students to grasp abstract concepts (Annanda Shofi Sulthoni & Ucik Fitri Handayani, 2025). Thus, this tradition not only reflects local wisdom within a cultural framework (Ibnu Fitrianto & Muhammad Farisi, 2025). This tradition can have a deeper impact, as it holds great potential to serve as a rich and meaningful source for contextual mathematics learning (Untara, 2025). Examining the *Nyadran* tradition through an ethnomathematical approach can broaden students' perspectives on mathematics as a discipline deeply connected to social and cultural life (Rudhito et al., 2024).

However, despite this potential for cultural integration, there is a gap between the current trend in ethnomathematics studies and the reality of low conceptual understanding and student learning outcomes in Indonesia. So far, the focus of material in ethnomathematics studies in Indonesia has been dominated by the exploration of geometric concepts, number patterns, and arithmetic in static physical objects such as building architecture, traditional tools, crafts, or batik motifs (Afriansyah & Sugiarti, 2025; Dewi et al., 2025; Putra et al., 2023). Conversely, ethnomathematics research exploring advanced combinatorics concepts, particularly cyclical permutations through dynamic cultural activities and social formations, remains limited and under-recognized. The limited scope of this material has a direct impact on classroom teaching practices, where cyclical permutations are often a topic with low learning

outcomes because students experience misconceptions, difficulty distinguishing circular position orientations, and are accustomed to memorizing abstract formulas without understanding their contextual meaning (Nopriana et al., 2023). This low spatial reasoning and visualization ability is also reflected in the results of the 2022 PISA, which reported that Indonesian students' mathematics scores dropped sharply to 366, placing Indonesia in 70th place out of 81 countries due to students' inability to solve non-routine problems that require problem-solving based on real contexts (OECD, 2023). This is reinforced by previous research, which shows that the main error of Indonesian students in solving context-based problems is the failure to interpret real situations and formulate them into appropriate mathematical concepts (Lamanna et al., 2022; Nopriana et al., 2023; Nursisto & Ismailmuza, 2022). In fact, several studies have shown that utilizing local cultural contexts has been effective in minimizing learning barriers and significantly improving students' mathematical learning outcomes (Hidayati & Prahmana, 2022; Prahmana et al., 2023). Therefore, this research, which examines ethnomathematics in the *Tenong* arrangement in the *Nyadran* tradition in Banyon Village, is urgently needed to fill the gap in cultural combinatorics literature while providing alternative concrete solutions to improve students' conceptual understanding and learning outcomes in the cyclical permutation material.

This study aims to describe and analyze the ethnomathematical elements embedded in the use of *Tenong* and the circular arrangement of food during the *Nyadran* tradition in Banyon Village, Klaten Regency, Central Java, Indonesia (Fauzi & Lu'luilmaknun, 2019). This study is not limited to exploring mathematics within the *Nyadran* tradition but also explores it through tests on students directly involved in the tradition. These tests aim to identify the extent to which the *Nyadran* tradition can develop students' problem-solving skills in cyclical permutation discussions. This study is expected to contribute to the development of culturally-based mathematics education (Masamah, 2019). It not only fosters conceptual understanding but also builds a meaningful connection between students and their own cultural heritage.

2. RESEARCH METHOD

The subjects in this study were selected using purposive sampling, taking into account individuals or groups with knowledge, involvement, and direct experience in the implementation of the *Nyadran* tradition in Banyon Village, Klaten Regency, Central Java, Indonesia. The main focus of the research subjects is the community members who actively participate in the *Nyadran* tradition, particularly cultural practitioners and local figures directly involved in the procession (Li et al., 2021). They are considered authentic and reliable sources of information, as they understand the philosophical meanings and the structure of food arrangement within the *Tenong* used during *Nyadran* (Wulanjari et al., 2024). The *Tenong* unique to this village not only embodies rich cultural values but also contains mathematical concepts, particularly cyclic permutations, which can be analyzed through an ethnomathematical approach.

In addition to the community as the primary source of data, this study also integrates ethnomathematical concepts directly into mathematics learning. Therefore, the researcher involved one 11th-grade high school student as an additional subject, tasked with solving contextual problems based on the *Nyadran* tradition (Utmeemang & Buaraphan, 2024). The student was selected based on criteria such as being cheerful, intelligent, and demonstrating a good understanding of mathematical concepts, although occasionally showing a lack of precision in writing numbers or solution steps. (Windasari et al., 2020). However, these minor errors did not affect the accuracy of the final answers, as the student was able to complete the problems using correct logical reasoning (Cresswell & Speelman, 2020). The inclusion of the student in this research aimed to explore the extent to which the concept of cyclic permutations, derived from cultural practices, could be understood and applied concretely in a mathematics learning context (Lathiifah et al., 2019). Thus, this study not only uncovers mathematical values within local culture but also highlights the potential for their application in context-based education rooted in local wisdom.

The data in this study were collected using several techniques aligned with the ethnographic approach and ethnomathematical inquiry.

1. Indirect observation was conducted due to the missed opportunity to observe the *Nyadran* tradition directly. Therefore, the observation technique was replaced with video recordings and photographic documentation obtained from reliable sources, such as YouTube, local cultural

figures, or village residents who documented the event (Yolan et al., 2024). This indirect observation aimed to study the patterns of *Tenong* usage and the community's interaction with the object within the context of the *Nyadran* ritual.

2. Interviews were conducted with informants involved in the implementation of *Nyadran*, such as cultural practitioners who understand its meaning and a community member who has previously used a *Tenong*. The purpose of these interviews was to gather information related to the philosophical meanings and mathematical concepts, particularly cyclic permutations, embedded in the patterns and arrangements of the *Tenong* (Sutra et al., 2025).
3. Documentation
 - Photos and videos obtained from local sources were also used to support the analysis of cyclic permutation patterns in the *Tenong*. This documentation served as a means of verifying data obtained from interviews and indirect observation.
 - Photos of the students' written responses, collected directly, were also used as part of the documentation.

With this combination of data collection techniques, the study is expected to produce comprehensive and valid data for analyzing the relationship between cultural traditions and the concept of cyclic permutations in the *Tenong* unique to Banyon Village, Klaten Regency, Central Java, Indonesia.

The data collected in this study were analyzed through several systematic stages. The first stage was data reduction, which involved sorting and selecting data relevant to cultural practices and the mathematical forms that emerge in the *Nyadran* tradition, particularly those related to the pattern of cyclic permutations in the *Tenong* (Sari et al., 2021). This process aimed to filter out the most important information aligned with the research focus. The second stage was data presentation, where the reduced data were presented in the form of descriptive narratives that depict the cultural context and the meaning of *Tenong* usage (Selmi et al., 2024). Additionally, the presentation was supported by visual documentation, including photos and diagrams of the *Tenong* arrangement patterns that illustrate the circular positions used in the tradition. This presentation helped clarify and reinforce the visualization of cyclic permutation patterns within the cultural practice. The final stage was concluding, conducted by connecting the identified cultural patterns with the concept of cyclic permutations in mathematics (Ahmad Arisman Nasution & Suparni, 2024). The analysis was performed contextually to demonstrate how such cultural practices implicitly contain mathematical principles (Abang Thamren et al., 2025). Thus, this study not only reveals cultural aspects but also highlights the connection between local traditions and mathematical concepts within the domain of ethnomathematics.

3. RESULTS AND DISCUSSION

Result

The selection of research subjects was based on a purposive sampling approach, in which data were collected through interviews with cultural leaders and practitioners in Banyon Village who were considered to have in-depth knowledge of the *Nyadran* tradition and the arrangement of food in the *Tenong*. In addition, visual documentation in the form of photos and videos of the tradition was also used as an important data source for analyzing cyclic permutation patterns (Senoprabowo et al., 2021). Subjects were chosen based on their active involvement in the tradition and their relevant understanding of the cultural patterns under study (Sheppard & Broughton, 2020). The findings revealed that the arrangement of food in the *Tenong*, unique to Banyon Village, follows a repeating circular pattern and exhibits characteristics of cyclic permutations. The food is not arranged randomly, but rather follows a rotation of positions from one type of food to another in a sequential manner with each occurrence of the tradition (Tamutu & Hasmah, 2024). This pattern reflects the principle of cyclic permutations in mathematics, where elements shift positions systematically without altering their overall relative order (Aini & Budiarto, 2022). For example, if there are five types of food in one *Tenong*, the food in the first position will move to the second position in the next ritual, and so on, until it returns to its original position.

These findings are supported by an analysis of visual documentation that clearly and consistently shows the rotational pattern of food positions, thereby reinforcing the idea that the *Nyadran* tradition implicitly adopts mathematical concepts within everyday cultural practices. Thus, both cultural figures and visual documentation provide in-depth and relevant information for understanding how the

principles of cyclic permutations are applied in a local cultural context (Faris & Hidajat, 2025). This study opens new insights into the close relationship between cultural practices and mathematical concepts, particularly within the field of ethnomathematics (Rosa & Clark Orey, 2019). The results enrich our understanding of the mathematical values embedded in local traditions, while also emphasizing the importance of preserving these traditions as authentic and contextual sources of knowledge.

Observation Results. The results of this observation indicate that the arrangement of food inside the *Tenong* is not done randomly, but instead follows a repeating and systematic circular pattern. Through the observed visualizations, the food is arranged in a circular formation within the *Tenong*. The number of food types served varies, usually between 4 to 7 types (Dosinaeng et al., 2020). Each type of food is placed in a specific position according to the existing layout. In the next *Nyadran* ceremony, the positions of these food items shift one step to the next position in a sequential manner, forming a continuous rotational pattern.

This pattern indicates the presence of a cyclic rotation of food positions, where each type of food shifts sequentially in a circular motion without altering the relative order among them. Observations also reveal a parallel rotation pattern for larger groups of food items, which can be interpreted as multiple simultaneous permutation cycles (disjoint cycles) (Nuryadi & Kholifa, 2020). By analyzing food positions through photographic documentation, this cyclic rotation pattern appears clear and consistent in each *Nyadran* ceremony (Fajriyeh & Zayyadi, 2023). This strengthens the assumption that the arrangement of food in the *Tenong* represents a natural application of the cyclic permutation concept within the *Nyadran* tradition in Banyon Village. Beyond the mathematical aspect, this observation also highlights the symbolic meaning of food positions as a representation of the life cycle and social harmony, preserved and passed down through generations in the *Nyadran* tradition.

Interview Results. Interviews were conducted with a cultural practitioner and a community member from Banyon Village. The purpose of the interviews was to gather information regarding the use of the *Tenong* in the *Nyadran* tradition procession within the context of ethnomathematics.

Interview with cultural leader

Researcher : "How is the food arranged in the *Tenong* during the *Nyadran* tradition in Banyon Village?"

Cultural Leader: "During *Nyadran*, each family brings a *Tenong* filled with food as an offering to the ancestors. The food isn't placed randomly; we arrange it in a circular and orderly way. There's a certain sequence that has become customary, and every year, the position of the food is rotated."

Researcher : "So, there is a specific pattern to the food arrangement in the *Tenong*, right?"

Cultural Leader: "Yes, exactly. For example, if the chicken is in the front this year, next year it might move to the side or the back. All kinds of food like vegetables, side dishes, and snacks take turns. This is done for balance and fairness. It's been practiced since our ancestors' time, every *Nyadran*."

Researcher : "Is there a particular meaning behind the rotation of food positions in *Nyadran*?"

Cultural Leader: "Yes, there is. For us, it symbolizes the cycle of life like a wheel. Everything takes turns, and nothing should remain in the same place forever. *Nyadran* is not only about praying for our ancestors, but also about teaching deep values of life."

Interview with resident

Researcher : "Do you usually take part in bringing a *Tenong* during the *Nyadran* ceremony in Banyon Village?"

Resident : "Yes, we join *Nyadran* every year. Our family prepares a *Tenong* filled with food to bring to the ancestors' graves. It usually contains rice, vegetables, side dishes, and traditional snacks."

Researcher : "Who arranges the food in the *Tenong*?"

Resident : "We arrange it together at home. There's a particular way to do it; it can't be done carelessly. The placement must be balanced. For example, main dishes like chicken or eggs are usually placed in the center or at the front, and the other food surrounds them. That's been a custom passed down from our elders."

Researcher : "Is the food arrangement always the same every year?"

- Resident : “No. We usually rotate the food positions in the *Tenong* each year. For instance, if the chicken was in the front last year, this year it might be placed to the side. Every dish gets a turn to occupy what’s considered the main position.”
- Researcher : “What’s the meaning behind this rotation of food positions in *Nyadran*?”
- Resident : “So that it’s fair and every dish gets an important spot. It also symbolizes that life goes in cycles; everything has its turn. During *Nyadran*, it’s not just about praying, but also about remembering togetherness and sharing with one another.”
- Researcher : “Have you ever realized that it’s like a rotating pattern or a cycle?”
- Resident : “Now that you mention it, yes, it’s like it keeps turning every year. But that’s just how it’s always been, so we follow it as part of the tradition.”

Discussion

Tenong in Cultural Tradition. The *Nyadran* tradition is a form of local wisdom that continues to be preserved and passed down through generations by the people of Banyon Village, Klaten Regency, Central Java, Indonesia. This practice is carried out as an expression of respect for ancestors and as a means to strengthen social bonds among community members (Anggreini et al., 2024). *Nyadran* is generally held prior to the month of Ramadan or during certain sacred times, such as before the harvest season or on major religious holidays (Rohman et al., 2024). All community members participate by bringing various dishes from home, neatly arranged in a special container called a *Tenong*. A *Tenong* is a traditional round container made of woven bamboo or metal, often covered with a batik cloth as a lid (Mahayanti et al., 2024). Functionally, the *Tenong* is used to carry food, but it also holds deep symbolic meaning. Its circular shape reflects the philosophy of unity, harmony, and the absence of hierarchy; every point on a circle occupies an equal position (Ishak, 2020). This aligns with the Javanese worldview that life is a shared space without caste or status distinctions (Yulianto et al., 2022). The *Tenong* also symbolizes determination, sincerity in giving, and the spirit of mutual cooperation in community life.

Based on the results of observations and interviews, there is an unwritten rule, orally passed down from generation to generation, that the arrangement of food in the *Tenong* is not done arbitrarily. The types of food commonly served include rice, *apem* cakes, free-range chicken, boiled eggs, and *rengginang* (crispy rice crackers). Each type of food is believed to carry its own symbolic meaning (Sekar Kirana Wulandari et al., 2024). For instance, *apem* symbolizes a plea for forgiveness (A. H. Putri et al., 2024); chicken as a source of protein, represents health for the community (Sembiring et al., 2022); Rice, as a staple food, symbolizes prosperity for all (Mu'min, 2020); eggs represent the unity of life that is inseparable from nature (Malo et al., 2022); and *rengginang* made from rejoined rice grains, symbolizes brotherhood (Vina Tri Agustinningrum & Sukarman Sukarman, 2024). This tradition illustrates that the *Tenong* is not merely a serving container but a vessel of noble values (Diana, 2023). Within every dish and its arrangement lie deep moral and spiritual messages. The value of sincerity is reflected in how each family brings food without expecting anything in return (Lamawuran et al., 2023). The value of equality is evident in how every type of food is treated with respect. And the value of togetherness is embodied in the moment when all residents sit in a circle, offer prayers, and share a meal from the same container, strengthening a sense of solidarity and brotherhood.

The *Nyadran* tradition, which uses the *Tenong* as the main container for offerings, not only represents a collective ritual but also embodies profound philosophical values deeply rooted in the Javanese worldview (F. H. M. Silalahi & Purwanto, 2025). Two key philosophies reflected in this practice are *urip iku mung mampir ngombe* (life is temporary and always rotating) and *sangkan paraning dumadi* (everything will return to its origin); both are not only upheld as life principles, but are also symbolically reflected in every aspect of the tradition, including the arrangement of food within the *Tenong*. (Hannan & Khotibum Umam, 2023).

The Javanese proverb *urip iku mung mampir ngombe*, which literally means life is just a stopover to drink, carries the message that life in this world is temporary. In the context of *Tenong* culture, this philosophy is reflected in the community’s attitude of not competing for position or status, but rather accepting with sincerity that each type of food in the *Tenong* will have its own turn to be placed at the front (Ambarwati et al., 2019). This demonstrates a life perspective of humility, contentment, and awareness that everything changes and rotates in due time. Just like in human life, no one remains in

the same position forever; everyone experiences a cycle of fate, happiness, sorrow, roles, and responsibilities (Safitri et al., 2022). Therefore, the rotation of food within the *Tenong* symbolically reflects the cycle of life: what is below will rise, and what is above will shift, in accordance with the natural order and the passage of time.

Meanwhile, the philosophy of *sangkan paraning dumadi*, which means that all things emanate from and ultimately revert to their source, adds spiritual depth to the arrangement and rotation of food within the *Tenong*. The annual rotation of food from one position to another is not merely a hereditary custom but also symbolizes the journey of living beings from the Creator and back to Him (Amanda Anyza & Listyo Yuwanto, 2023). When the food in the *Tenong* returns to its original position after several cycles, it reflects the eternal cycle of life's destiny (Herawati, 2023). The circular arrangement inside the *Tenong* also represents the unity of life's origin and purpose, emphasizing that in togetherness and equality, humans will always return to their essence and source.

Thus, the *Tenong* tradition is not merely a part of customary rituals but a philosophical expression of Javanese views on life, time, and existence. The *Tenong* serves as a symbol that imparts wisdom about accepting the cycles of life, understanding the origin and purpose of existence, and fostering social harmony rooted in equality, sincerity, and profound spiritual awareness.

Cyclic Permutation Simulation. In everyday life, we often encounter arrangements of objects or people in a circular formation such as sitting around a round table, arranging food in a circle, or implementing a rotating shift system. The mathematical concept that describes such arrangements is known as cyclic permutation (Engen & Vatter, 2021). Unlike regular permutations, which count the number of arrangements in a straight line, cyclic permutations only count arrangements that are truly distinct in a circular form (Ball & Serra, 2024). This is because in a circle, there is no absolute beginning or end, so arrangements that differ only by their starting point are considered the same.

General Formula:

$$P_n = (n - 1)!$$

Explanation:

P_n : the number of ways to arrange n objects in a circle.

n : the number of known objects

! : factorial symbol (product of all positive integers up to n)

For example, if 4 people are sitting in a circle, the number of unique arrangements is: $P_4 = (4 - 1)! = 3! = 3 \times 2 \times 1 P_4 = 6$. So, the answer is 6 distinct circular arrangements, because rotations of the same arrangement are not counted as new ones.

For example, in one *Tenong*, there are six types of food, namely: A as Rice; B as Chicken; C as Eggs; D as Vegetable; E as *Apem*; F as *Rengginang*.

In the first year, the arrangement in clockwise order is as follows:

1. Position 1 (very front): A
2. Position 2 (front right): B
3. Position 3 (back right): C
4. Position 4 (back): D
5. Position 5 (back left): E
6. Position 6 (front left): F

In the **second year**, there is a one-position shift in the clockwise direction:

1. Position 1: F
2. Position 2: A
3. Position 3: B
4. Position 4: C
5. Position 5: D
6. Position 6: E

And so on, until after 6 years, the arrangement returns to its original position. This demonstrates a single cyclic permutation of 6 elements; see the following illustration in Figure 1.

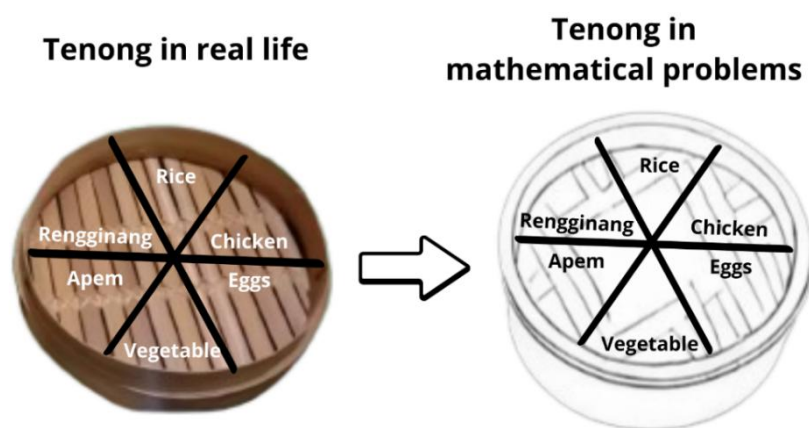


Fig 1. Illustration of *Tenong* in *Nyadran* Tradition in Banyon Village

The context in Figure 1 can be used as a cyclic permutation problem under different conditions as follows.

In the *Nyadran* tradition of Banyon Village, the community arranges food in a circular *Tenong* containing 5 types of food: fruit, *apem*, *rengginang*, chicken, and eggs, placed in clockwise order. Each year, the food positions are rotated one step clockwise so that each food item takes turns occupying the front position (the side facing the prayer direction).

1. How many different circular arrangements can be formed from the arrangement of the food items?
2. If this year (2025) the front position is occupied by fruit, in what year will chicken be in the front position?
3. After how many years will the arrangement return to its original order?

To solve the first question, Grade XI students need to first understand the concept of cyclic permutations, which involve arranging objects in a circle where one arrangement is considered the same as another if they differ only by rotation. In the problem, it is stated that there are 5 types of food: fruit, *apem*, *rengginang*, chicken, and egg, arranged in a circle inside a *Tenong*. Since this is a circular arrangement, the number of distinct arrangements can be calculated using the formula $(n - 1)!$, where n is the number of objects being arranged. So, $(5 - 1)! = 4! = 4 \times 3 \times 2 \times 1 = 24$ distinct circular arrangements can be formed. Next, for the second question, it is stated that in the year 2025 the front position is occupied by fruit, and the food positions are rotated one step clockwise each year. Since chicken is in the fourth position after fruit, it will move to the front in the fourth year after 2025, which is in the year 2028.

For the third question, since the rotation is done one step each year in a five-element arrangement, it takes 5 years for all positions to rotate back to their original places. Therefore, the arrangement will return to its initial state in the year 2030. This explanation demonstrates how mathematics, particularly cyclic permutations, can be applied to logically and systematically understand patterns within cultural traditions.

This problem is designed for 11th-grade high school students, as the concept of cyclic permutations is part of the advanced mathematics curriculum introduced at that level. By incorporating local cultural context through the *Nyadran* tradition, students are encouraged to understand mathematical concepts in a contextual and applicable manner (Sulistyowati et al., 2024). Through the arrangement of food in the *Tenong* that rotates each year, students not only learn to calculate cyclic permutations using the formula $(n - 1)!$, but also to analyze rotational patterns and determine the time it takes for the arrangement to return to its original order (Sabon et al., 2021). Questions such as determining when the chicken reaches the front position and how long it takes for the arrangement to reset encourage students to think logically, systematically, and to meaningfully connect cultural practices with abstract mathematical concepts.

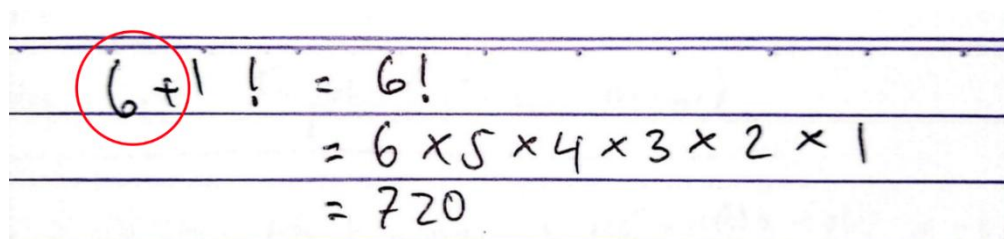
To assess the extent of students' understanding of the given problem, their answers were analyzed based on the work submitted on paper. Each student was asked to write down the steps of their solution, including the formula used, the logical reasoning behind their calculations, and the accurate

determination of the year based on the rotation pattern of food arrangements in the *Nyadran* tradition (Gombo, 2024; Irfan et al., 2023; Sulistyowati, Kuncoro, Nugraheni, et al., 2019; Sulistyowati, Kuncoro, Setiana, et al., 2019). This analysis of the written responses provides a concrete picture of how students apply the concept of cyclic permutations within a local cultural context and reflects their individual level of mathematical thinking skills (Trisnawati, 2022). The following section presents examples of student responses that illustrate their thought processes and final answers to the completed problem.

Question Number 1

How many different circular arrangements can be formed from the arrangement of the food items?

Student's Answer:



$$(6+1)! = 6!$$

$$= 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

$$= 720$$

Fig 2. Student's Answer

Based on the analysis of one student's answer to the question in Figure 2, it is evident that the final answer given was incorrect. The student wrote the final result as 720, whereas according to the concept of circular permutation, the correct answer should be 24. This result is obtained from the formula $(n - 1)!$ with $n = 5$, so $(5 - 1)! = 4! = 4 \times 3 \times 2 \times 1 = 24$. The error occurred because the student applied the formula for linear permutation instead of circular permutation. In addition, the solution lacked the correct formula notation, and there were mistakes in the initial numerical setup from the problem, as well as using the addition sign (+) where the subtraction sign (−) should have been used. Nevertheless, the steps taken indicate that the student has begun to understand the pattern of rotation and the logical structure of the problem, although there was a misapplication of the formula.

Learning circular permutations through an ethnomathematical approach using the context of the *Nyadran* tradition still provides a positive impact on students' thinking processes. Culturally contextualized problems make students more engaged and help them visualize the problem in a concrete way. The student also demonstrated the ability to calculate arrangement possibilities based on the yearly rotation of food in the *Tenong* (Herizal, 2021). This shows that the ethnomathematical approach can strengthen the connection between abstract mathematical concepts and students' cultural realities (Gustina et al., 2025). It helps enhance conceptual understanding, accuracy, and analytical skills in learning mathematics more meaningfully.

Question Number 2

If this year (2025) the front position is occupied by fruit, in what year will chicken be in the front position?

Student's Answer:

☐	Buah = 2025	☐	Fruit = 2025
☐	Apem = 2026	☐	Apem = 2026
☐	Penggirang = 2027	☐	Penggirang = 2027
☐	Ayam = 2028	☐	Chicken = 2028
☐	Telur = 2029	☐	Eggs = 2029
☐	Jadi ayam akan ada diposisi depan pada tahun 2028	☐	So the chicken will be in the front position in 2028
in Bahasa		in English	

Fig 3. Student's Answer

Based on the analysis of the student's answer in Figure 3, it shows that the student was able to accurately determine the position of food items in specific years based on the rotation pattern in circular

permutations. For example, when asked to determine in which year the chicken would be in the front position if the starting year (2025) was occupied by the fruit, the student was able to calculate and conclude that the chicken would be in the front in 2028, which is three years after the initial position. This step indicates that the student not only understood the concept of circular position shifting but also successfully applied it in the context of time, based on the number of elements and the direction of rotation stated in the problem. This accuracy reflects a solid understanding of cyclic patterns and their application in real life. Interestingly, through a learning process that connects mathematical concepts with local culture such as the *Nyadran* tradition, students not only showed more focus in understanding the material, but also demonstrated the ability to confidently predict which type of food would be in the front position in the coming years. This indicates that the ethnomathematical approach not only fosters learning interest but also cultivates analytical thinking and skills in projecting long-term patterns (Rachmawati & Purwaningrum, 2019). In other words, students are not just learning mathematics they are practicing how to identify patterns in real life in a meaningful and enjoyable way.

Question Number 3

After how many years will the arrangement return to its original order?

Student's Answer:

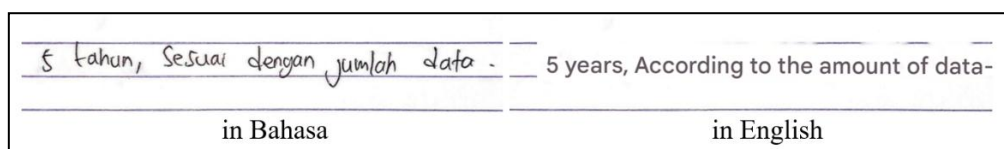


Fig 4. Student's Answer

Based on the analysis of the student's answer in Figure 4, it shows that the student successfully answered the question correctly, particularly in determining when the food arrangement in the *Tenong* would return to its original position. The student demonstrated an understanding that, since there are 5 types of food arranged in a circle and the positions rotate one step clockwise each year, it would take 5 years for the arrangement to return to its original order. This response indicates that the student comprehends the fundamental principle of circular permutations and is able to relate it to the context of annual rotation in the *Nyadran* tradition. This accuracy reflects the student's ability to recognize patterns and apply mathematical concepts logically and contextually (Saminanto et al., 2024). Furthermore, the student was also able to accurately predict when the food arrangement would return to its initial form, specifically in the year 2030 if the starting position was in 2025. This ability serves as evidence that ethnomathematics-based learning not only strengthens conceptual understanding of mathematics but also develops a systematic way of thinking, enabling students to read and forecast cyclical events mathematically (Hia et al., 2025). This makes the learning process more meaningful and relevant by connecting it to real-life experiences.

Thus, in line with the understanding of (Rosa et al., 2016) through learning based on local culture such as the *Nyadran* tradition in Banyon Village, students not only successfully comprehend the concept of circular permutations mathematically, but also become aware that such concepts exist in their real-life experiences. This understanding reinforces the belief that mathematics is not merely a collection of numbers and formulas, but a living and meaningful part of cultural heritage (D'Ambrósio & Knijnik, 2019; Ridho'i, 2024). The integration of ethnomathematics into learning provides an educational experience that is closer to students' realities, enhances their sense of ownership over the material being studied, and enriches their understanding of the connection between science and culture (Fatimah et al., 2024). Therefore, such an approach is highly relevant for implementation in mathematics education, especially in supporting contextual learning rooted in local wisdom.

4. CONCLUSION

The results of this study indicate that the *Nyadran* Tradition in Banyon Village is not only rich in cultural and spiritual values but also reflects a mathematical concept, namely cyclic permutations in the arrangement of food in *Tenong*. The results of the interview indicate that the round shape of *Tenong*

symbolizes equality in a society without hierarchy. The annual rotation of food symbolizes social harmony and Javanese philosophical values, namely *urip iku mung mampir ngombe* (life is temporary and always rotating) and *sangkan paraning dumadi* (everything will return to its origin). The observation results show that the arrangement of food in Tenong is not random but follows a systematic circular pattern. Every year, the position of the food shifts one step clockwise sequentially. This characteristic reflects the principle of cyclic permutation, where elements shift positions systematically without changing their relative order. The results of student tests in solving cyclic permutation problems in the context of Tenong in the Nyadran tradition show that students tend to better understand the given context because it is in accordance with everyday life, although there are errors in finding possible arrangements in Tenong due to reading errors. Through an ethnomathematics approach, students are able to understand and apply the concept of cyclic permutations contextually, such as counting the number of arrangements, determining the position of food in a given year, and identifying when the arrangement returns to its original form. This learning experience not only strengthens students' understanding of mathematical content but also fosters awareness of the presence of mathematics in their daily lives and cultural practices. This approach aligns with the Indonesian National Curriculum, which emphasizes meaningful, contextual learning based on local wisdom.

ACKNOWLEDGEMENTS

The authors would like to thank LPPM Universitas Sarjanawiyata Tamansiswa for providing financial assistance in conducting research and publication. The authors would like to thank all parties who have assisted in conducting the research. Hopefully, this research can be a reference for further research related to the ethnomathematical study of tradition in Indonesia.

REFERENCES

- Abang Thamren, Sandie, & Jamilah. (2025). Eksplorasi Konsep Geometri Pada Ikan Keturi Piring Dalam Tradisi Masyarakat Melayu Jangkang. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, 12(1), 97–109. <https://doi.org/10.31316/j.derivat.v12i1.7579>
- Abasi, A. U. (2025). *Effects of Ethno-mathematics Approach on Achievement in Number and Numeration concept among Junior Secondary School Students in Akwa Ibom State*. <https://doi.org/10.5281/ZENODO.15305144>
- Afriansyah, E. A., & Sugiarti, S. (2025). Ethnomathematics in Sundanese Crafts: A Systematic Literature Review. *Mosharafa: Jurnal Pendidikan Matematika*, 14(2), 571–582. <https://doi.org/10.31980/mosharafa.v14i2.3432>
- Ahmad Arisman Nasution & Suparni. (2024). Eksplorasi Etnomatematika Pada Kipang Panyabungan Sebagai Makanan Khas Mandailing Natal. *Jurnal Pendekar Nusantara*, 1(2). <https://doi.org/10.37776/pend.v1i2.1285>
- Aini, N. N., & Budiarto, M. T. (2022). Literasi Matematis Berbasis Budaya Mojokerto Dalam Perspektif Etnomatematika. *MATHEdunesa*, 11(1), 198–209. <https://doi.org/10.26740/mathedunesa.v11n1.p198-209>
- Amanda Anyza & Listyo Yuwanto. (2023). Suwung: Pencarian Kesempurnaan Hidup Masyarakat Jawa. *Jurnal Budaya Nusantara*, 6(1), 221–227. <https://doi.org/10.36456/JBN.vol6.no1.7503>
- Ambarwati, P., Wardah, H., & Sofian, M. O. (2019). Nilai Sosial Masyarakat Madura dalam Kumpulan Syair Lagu Daerah Madura. *Satwika : Kajian Ilmu Budaya Dan Perubahan Sosial*, 3(1), 54–68. <https://doi.org/10.22219/satwika.v3i1.8682>
- Anggreini, M. S., Laili, B. N., Albani, A., Anggraeni, A. P., Sugiantoro, S., & Setyawan, K. G. (2024). Tradisi Grebek Suro Sebagai Bentuk Rasa Syukur Masyarakat Trowulan Mojokerto. *Maharsi*, 6(3), 68–77. <https://doi.org/10.33503/maharsi.v6i3.43>
- Annanda Shofi Sulthoni & Ucik Fitri Handayani. (2025). Eksplorasi etnomatematika pada masjid nurul anwar untuk pembelajaran bangun datar dan ruang di mts mambaul ulum banjarejo. *Jurnal Inovasi Pembelajaran Matematika: PowerMathEdu*, 4(1). <https://doi.org/10.31980/pme.v4i1.2655>
- Annisa Ul Aliyah, T. R. (2022). Nilai-nilai estetika dalam tradisi nyadran di dusun blambangan, desa gedangan, kecamatan cepogo, kabupaten boyolali. *Indonesian Journal of Muhammadiyah Studies (IJMUS)*, 3(2), 90–98. <https://doi.org/10.62289/ijmus.v3i2.45>
- Apisari, E., Nurhayati, & Agus Subiyanto. (2024). Pola Komunikasi pada Tradisi Genduren di Dusun Talang, Kabupaten Magelang (Pendekatan Etnografi Komunikasi). *Kawruh: Journal of Language Education, Literature and Local Culture*, 6(2), 91–102. <https://doi.org/10.32585/kawruh.v6i2.5936>
- Arumugam, I. (2022). Laying Out Feast-Offerings: Offering Meat, Feasting Together and Sharing with the Gods. *Religions of South Asia*, 15(3). <https://doi.org/10.1558/rosa.21396>
- Astuti, W., & Madawistama, S. T. (2025). Etnomatematika Permainan Tradisional Congklak Sebagai Penerapan Konsep Matematika. *Jurnal Inovasi Pendidikan Dan Sains*, 6(1), 34–40. <https://doi.org/10.51673/jips.v6i1.2391>
- Atu, L. F., Solosumantoro, H., & Langgor, M. (2023). Konsep Filosofis Budaya Lonto Leok di Manggarai dalam Perspektif Filsafat Dialogis Martin Buber. *Borneo Review*, 2(2), 116–124. <https://doi.org/10.52075/br.v2i2.260>
- Ball, S., & Serra, O. (2024). Enumeration with Symmetries. In S. Ball & O. Serra, *A Course in Combinatorics and Graphs* (pp. 33–53). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55384-4_3
- Borga, J. (2021). *Random Permutations—A geometric point of view* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2107.09699>

- Brondizio, E. S., Aumeeruddy-Thomas, Y., Bates, P., Carino, J., Fernández-Llamazares, Á., Ferrari, M. F., Galvin, K., Reyes-García, V., McElwee, P., Molnár, Z., Samakov, A., & Shrestha, U. B. (2021). Locally Based, Regionally Manifested, and Globally Relevant: Indigenous and Local Knowledge, Values, and Practices for Nature. *Annual Review of Environment and Resources*, 46(1), 481–509. <https://doi.org/10.1146/annurev-environ-012220-012127>
- Chen, X., Chen, F., Yang, Q., Gong, W., Wang, J., Li, Y., & Wang, G. (2023). An environmental food packaging material part I: A case study of life-cycle assessment (LCA) for bamboo fiber environmental tableware. *Industrial Crops and Products*, 194, 116279. <https://doi.org/10.1016/j.indcrop.2023.116279>
- Cresswell, C., & Speelman, C. P. (2020). Does mathematics training lead to better logical thinking and reasoning? A cross-sectional assessment from students to professors. *PLOS ONE*, 15(7), e0236153. <https://doi.org/10.1371/journal.pone.0236153>
- D'Ambrósio, U., & Knijnik, G. (2019). Ethnomathematics. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 1–6). Springer International Publishing. https://doi.org/10.1007/978-3-319-77487-9_60-3
- Dewi, A. M., Ramadhani, H. M., Sarirah, F. H., & Fahmy, A. F. R. (2025). Systematic Literature Review: Exploring the Ethnomathematics of Javanese Traditional Houses. *Ethnomathematics Journal*, 6(1), 18–32. <https://doi.org/10.21831/ej.v6i1.68128>
- Diana, E. (2023). Eksplorasi Nilai-Nilai Luhur dalam Tradisi Lisan “Berasan” Adat Perkawinan Kota Bengkulu. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 6(1), 205–222. <https://doi.org/10.30872/diglosia.v6i1.550>
- Dosinaeng, W. B. N., Lakapu, M., Jagom, Y. O., Uskono, I. V., Leton, S. I., & Djong, K. D. (2020). Etnomatematika Untuk Siswa Sekolah Menengah: Eksplorasi Konsep-Konsep Geometri Pada Budaya Suku Boti. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 9(3), 739. <https://doi.org/10.24127/ajpm.v9i3.2900>
- Dzarna, Mijianti, Y., & Dewi, E. S. (2022). Makna Simbolik Makanan Khas Bakalan: Studi Etnografi. *BELAJAR BAHASA: Jurnal Ilmiah Program Studi Pendidikan Bahasa Dan Sastra Indonesia*, 7(1), 15–24. <https://doi.org/10.32528/bb.v7i1.6>
- Engen, M., & Vatter, V. (2021). Containing All Permutations. *The American Mathematical Monthly*, 128(1), 4–24. <https://doi.org/10.1080/00029890.2021.1835384>
- Fajriyeh, L., & Zayyadi, M. (2023). Etnomatematika: Eksplorasi Budaya Roket TaseaTM Pantai Jumiang Pamekasan Madura. *JEMS: Jurnal Edukasi Matematika Dan Sains*, 11(2), 458–467. <https://doi.org/10.25273/jems.v11i2.16280>
- Faris, A. I., & Hidayat, R. (2025). Fungsi Musik pada Atraksi Tarung Rotan dalam Ritual Ojhung di Desa Blimbing Kabupaten Bondowoso. *Journal of Language Literature and Arts*, 5(5), 557–568. <https://doi.org/10.17977/um064v5i52025p557-568>
- Fatimah, S., Zulfi Fajriyah, R., Fatimah Zahra, F., & Prasetyo, S. P. (2024). Integrasi Etnomatematika dalam Pembelajaran Matematika di Sekolah Dasar Berbasis Kesenian Tari Budaya Lampung. *Al-Madrasah Jurnal Pendidikan Madrasah Ibtidaiyah*, 8(4), 1631. <https://doi.org/10.35931/am.v8i4.3721>
- Fauzi, A., & Lu'luilmaknun, U. (2019). Etnomatematika Pada Permainan Dengklag Sebagai Media Pembelajaran Matematika. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 8(3), 408. <https://doi.org/10.24127/ajpm.v8i3.2303>
- Gombo, M. (2024). Matematika Dan Kearifan Lokal: Noken Sebagai Media Edukatif Di Tanah Papua. *Haga: Jurnal Pengabdian Kepada Masyarakat*, 3(1), 127–138. <https://doi.org/10.57094/haga.v3i1.2640>
- Gropp, A., Yariv, L., Haim, N., Atzmon, M., & Lipman, Y. (2020). *Implicit Geometric Regularization for Learning Shapes* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2002.10099>
- Gustina, D. M., Mariana, N., & Wiryanto, W. (2025). Augmented Reality-Based Ethnomathematics Learning Media to Enhance Spatial Ability in 3D Geometry for Fifth Grade Elementary Students. *Journal of Innovation and Research in Primary Education*, 4(2), 273–280. <https://doi.org/10.56916/jirpe.v4i2.1229>
- Handoko, H. (2019). Dolung dolung Kudapan Tradisional Berbahan Dasar Tepung Beras dan Kolang Kaling Buah Aren Pangan yang Menyehatkan Dalam Mengembangkan Wisata Kuliner Kota Wisata Parapat Kabupaten Simalungun. *Jurnal Akademi Pariwisata Medan*, 7(2), 1–10. <https://doi.org/10.36983/japm.v7i2.43>
- Hannan, A. & Khotibum Umam. (2023). Tinjauan Sosiologi Terhadap Relasi Agama Dan Budaya Pada Tradisi Koloman Dalam Memperkuat Religiusitas Masyarakat Madura. *RESIPROKAL: Jurnal Riset Sosiologi Progresif Aktual*, 5(1), 57–73. <https://doi.org/10.29303/resiprokal.v5i1.284>
- Herawati, Y. (2023). Makna Dan Nilai-Nilai Moral Dalam Sastra Daerah Tarsulan Perkawinan Di Kutai Kartanegara. *LOA: Jurnal Ketatabahasaan Dan Kesusastraan*, 18(1). <https://doi.org/10.26499/loa.v18i1.5891>
- Herizal, H. (2021). Penguatan Konsep Kombinatorika Siswa SMAN 1 Muara Batu Kabupaten Aceh Utara dalam Menghadapi KSN-K Bidang Matematika 2021. *Jurnal Pengabdian UNDIKMA*, 2(2), 174. <https://doi.org/10.33394/jpu.v2i2.4207>
- Hia, E., Amin Otoni Harefa, Yakin Niat Telaumbanua, & Sadiana Lase. (2025). Pengaruh Pendekatan Etnomatematika Terhadap Kemampuan Pemecahan Masalah Matematis Siswa UPTD SMP Negeri Lolofitu Moi. *Emasains: Jurnal Edukasi Matematika Dan Sains*, 14(1), 1–14. <https://doi.org/10.59672/emasains.v14i1.4555>
- Hidayati, F. N., & Prahmana, R. C. I. (2022). Ethnomathematics' research in Indonesia during 2015–2020. *Indonesian Journal of Ethnomathematics*, 1(1), 29–42. <http://doi.org/10.48135/ije.v1i1.29-42>
- Ibnu Fitrianto & Muhammad Farisi. (2025). Integrating Local Wisdom into 21st Century Skills: A Contextual Framework for Culturally Relevant Pedagogy in Rural Classrooms. *International Journal of Post Axial: Futuristic Teaching and Learning*, 109–121. <https://doi.org/10.59944/postaxial.v3i2.444>
- Irfan, M., Widodo, S. A., Sulistyowati, F., Arif, D. F., & Syaifuddin, M. W. (2023). Student Errors in Solving Higher Order Thinking Skills Problems: Bridge Context. *IndoMath: Indonesia Mathematics Education*, 5(2), 185–194. <https://indomath.org/index.php/indomath/article/view/39/pdf>
- Ishak, M. I. (2020). Makna Simbolik Dibalik Rumah Adat Masyarakat Waibalun Kecamatan Larantuka Kabupaten Flores Timur. *Jurnal Spektrum Komunikasi*, 8(1), 1–9. <https://doi.org/10.37826/spektrum.v8i1.58>
- Iswandi, F., Hakeng, L., & Widodo, A. (2025). Paham Ketuhanan Dalam Ritual “Barong Wae” di Manggarai, Nusa Tenggara Timur. *Jurnal Adat Dan Budaya Indonesia*, 7(1), 55–65. <https://doi.org/10.23887/jabi.v7i1.86869>

- Jainuri, M., Haryadi, B., Haryanto, H., & Zurweni, Z. (2025). Geometri Simbolik dalam Penataan Hidangan Tradisi Tabir Lamo: Pendekatan Etnomatematika Visual dan Implikasinya dalam Pendidikan Matematika. *JagoMIPA: Jurnal Pendidikan Matematika Dan IPA*, 5(2), 464–478. <https://doi.org/10.53299/jagomipa.v5i2.1555>
- Lamanna, L., Gea, M. M., & Batanero, C. (2022). Do secondary school students' strategies in solving permutation and combination problems change with instruction? *Canadian Journal of Science, Mathematics and Technology Education*, 22(3), 602–616. <https://doi.org/10.1007/s42330-022-00228-z>
- Lamawuran, Y. D., Tokan, F. B., & Ratumakin, P. A. K. L. (2023). Tradisi Bu'a Oring: Alternatif Penanganan Stunting Berbasis Kearifan Lokal Di Kabupaten Flores Timur. *Jurnal Agregasi : Aksi Reformasi Government Dalam Demokrasi*, 11(2), 112–128. <https://doi.org/10.34010/agregasi.v11i2.11228>
- Lathifah, I. J., Apriani, F., & Agustine, P. C. (2019). Pelatihan pembuatan bahan ajar untuk pembelajaran matematika dengan pendekatan matematika realistik Indonesia. *Transformasi: Jurnal Pengabdian Masyarakat*, 15(2), 85–94. <https://doi.org/10.20414/transformasi.v15i2.1255>
- Li, X., Kim, J. S., & Lee, T. J. (2021). Contribution of Supportive Local Communities to Sustainable Event Tourism. *Sustainability*, 13(14), 7853. <https://doi.org/10.3390/su13147853>
- Lu, Y. (2022). Analysis of “Harmony” and “Unity” Concepts in Indonesian and Chinese Culture. *MANDARINABLE : Journal of Chinese Studies*, 1(1), 10–16. <https://doi.org/10.20961/mandarinable.v1i1.326>
- Mahayanti, K., Suja, I. W., & Arnyana, I. B. P. (2024). Menjaga warisan leluhur: Keindahan tradisi Mapeed dalam Upacara Dewa Yadnya di Desa Jagapati. *Jurnal Pembelajaran, Bimbingan, Dan Pengelolaan Pendidikan*, 4(12), 25. <https://doi.org/10.17977/um065.v4.i12.2024.25>
- Malo, H. A. I., Rujā, I. N., & Perguna, L. A. (2022). Makna Simbolik dalam Tradisi Gerep Ruha di Desa Tenda. *Jurnal Ilmiah Ilmu Sosial*, 8(2), 208–219. <https://doi.org/10.23887/jiis.v8i2.53775>
- Masamah, U. (2019). Pengembangan Pembelajaran Matematika Dengan Pendekatan Etnomatematika Berbasis Budaya Lokal Kudus. *JURNAL PENDIDIKAN MATEMATIKA (KUDUS)*, 1(2). <https://doi.org/10.21043/jpm.v1i2.4882>
- Mauleto, K. (2025). Etnomatematika pada Tarian Te'o Renda. *Leibniz: Jurnal Matematika*, 5(01), 16–24. <https://doi.org/10.59632/leibniz.v5i01.480>
- Meliasari, W. O., & Wardana, M. D. K. (2023). Pengaruh Media Permainan Dakon Berbasis Etnomatematika terhadap Hasil Belajar Operasi Hitung Bilangan Cacah. *Emergent Journal of Educational Discoveries and Lifelong Learning (EJEDL)*, 3(1). <https://doi.org/10.47134/emergent.v3i1.13>
- Mubarok, A. R. (2024). Islam and Samin: Study of the Relationship between Religion, Beliefs, and Values of the Samin Community. *JUSS (Jurnal Sosial Soedirman)*, 7(2), 258. <https://doi.org/10.20884/juss.v7i2.13118>
- Muhammad Yusuf, Sri Rahayu, & Muhammad Amin. (2023). Tradisi Baritan Masyarakat Muslim Pedesaan Wonosobo: Dialektika Komunitas-Struktur Perspektif Victor Turner. *Bulletin of Indonesian Islamic Studies*, 2(2), 147–168. <https://doi.org/10.51214/biis.v2i2.597>
- Muliadi, E. (2019). MASYARAKAT BERWAWASAN LINGKUNGAN DALAM KONSEP TRADISI MASYARAKAT ISLAM WETU TELU.docx. *Jurnal Penelitian Tarbawi*, 3(2), 23–49. <https://doi.org/10.37216/tarbawi.v3i2.155>
- Mu'min, U. A. (2020). Spiritualitas Karakter Tuang dalam Budaya Masyarakat Kampung Adat Cireunde. *Islamadina : Jurnal Pemikiran Islam*, 53. <https://doi.org/10.30595/islamadina.v0i0.6325>
- Nopriana, T., Hermanand, T., Suryadi, D., & Martadiputra, B. A. P. (2023). Combinatorics problems for undergraduate students: Identification of epistemological obstacles and model of combinatorial thinking. *AIP Conference Proceedings*, 2811(1), 020021. <https://doi.org/10.1063/5.0142261>
- Nursisto, M., & Ismailmuza, D. (2022). ANALISIS KESALAHAN SISWA DALAM MENYELESAIKAN MASALAH PERMUTASI DAN KOMBINASI DI SMAN MODEL TERPADU MADANI PALU. *Jurnal Elektronik Pendidikan Matematika Tadulako*, 10(1), 76–83. <https://jurnal.fkipuntad.com/index.php/jpmt/article/view/2579>
- Nuryadi, N., & Kholifa, I. (2020). Etnomatematika: Eksplorasi gamelan Jawa karawitan dengan pendekatan science, technology, engineering, and mathematics (STEM). *Jurnal Pendidikan Surya Edukasi (JPSE)*, 6(2), 140–148. <https://doi.org/10.37729/jpse.v6i2.6810>
- OECD. (2023). *PISA 2022 Results Indonesia*. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbc5-en/indonesia_c2e1ae0e-en.html
- Payadnya, I. P. A. A., Prahmana, R. C. I., Lo, J.-J., Noviyanti, P. L., & Atmaja, I. M. D. (2023). Designing area of circle learning trajectory based on “what-if” questions to support students' higher-order thinking skills. *Journal on Mathematics Education*, 14(4), 757–780. <https://doi.org/10.22342/jme.v14i4.pp757-780>
- Prahmana, R. C. I., Arnal-Palacián, M., Risdianti, I., & Ramadhani, R. (2023). Trivium curriculum in Ethno-RME approach: An impactful insight from ethnomathematics and realistic mathematics education. *Jurnal Elemen*, 9(1), 298–316. <https://doi.org/10.29408/jel.v9i1.7262>
- Prahmana, R. C. I., & D'Ambrosio, U. (2020). Learning Geometry And Values From Patterns: Ethnomathematics On The Batik Patterns Of Yogyakarta, Indonesia. *Journal on Mathematics Education*, 11(3), 439–456. <https://doi.org/10.22342/jme.11.3.12949.439-456>
- Putra, A., Zulkardi, Z., Putri, R. I. I., Laswadi, L., & Nusantara, D. S. (2023). Scoping Literature Review: What Activities Can Help Students Discover Permutations? *Edumatika: Jurnal Riset Pendidikan Matematika*, 6(2), 105–117. <https://doi.org/10.32939/ejrpm.v6i2.3167>
- Putri, A. H., Maulidini, N. E., Wardani, S., Susanti, T. A., Setyawan, K. G., & Sugiantoro, S. (2024). Tradisi Tahunan Grebeg Apem di Kabupaten Jombang Sebagai Simbol Kebersamaan dan Keberkahan. *Maharsi*, 6(3), 78–86. <https://doi.org/10.33503/maharsi.v6i3.45>
- Putri, S. D., Satria, I., & Heriadi, M. (2025). Makna Festival Budaya Umbung Kuteui Masyarakat Suku Rejang Kabupaten Kepahiang: (Kajian Semiotika). *Jurnal Penelitian Ilmu Pendidikan Indonesia*, 4(1), 262–270. <https://doi.org/10.31004/jpion.v4i1.371>

- Rachmawati, F., & Purwaningrum, J. P. (2019). Model discovery learning berbasis etnomatematika pada bangun ruang untuk Menumbuhkan kemampuan Literasi dan Karakter Nasionalisme Pada Generasi Z 4.0. *AKSIOMA : Jurnal Matematika Dan Pendidikan Matematika*, 10(2), 254–260. <https://doi.org/10.26877/aks.v10i2.4837>
- Ridho'i, M. (2024). Eksplorasi Unsur Geometri dalam Arsitektur Masjid Agung KH Anas Mahfud sebagai Sumber Belajar Berbasis Etnomatematika. *Jurnal Ilmu Pendidikan Muhammadiyah Kramat Jati*, 5(1), 383–390. <https://doi.org/10.55943/jipmukjt.v5i1.364>
- Rodríguez-Nieto, C. A., Escobar-Ramírez, Y. C., Font Moll, V., & Aroca Araújo, A. (2023). Ethnomathematical and Mathematical Connections Activated by a Teacher in Mathematical Problems Posing and Solving. *Acta Scientiae*, 25(1), 86–121. <https://doi.org/10.17648/acta.scientiae.7356>
- Rohman, A. D., Afiah, K., Riayana, R., & Huda, M. F. (2024). Nyadran: Tradisi Penghormatan Leluhur dalam Bingkai Nilai-Nilai Islam di Dusun Silawan Desa Kutorjo. *PRAXIS: Jurnal Pengabdian Kepada Masyarakat*, 2(3), 171–176. <https://doi.org/10.47776/praxis.v2i3.1018>
- Rosa, M., & Clark Orey, D. (2019). Etnomatemática e a subversão responsável de sua ação pedagógica: Uma investigação apoiada em três abordagens antropológicas. *Revista Brasileira de Estudos Pedagógicos*, 100(254). <https://doi.org/10.24109/2176-6681.rbep.100i254.3939>
- Rosa, M., D'Ambrosio, U., Orey, D. C., Shirley, L., Alanguí, W. V., Palhares, P., & Gavarrete, M. E. (2016). *Current and Future Perspectives of Ethnomathematics as a Program*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-30120-4>
- Rudhito, M. A., Kristanti, R., & Syahdana, H. (2024). Kajian Etnomatematika dalam Budaya Merti Dusun Temanggung Kabupaten Wonosobo dan Implementasinya dalam Pembelajaran Matematika. *Postulat : Jurnal Inovasi Pendidikan Matematika*, 5(1), 65. <https://doi.org/10.30587/postulat.v5i1.7628>
- Sabon, Y. O. S., Putro, N. H. P. S., & Rahim, A. (2021). Etnomatematika Dan Nilai Karakter Dalam Permainan Tradisional Kebetuk. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(4), 2079. <https://doi.org/10.24127/ajpm.v10i4.3550>
- Safitri, P. I., Zuriyati, Z., & Rahman, S. (2022). Peribahasa Masyarakat Jawa Sebagai Cermin Kepribadian Perempuan Jawa. *Lingua Rima: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(3), 211. <https://doi.org/10.31000/lgrm.v11i3.7307>
- Saminanto, S., Suwarno, M., Septianah, S., & Aistafania, A. (2024). Proses Penalaran Matematis Siswa SMP dalam Statistika. *Square : Journal of Mathematics and Mathematics Education*, 6(2), 159–169. <https://doi.org/10.21580/square.2024.6.2.23575>
- Sari, T. A. M., Sholehatur, A. N., Rahma, S. A., & Prasetyo, R. B. (2021). Eksplorasi Etnomatematika pada Seni Batik Madura dalam Pembelajaran Geometri. *Journal of Instructional Mathematics*, 2(2), 71–77. <https://doi.org/10.37640/jim.v2i2.1032>
- Sekar Kirana Wulandari, Lestari, A., Nadila Putri Budi Sugiarti, Siti Komariah, & Mirna Nur Alia Abdullah. (2024). Filosofi Jenang Procot sebagai Makanan Khas Upacara Tingkeban Masyarakat Suku Jawa. *SOSMANIORA: Jurnal Ilmu Sosial Dan Humaniora*, 3(2), 256–261. <https://doi.org/10.55123/sosmaniora.v3i2.3529>
- Selmi, S., Gasong, D., & Dewi, R. (2024). Interferensi Bahasa Indonesia ke dalam Bahasa Toraja Serta Dampaknya Terhadap Identitas Budaya Toraja di Lingkungan Tallunglipu Kabupaten Toraja Utara. *Indonesian Research Journal on Education*, 4(3), 887–894. <https://doi.org/10.31004/irje.v4i3.903>
- Sembiring, J. M. M., Fitriani, E., & Febrianto, A. (2022). Makna Upacara Utang Tahunan Masyarakat Melayu Indragiri Hulu. *Culture & Society: Journal Of Anthropological Research*, 4(2), 124–139. <https://doi.org/10.24036/csjar.v4i2.113>
- Senoprabowo, A., Khamadi, K., & Septian, Y. A. (2021). Komik Digital Warak Ngendog Untuk Memperkenalkan Nilai Kearifan Lokal Kepada Anak Di Kota Semarang. *Prosiding Seminar Nasional Desain Komunikasi Visual*, 1, 1–13. <https://doi.org/10.33479/sndkv.v1i.76>
- Sheppard, A., & Broughton, M. C. (2020). Promoting wellbeing and health through active participation in music and dance: A systematic review. *International Journal of Qualitative Studies on Health and Well-Being*, 15(1), 1732526. <https://doi.org/10.1080/17482631.2020.1732526>
- Silalahi, F. H. M., & Purwanto, E. (2025). Sacred Harmony: Exploring Pikukuh Tilu Philosophy in the Spiritual, Social, and Environmental Practices of the Baduy People. *Pharos Journal of Theology*, 106.2. <https://doi.org/10.46222/pharosjot.106.2022>
- Silalahi, L. C., Rizal, Muh., & Sugita, G. (2020). Analisis Kemampuan Spasial Siswa Berkemampuan Matematika Tinggi Dalam Menyelesaikan Masalah Geometri Bangun Ruang Sisi Datar. *Aksioma*, 9(2), 112–125. <https://doi.org/10.22487/aksioma.v9i2.521>
- Sulistiyowati, F., Istiqomah, I., Putri, N. A. H., Septiani, D., Kusumaningrum, B., & Kuncoro, K. S. (2024). *Interactive e-learning using web-wix for joyful learning linear equations*. 020033. <https://doi.org/10.1063/5.0194657>
- Sulistiyowati, F., Kuncoro, K. S., Nugraheni, P., Hernowo, H., & Setyawan, F. (2019). The problems of teaching fractional arithmetic operations for disabled student using Realistic Mathematics Education. *Journal of Physics: Conference Series*, 1188(1), 012030. <https://doi.org/10.1088/1742-6596/1188/1/012030>
- Sulistiyowati, F., Kuncoro, K. S., Setiana, D. S., & Purwoko, R. Y. (2019). Solving high order thinking problem with a different way in trigonometry. *Journal of Physics: Conference Series*, 1315(1), 12001. <https://doi.org/10.1088/1742-6596/1315/1/012001>
- Sunarno, & Rahmawati Zakiyah, A. (2024). Modal Sosial Masyarakat Multikultural: Persepsi Komunitas di Dusun Sumberjo, Kabupaten Kediri. *MEDIAPSI*, 10(2), 302–317. <https://doi.org/10.21776/ub.mps.2024.0010.02.961>
- Sutra, S., Nur, F., & Yuliani, N. (2025). Eksplorasi Etnomatematika Konsep Geometri Pada Kue Apang Makanan Tradisional Suku Bugis. *JUPIKA: JURNAL PENDIDIKAN MATEMATIKA*, 8(1), 1–8. <https://doi.org/10.37478/jupika.v8i1.5218>

- Tamutu, N. A., & Hasmah, H. (2024). Eksistensi Seni Kerajinan Dari Limbah Ban Bekas Di Desa Pentadio Timur Kecamatan Telaga Biru Kabupaten Gorontalo. *Jambura: Jurnal Seni Dan Desain*, 3(2), 63–72. <https://doi.org/10.37905/jjsd.v3i2.21720>
- Trisnawati, T. (2022). Pengembangan Bahan Ajar Interaktif dengan Pendekatan Etnomatematika Berbasis Budaya Lokal di Banten pada Pokok Bahasan Barisan dan Deret untuk Siswa SMP. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 6(1), 282–290. <https://doi.org/10.31004/cendekia.v6i1.739>
- Triyono, R., Putri, D. A., Ardiningrum, D., Nuramalia, S., Patmawati, H., & Amarulloh, S. I. (2025). Eksplorasi Etnomatematika: Konsep Simetri Pada Pola Gerak Tari Merak. *De Fermat : Jurnal Pendidikan Matematika*, 7(2), 67–82. <https://doi.org/10.36277/deferemat.v7i2.2237>
- Untara, I. M. G. S. (2025). Transformasi Ilmu Wariga Dalam Masyarakat Adat Buleleng Antara Tradisi Dan Modernitas. *Jurnal Penelitian Agama Hindu*, 9(3), 88–107. <https://doi.org/10.37329/jpah.v9i3.4216>
- Utmeemang, R., & Buaraphan, K. (2024). Effects of small-scale chemistry STEM integrated with local contexts for enhancing grade 11 students' learning achievement and learning and innovation skills. *Ecletica Quimica*, 49. <https://doi.org/10.26850/1678-4618.eq.v49.2024.e1561>
- Vina Tri Agustinningrum & Sukarman Sukarman. (2024). Tradisi Ganjuran Di Desa Ngambeg Kecamatan Pucuk Kabupaten Lamongan (Teori Folklor). *Morfologi: Jurnal Ilmu Pendidikan, Bahasa, Sastra Dan Budaya*, 2(5), 120–138. <https://doi.org/10.61132/morfologi.v2i5.927>
- Windasari, I. Y., Prasetyowati, D., & Shodiqin, A. (2020). Analisis Pemahaman Konsep Berdasarkan Teori Apos pada Materi Barisan Geometri di Kelas XI SMA Negeri 1 Godong. *Imajiner: Jurnal Matematika Dan Pendidikan Matematika*, 2(5), 417–427. <https://doi.org/10.26877/imajiner.v2i5.6664>
- Wulanjari, T. P., Kudubun, E. E., & Siahainenia, R. R. (2024). Makna Tradisi Sadranan Bagi Masyarakat di Dukuh Dungus dalam Prespektif Interaksi Simbolik Geogre Herbert Mead. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, 7(6), 6076–6084. <https://doi.org/10.54371/jiip.v7i6.4905>
- Yanich, M., Maykova, V., Pesotzky, V., & Molchan, E. (2019). Spiritual and Moral Values as a System-forming Factor of Social Systems. *Proceedings of the 2nd International Conference on Contemporary Education, Social Sciences and Ecological Studies (CESSSES 2019)*. Proceedings of the 2nd International Conference on Contemporary Education, Social Sciences and Ecological Studies (CESSSES 2019), Moscow, Russia. <https://doi.org/10.2991/cesses-19.2019.264>
- Yogasmara Binsar Putra Wimansyah, Ida Ayu Alit Laksmiwati, & Ida Bagus Oka Wedasantara. (2025). Etnomatematika Weti pada Sa'o di Kampung Adat Gurusina, Kabupaten Ngada, Provinsi Nusa Tenggara Timur. *Dharma Acariya Nusantara: Jurnal Pendidikan, Bahasa Dan Budaya*, 3(1), 104–118. <https://doi.org/10.47861/jdan.v3i1.1816>
- Yolan, Suparman, S., Besse Herdiana, & Muhammad Nuruahmad. (2024). Kemampuan Menulis Karangan Argumentasi Dengan Menggunakan Media Gambar Pada Siswa Kelas Vii Smp Negri Ii Walenrang. *Jurnal Vokatif: Pendidikan Bahasa, Kebahasaan, Dan Sastra*, 1(2), 102–108. <https://doi.org/10.51574/vokatif.v1i2.1736>
- Yulianto, J. E., Hodgetts, D., King, P., & Liu, J. H. (2022). The assemblage of inter-ethnic marriages in Indonesia. *Journal of Community & Applied Social Psychology*, 32(4), 706–720. <https://doi.org/10.1002/casp.2587>

Development of a Physics Diagnostic Assessment Framework based on Solo Taxonomy for Senior High School Student

Sabani¹, Irham Ramadhani², Wawan Bunawan³, Seila Agustina⁴, Elang Mulia Lesmana⁵, Dyah Inggit Murtinigrum⁶

^{1,2,3}Physics Education, Universitas Negeri Medan, Medan, Indonesia

⁴Bilingual Physics Education, Universitas Negeri Medan, Medan, Indonesia

^{5,6}Teacher Professional Education, Universitas Negeri Medan, Medan, Indonesia

¹sabani@unimed.ac.id, ²irhamramadhani@unimed.ac.id, ³wawanbunawan@unimed.ac.id,

⁴seilaagustina@mhs.unimed.ac.id, ⁵elang.mulia179@guru.sma.belajar.id,

⁶dyah.inggit14@guru.sma.belajar.id

ABSTRACT

The paradigm shift toward Deep Learning in Indonesia demands assessment tools that capture the quality of students' cognitive structures rather than mere procedural memorization. This study aims to develop and validate a physics diagnostic test framework based on the SOLO taxonomy to evaluate high school students' understanding of fluid concepts. Utilizing a Research and Development (R&D) design adapted from the Borg and Gall model, the study involved 80 eleventh-grade students in Medan. Data were analyzed using Classical Test Theory via SPSS and Microsoft Excel. Psychometric evaluations optimized the initial test instrument diagnostic based on SOLO Taxonomy into 12 valid items with validity coefficients ranging from 0.480 to 0.815. The final instrument demonstrated excellent internal consistency with a reliability coefficient of 0.887. Item discrimination was mostly very good (78.57%), with a balanced difficulty distribution (28.57% easy, 57.14% moderate, 14.28% difficult). The final framework successfully mapped cognitive structures across unistructural, multistructural, relational, and extended abstract levels. This valid and reliable diagnostic tool effectively measures students' structural understanding to support the deep learning paradigm.

Keywords: SOLO Taxonomy, Diagnostic Instrument, Fluid Concepts, Physics Assessment.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Irham Ramadhani,

Physics Education,

Universitas Negeri Medan,

Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang 20371, Indonesia.

irhamramadhani@unimed.ac.id

1. INTRODUCTION

Physics learning at the Senior High School (SHS) level aims to equip students with an understanding of scientific concepts related to natural phenomena. One of the essential topics is fluids, which include static and dynamic fluids and are closely related to everyday life, such as the working principles of pumps, buoyant force, blood pressure, and water flow. The topic of fluids is important because it provides a fundamental understanding of how pressure, buoyancy, and fluid motion govern various natural phenomena and technological applications. Understanding fluid concepts enables students to explain and analyze real-world phenomena, such as the operation of hydraulic systems, water distribution, aircraft lift, and the circulation of blood in the human body. Therefore, learning fluids is essential not only for developing students' conceptual understanding of physics but also for connecting scientific principles with practical applications in everyday life.

Ideally, physics learning should not only emphasize formulas and problem-solving procedures but also develop students' scientific thinking and deep conceptual understanding. However, various studies indicate that students' scientific literacy and conceptual understanding, particularly in fluid topics, remain low. Data from the Programme for International Student Assessment (PISA) 2022 show that only about 34% of Indonesian students reached level 2 or higher in scientific literacy, far below the OECD average of 76%. This suggests that many students still struggle to interpret scientifically, reasoning.

These difficulties are also evident in students' understanding of static fluid concepts such as hydrostatic pressure, pascal's law, and archimedes' principle. Students are generally able to recall and apply formulas but struggle to integrate concepts into a coherent framework. From the perspective of the SOLO taxonomy (structure of the observed learning outcomes), students' understanding tends to

remain at the unistructural and multistructural levels, where concepts are understood separately without meaningful integration. previous studies have reported significant levels of misconceptions in fluid topics, such as hydrostatic pressure (16.19%), pascal's law (39.04%), and archimedes' principle (45.71%) (Ahmadi et al., 2024). Other studies also found that up to 88.5% of students experience misconceptions in static fluid materials (Atmaja & Samsudin, 2024; Hutasoit et al., 2025; Indrasari, 2025; Muki et al., 2025; Putri & Nasbey, 2023; Kurniawan, 2023; Koto & Ilhami, 2023; Fatmawati et al., 2025). This condition indicates that conventional assessments, which mainly focus on memorization and procedural skills, are insufficient to reveal the depth of students' conceptual understanding.

Initially observations through a survey revealed that most students actually demonstrate a tendency toward conceptual and analytical thinking. Approximately 76.7% of students reported that they understand physics concepts better through relationships between concepts rather than memorization, and 80% feel challenged by multi-step problems. Additionally, 76.6% stated that higher-order thinking activities improve their understanding. However, nearly half of the students believe that current assessments still emphasize memorization and do not fully reflect their thinking abilities. This indicates a mismatch between students' cognitive characteristics and the forms of assessment used in schools. one potential solution is the implementation of diagnostic tests based on the SOLO taxonomy. this taxonomy, developed by Biggs and Collis, classifies students' levels of understanding based on the complexity of their responses, ranging from prestructural, unistructural, and multistructural to relational and extended abstract levels (Biggs & Collis, 1982). Unlike bloom's taxonomy, which focuses on cognitive processes, the SOLO taxonomy emphasizes the quality and structure of understanding (Fernandez & Guzon, 2025). Therefore, SOLO-based diagnostic tests are more suitable for identifying students' conceptual development levels.

The urgency of developing assessment instruments capable of measuring depth of understanding is increasingly relevant in the context of national education transformation, particularly in the implementation of the Deep Learning paradigm, which is currently the policy direction for improving the quality of learning in Indonesia. Deep learning emphasizes a meaningful, conscious, and enjoyable learning process through the active involvement of students in acquiring knowledge, connecting various concepts, and transferring understanding to new situations (Sabani et al., 2026). In this context, learning success is no longer simply measured by the ability to recall information or solve routine problems, but rather by students' ability to demonstrate integrated conceptual understanding, conduct scientific reasoning, and apply concepts in various real-life contexts (Marton & Saljo, 1976; Biggs, 1987; Ministry of Education and Culture, 2025). The Deep Learning academic paper published by the Ministry of Primary and Secondary Education also emphasizes that this approach requires students to understand, apply, and reflect on knowledge meaningfully, rather than simply memorizing information.

This need is also reflected in the assessment system for the Teacher Professional Education Program (PPG), both the PPG for Specific Teachers and the PPG for Prospective Teachers. In recent years, evaluations of the implementation of the PPG (School Teacher Education Program) have shown that some assessment instruments used are still oriented toward measuring material mastery and procedural skills, while aspects such as depth of conceptual understanding, the ability to integrate various concepts, and the ability to reflect and transfer knowledge have not been optimally measured (Brookhrat, 2010; UNESCO, 2021). Prospective and professional teachers are required not only to master learning content but also to possess higher-order thinking skills that enable them to analyze student misconceptions, design meaningful learning, and develop adaptive learning strategies for various classroom contexts. The in-depth learning orientation emphasizes holistic understanding and the interconnectedness of concepts, further emphasizing the importance of assessments that capture the quality of this understanding (Biggs & Collis, 1982; Biggs & Tang, 2011).

In PPG assessments, participants are often challenged to demonstrate integrated pedagogical and professional skills. However, available instruments generally fail to map in detail the level of development of participants' understanding, from partial understanding to the ability to integrate various concepts and generate new generalizations. As a result, assessment results often only quantitatively describe the level of material mastery without providing diagnostic information about the cognitive structures underlying participants' responses. This situation presents a unique challenge because teachers, who will be tasked with developing critical thinking skills and in-depth understanding in students, must first demonstrate the characteristics of this understanding in their professional education

process. In this perspective, the SOLO Taxonomy provides a highly relevant framework to support in-depth learning-based assessment. The SOLO levels allow for the identification of the development of understanding quality ranging from pre-structural to extended abstract, thus providing a more accurate picture of the development of an individual's ability to connect, integrate, and generalize concepts (Biggs & Collis, 1982; Biggs & Tang, 2011). The alignment between learning principles and the characteristics of the SOLO Taxonomy makes this approach potentially a foundation for developing assessment instruments that are not only relevant for students in schools but also for PPG participants. Thus, the development of a SOLO Taxonomy-based diagnostic instrument for Fluids not only addresses the need for conceptual assessment at the high school level but also contributes to the development of assessment models that align with the demands of in-depth learning and the strengthening of professional teacher competencies in Indonesia.

Previous studies have shown that SOLO based instruments are effective in identifying misconceptions and students' thinking patterns in physics. for example, studies on fluid topics have demonstrated that SOLO based diagnostic tests can reveal misconceptions in hydrostatic pressure and bernoulli's principle (Ahmadi et al., 2024; Hartanti et al., 2021). however, most of these studies focus primarily on identifying misconceptions rather than developing and validating instruments that meet psychometric criteria such as validity, reliability, discrimination power, and difficulty level (Ole & Gallos, 2023; Sedhandoyo, 2024; Eryanto & Prasetyono, 2025; Guzon & Fernandez, 2025; Yuberti et al., 2021; Inggit et al., 2021). Furthermore, limited research integrates students' thinking characteristics into the design of SOLO based diagnostic instruments.

Based on these problems, this study aims to develop a diagnostic test instrument based on the SOLO taxonomy for fluid topics in senior high school and to evaluate its quality using classical test theory, including validity, reliability, discrimination power, and difficulty level. This study is expected to contribute to the development of assessment tools that can measure students' cognitive structures more comprehensively. The results of this study are expected to provide several benefits. For students, the instrument can enhance higher-order thinking skills and deepen conceptual understanding, for teachers, it can serve as an alternative assessment tool that provides more diagnostic and meaningful information about students' learning. For schools, it can support the improvement of physics learning quality. Additionally, this study can serve as a reference for future researchers in developing SOLO-based assessment frameworks in different contexts.

2. RESEARCH METHOD

This study employed a research and development (R&D) design aimed at developing and validating a diagnostic test instrument based on the SOLO taxonomy for fluid topics in senior high school physics. The development procedure followed the Borg and Gall model adapted by Sugiyono (2023), which consists of ten stages: (1) Identifying potential and problems, (2) Data collection, (3) Preliminary product development, (4) Design validation, (5) Design revision, (6) Limited trial, (7) Product revision, (8) Field testing, (9) Final product revision, and (10) Mass production. The research was conducted in April 2026 at a public high school in Medan, North Sumatra. This location was purposively selected because SOLO taxonomy-based assessments had not been previously implemented there, and the students demonstrated adequate academic readiness.

The subjects included 80 students selected using cluster sampling techniques, divided into three classes with first class 20 subjects, each 10 students as subject for first preliminary field testing and second preliminary field testing, second and three class 30 students for each class as wider field testing. The variables measured in this study included validity, reliability, item difficulty, and discrimination power. the developed instrument was a diagnostic test consisting of 17 multiple-choice items with various representations, including verbal, numerical, graphical, and visual forms, designed to assess students' conceptual understanding based on SOLO levels (S2–S5).

Data collection techniques included literature study and test administration. The instrument development process involved expert validation by two lecturers (assessment expert and physics expert), followed by a first preliminary field test with 10 students, second preliminary field test with 10 students, and wider field test with 60 students. the tools used in data processing included microsoft excel for discrimination power, and SPSS for statistical analysis of validity, reliability, and item difficulty.

Data analysis was conducted using both qualitative and quantitative approaches. Qualitative analysis involved expert judgment focusing on content, construction, and language validity using content validity ratio (CVR). The steps of the research procedure can be seen in Fig 1 below.

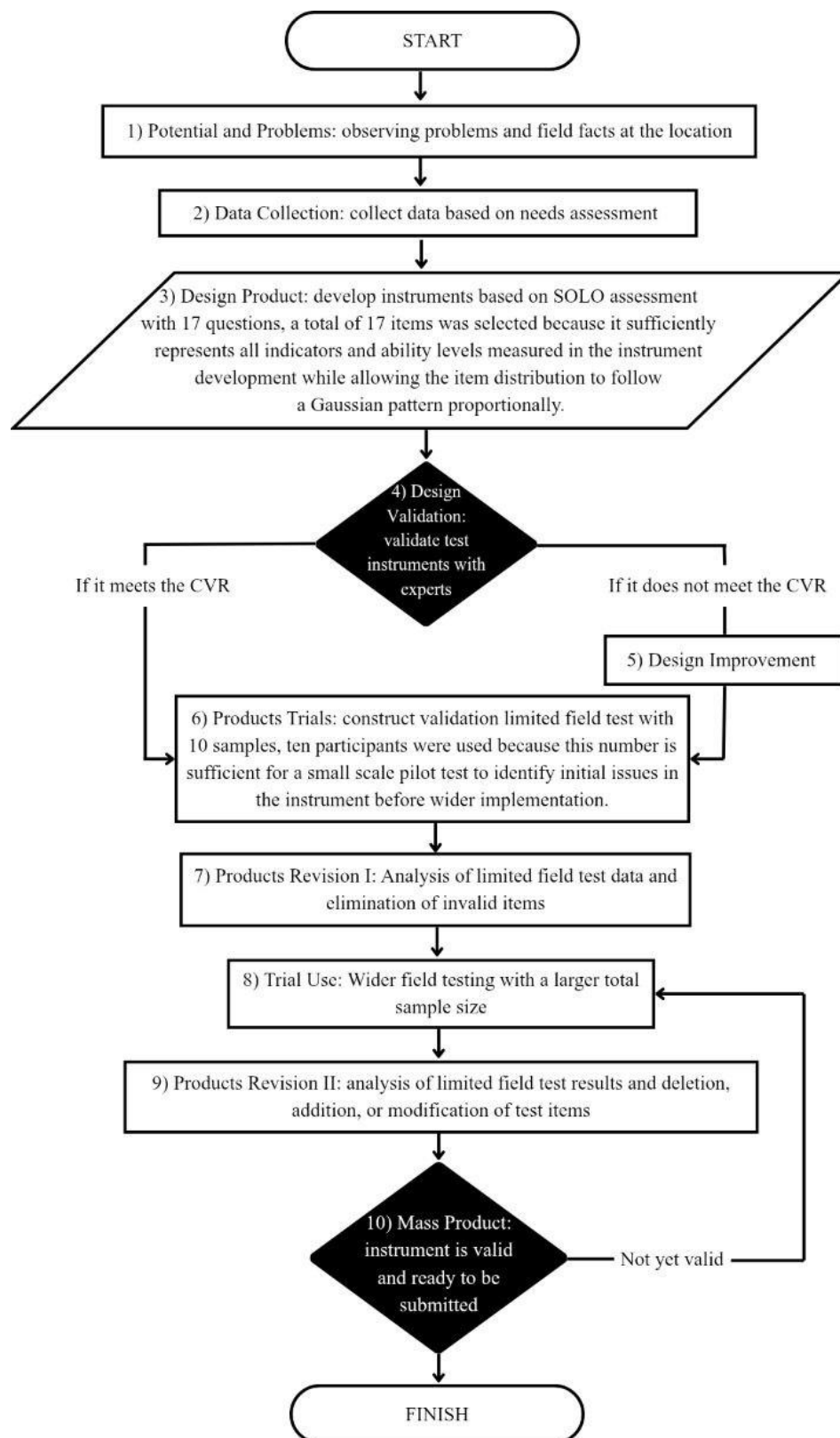


Fig 1. Research Procedure Steps

3. RESULTS AND DISCUSSION

In this instrument development study, the stages were carried out following the Borg and Gall procedure, starting from identifying potential and problems to producing a final product that was feasible for use. At the initial stage (potential and problem), it was found that the school already had assessment instruments to measure students' cognitive understanding; however, these instruments did not specifically classify the structure of understanding based on the SOLO Taxonomy levels. Therefore, an additional instrument was needed as a complement to provide a more structured representation of students' levels of understanding. At the data collection stage, literature studies and needs analysis were conducted to develop indicators and test blueprints.

A. Validation Results by Experts

At the product design stage, an initial instrument consisting of 17 items based on SOLO levels was developed. The instrument then underwent content validation by two experts, resulting in revisions that reduced the items to 14, all of which were deemed suitable for the trial. The product trial stage was conducted through the first preliminary field testing involving 10 students, and the results of the analysis showed that 8 items were invalid, namely items number 1, 2, 3, 5, 8, 9, 10, and 13, that have been studied by experts are processed using the CVR index with a scale of 1, 2, and 3, where if the conclusion of E/ Essential (can be used without revision) is on a scale of 3, ET/ Almost Essential (can be used with minor revisions) on a scale of 2, and T/ Not Essential (rejected) on a scale of 1.

If only 1 validator accepts:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} = \frac{1 - \frac{2}{2}}{\frac{2}{2}} = 0$$

If there are 2 validators who accept:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} = \frac{2 - \frac{2}{2}}{\frac{2}{2}} = 1$$

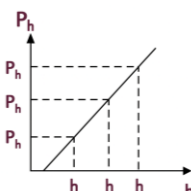
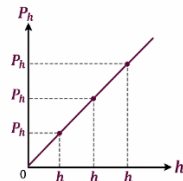
Because the calculation is valid based on the formula, there must be 2 validators who accept it so that one question item is declared valid. The categories of test item analysis results qualitatively presented in Table 1.

Table 1. Item Analysis Results Qualitatively

Category	Question Number	Questions Total
Accepted	1, 12, 14, 15, 16, 17.	6 quesrions
Revised	2, 3, 4, 5, 7, 8, 9, 10.	8 questions
Rejected	6, 11, 13.	3 questions

There were 3 items that did not meet the required criteria (rejected), namely item number 6 at SOLO level 5, as well as items number 11 and 13 at SOLO level 4. Thus, from the total items developed, 14 items remained and were considered suitable for use, with a distribution of 2 items at SOLO level 2, 5 items at SOLO level 3, 3 items at SOLO level 4, and 4 items at SOLO level 5. The rejection of these items was based on expert validation results. Item number 6 was rejected because the image provided was not consistent with the relevant theoretical concept, which could potentially lead to misconceptions. Meanwhile, items number 11 and 13 were rejected because they were considered illogical in terms of context and problem structure, making them unsuitable as instruments for measuring students' cognitive abilities. Therefore, these items were eliminated to ensure that the instrument met quality standards and could accurately assess students' understanding. After this stage, the researcher has 14 questions left that will be tested on students empirically with the order of the question numbers raised and rearranged after there are three questions that are deleted. Table 2 shows examples of question instruments before and after the revision.

Table 2. Revision of Test Instruments After Expert Validation

Initial Questions	Suggestions for Improvement	Questions After Revision
Grafik hubungan antara tekanan hidrostatik (P_h) dan kedalaman (h) disajikan sebagai berikut.  Berdasarkan grafik di atas, disimpulkan bahwa: 1. Antara tekanan hidrostatik (P_h) dan kedalaman (h) sebanding. 2. Tekanan hidrostatik besar (P_h) relatif terhadap lurus ke kedalaman (h). 3. Hubungan tekanan hidrostatik (P_h) dengan kedalaman (h) adalah linier. 4. Hubungan tekanan hidrostatik (P_h) dengan kedalaman (h) tidak linier. <i>Opsi:</i> A. Pernyataan 1, 2, dan 3 adalah benar. Pernyataan 4 salah. B. Pernyataan 1 dan 3 benar. Pernyataan 2 dan 4 adalah salah. C. Pernyataan 1 dan 3 adalah salah. Pernyataan 2 dan 4 adalah benar D. Pernyataan 4 benar. E. Tidak ada pernyataan yang benar.	Fix the graph so that the lines are linear from point 0, change statement 1 to make it more difficult.	Grafik hubungan antara tekanan hidrostatik (P_h) dan kedalaman (h) disajikan sebagai berikut.  Berdasarkan grafik di atas, disimpulkan bahwa: 1. Tekanan yang dialami suatu titik akan semakin besar jika titik tersebut berada pada kedalaman yang lebih tinggi dari permukaan fluida.. 2. Tekanan hidrostatik besar (P_h) berbanding lurus dengan kedalaman (h). 3. Hubungan tekanan hidrostatik (P_h) dengan kedalaman (h) adalah linier. 4. Hubungan tekanan hidrostatik (P_h) dengan kedalaman (h) tidak linier. Berdasarkan pernyataan di atas, manakah opsi yang benar di bawah ini? <i>Opsi:</i> A. Pernyataan 1, 2, dan 3 benar. Pernyataan 4 salah. B. Pernyataan 1 dan 3 benar. Pernyataan 2 dan 4 salah. C. Pernyataan 1 dan 3 salah. Pernyataan 2 dan 4 benar D. Pernyataan 4 benar. E. Tidak ada pernyataan yang benar.

B. Validity Pearson Correlation

In line with classical test theory, item validity was determined by comparing the calculated r -value (r -count) against the critical r -table value, alongside a significance threshold of $p < 0.05$. Following the first preliminary field testing, the analysis revealed that seven items had a very low correlation, two items had a low correlation, three items had a moderate correlation, and two items had a high correlation. Based on these results, the poorly performing items were removed, and a renumbering process was conducted to ensure that the deleted items did not reappear in subsequent analyses. The full validity interpretation for this first preliminary field testing is presented in Table 3.

Table 3. The interpretation validity for first preliminary field testing

Number of Questions	Pearson	Nilai Sig.	Summary	Interpretation
Q1	0	0	Not Valid	Very Low
Q2	-0.155	0.699	Not Valid	Very Low

Q3	0.203	0.573	Not Valid	Low
Q4	0.583	0.077	Valid	Enough
Q5	0	0	Not Valid	Very Low
Q6	0.62	0.056	Valid	High
Q7	0.785	0.007	Valid	High
Q8	-0.155	0.669	Not Valid	Very Low
Q9	0.155	0.669	Not Valid	Very Low
Q10	0	0	Not Valid	Very Low
Q11	0.807	0.005	Valid	Very Low
Q12	0.581	0.078	Valid	Enough
Q13	0.349	0.323	Not Valid	Low
Q14	0.426	0.219	Valid	Enough

The preliminary field testing that was carried out at first still had many invalid items, namely questions number 1, 2, 3, 5, 8, 9, 10, and 13. Therefore these questions are revised and then retested. However, because the limited test only used a small subject of 10, the average significant result could be above 0.05. The resulting correlation values are compared with the r table. The interpretation validity result for second preliminary field testing in Table 4.

Table 4. The interpretation validity for Second preliminary field testing

Number of Questions	Pearson	Nilai Sig.	Summary	Interpretation
Q1	0.583	0.077	Valid	Enough
Q2	0.653	0.041	Valid	High
Q3	0.515	0.128	Valid	Enough
Q4	0.730	0.016	Valid	High
Q5	0.515	0.128	Valid	Enough
Q6	0.730	0.016	Valid	High
Q7	0.620	0.056	Valid	High
Q8	0.408	0.242	Valid	Enough
Q9	0.467	0.173	Valid	Enough
Q10	0.454	0.188	Valid	Enough
Q11	0.515	0.128	Valid	Enough
Q12	0.678	0.031	Valid	High
Q13	0.515	0.128	Valid	Enough
Q14	0.675	0.032	Valid	High

The revision carried out was able to improve the quality of the instrument, so that the instrument was declared suitable for collecting field test research data more widely, the main field testing used 14 questions that had been revised based on the results of preliminary field testing. A main field testing was carried out in class XI with 60 subjects. The interpretation validity result for main field testing in table 5.

Table 5. The interpretation validity for Main Field Testing

Number of Questions	Pearson	Nilai Sig.	Summary	Interpretation
Q1	0.547	0.000	Valid	Enough
Q2	0.584	0.000	Valid	Enough
Q3	0.692	0.000	Valid	High
Q4	0.815	0.000	Valid	High
Q5	0.670	0.000	Valid	High
Q6	0.621	0.000	Valid	High
Q7	0.657	0.000	Valid	High
Q8	0.719	0.000	Valid	High

Q9	0.626	0.000	Valid	High
Q10	0.347	0.007	Not Valid	Low
Q11	0.480	0.000	Valid	Enough
Q12	0.795	0.000	Valid	High
Q13	0.179	0.172	Not Valid	Very Low
Q14	0.811	0.000	Valid	High

Based on the results of the field test analysis of 14 questions, validity test showed that 12 questions were declared valid with correlation coefficients in the category of sufficient to high (0.480–0.815), while two items, namely Q10 and Q13, were invalid because they had low to very low coefficients.

C. Reliability Cronbach's Alpha

According to Heale and Twycross (2015), an instrument is considered reliable if its Cronbach's alpha coefficient is ≥ 0.70 . Based on the first preliminary field testing, the SPSS analysis yielded an alpha value of 0.509, indicating that the initial items were unreliable. However, after revisions were made and the second preliminary field test was conducted, the Cronbach's alpha rose to 0.845, demonstrating acceptable reliability. This reliability was sustained into the main field testing, which produced a highly consistent alpha value of 0.876.

D. Distinguishing Power

The analysis of the distinguishing power of item question aims to distinguish between students who have high ability and students who have low ability, or to distinguish between the upper and lower groups (Çelen & Aybek, 2022; Rosyad, 2024). The results of the analysis of the distinguishing of test instruments in the first preliminary field testing based on the interpretation of the differential power index where 3 very good items (numbers 3, 7, and 11), 6 items are enough (numbers 4, 6, 8, 12, 13, and 14), 5 bad items (numbers 1, 2, 5, 9, and 10) as show in Table 6.

Table 6. Distinguishing Power for First Preliminary Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
0	-0.2	0.4	0.2	0	0.2	0.4	0.2	-0.2	0	0.8	0.2	0.2	0.2

The results of the analysis of the distinguishing of test instruments in the second preliminary field testing where there are 8 very good questions (numbers 1, 2, 4, 6, 7, 8, 12, and 14), 3 enough questions (numbers 3, 5, and 9), and 3 bad questions (numbers 10, 11, and 13) as show in Table 7.

Table 7. Distinguishing Power for Second Preliminary Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
0.4	0.8	0.2	0.8	0.2	0.8	0.4	0.6	0.2	0	0	0.6	0	0.8

The results for the analysis of the distinguishing test instruments after instrument revision in the main field testing based on the interpretation of the differential power index where there are 11 very good questions (numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 12, and 14), 1 question item was enough (number 10), and 2 items of bad questions (numbers 11 and 13) as show in Table 8.

Table 8. Distinguishing Power for Main Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
0.81	0.75	0.81	0.93	0.75	0.68	0.81	0.87	0.87	0.25	0.18	0.62	0.18	0.62

E. Level of Difficulty

An ideal test item was one that possesses a moderate level of difficulty (Çelen & Aybek, 2022; Rosyad, 2024). In this study, the difficulty level (P) was classified as moderate if it fell within the range of $0.31 < P \leq 0.70$. Based on this interpretation, the first preliminary field testing revealed 4 difficult items (numbers 5, 9, 10, and 13), 2 moderate items (numbers 11 and 12), and 8 easy items (numbers 1, 2, 3, 4, 6, 7, 8, and 14) as show in Table 9.

Table 9. Level of Difficulty for First Preliminary Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
1.000	.900	.800	.700	.000	.900	.800	.900	.100	.000	.600	.500	.100	.900

Following revisions, the second preliminary field testing showed a shift in the difficulty distribution, resulting in 1 difficult item (number 3), 10 moderate items (numbers 2, 4, 5, 6, 7, 8, 9, 10, 12, and 14), and 3 easy items (numbers 1, 11, and 13) as show in Table 10.

Table 10. Level of Difficulty for Second Preliminary Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
.800	.400	.100	.600	.500	.600	.600	.700	.700	.600	.800	.500	.800	.600

Finally, the main field testing produced a distribution of 2 difficult items (numbers 10 and 13), 8 moderate items (numbers 1, 2, 3, 5, 6, 7, 8, and 9), and 4 easy items (numbers 4, 11, 12, and 14) as show in Table 11.

Table 11. Level of Difficulty for Main Field Testing

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
.566	.483	.583	.700	.666	.583	.583	.633	.350	.116	.950	.800	.083	.816

F. Data Analysis After Item Deleted

Based on the results of the field test analysis of 14 questions, the quality of the instrument in general was classified as good because it has met most of the criteria of validity, reliability, differentiation, and difficulty. The results of the validity test showed that 12 questions were declared valid with correlation coefficients in the category of sufficient to high (0.480–0.815), while two items, namely Q10 and Q13, were invalid because they had low to very low coefficients. Cronbach's Alpha reliability value of 0.887 indicates that the instrument has good internal consistency and suitable for use as a measuring tool. In terms of differentiation, most of the items are in the very good category so that they are able to distinguish between students with high and low ability, but Q10 has enough differentiation and Q11 and Q13 have low differentiation. Based on the level of difficulty, the majority of the questions are in the medium category which shows that the composition of the questions is quite proportional, although there are some easy questions (Q4, Q11, Q12, Q14) and difficult questions (Q10 and Q13) that need to be adjusted to make the distribution of difficulty levels more balanced. Overall, the instrument can be declared suitable for use for research data collection, but items Q10 and Q13 can be deleted because they do not meet the valid category so that the quality of the instrument becomes more optimal and are able to measure students' abilities more accurately, as well as low differentiation, this question are also classified as too difficult and almost impossible for students to answer, therefore the following are the results of validity, reliability, distinguishing power, and difficulty level for the fix question after the question item was deleted in table 12.

Table 12. Data Analysis After Item Deleted

No	Level SOLO	Validity	Reliability	Distinguishing Power	Difficulty Level
1	S4	0.547	0.887	0.75	0.5667
2	S5	0.584		0.75	0.4883
3	S4	0.692		0.875	0.5833
4	S3	0.815		0.875	0.7000
5	S4	0.670		0.75	0.6667
6	S2	0.621		0.75	0.5833
7	S5	0.657		0.75	0.5833
8	S3	0.719		0.875	0.6333
9	S5	0.626		0.875	0.3500
11	S5	0.480		0.1875	0.9500
12	S2	0.795		0.6875	0.8000

14	S3	0.811	0.6875	0.8167
----	----	-------	--------	--------

Therefore, it can be concluded that the total number of questions that are suitable for use is 12 items with a distribution of 2 questions at the SOLO 2 (unistructural) level, 3 questions at the SOLO 3 level (multistructural), 3 questions at the SOLO 4 (relational) level, and 4 questions at the SOLO 5 (extended abstract) level.

The results indicate that the SOLO Taxonomy-based assessment instrument developed in this study demonstrates satisfactory psychometric quality and can be used to assess students' conceptual understanding of fluid mechanics. The finalized instrument comprised 12 valid items, with validity coefficients ranging from 0.480 to 0.815 and a reliability coefficient of 0.887. These findings demonstrate that the instrument has adequate consistency in measuring the intended construct and can provide reliable information about students' levels of understanding. Biggs and Collis (1982) explained that assessment using the SOLO Taxonomy is intended to examine not only whether students arrive at correct answers but also the complexity and structure of the thinking underlying those responses. In this regard, the acceptable validity and reliability obtained in the present study indicate that the instrument was able to identify differences in students' cognitive development across the SOLO levels.

Further evidence of the instrument's quality is provided by the item discrimination results. The mean discrimination index of 0.734 demonstrates that the items were generally effective in distinguishing students with higher conceptual understanding from those with lower understanding. This result supports the findings of Mindayani et al. (2019), who stated that SOLO-based assessment should be able to differentiate students' cognitive performance because each SOLO category reflects a distinct level of conceptual development. The relatively high discrimination values obtained for most items indicate that the instrument was sufficiently sensitive to differences in students' mastery of fluid concepts, making it useful for both diagnostic and evaluative assessment.

The SOLO analysis revealed that the multistructural level (S3) was the most dominant, while the percentages at the relational (S4) and extended abstract (S5) levels were comparatively lower. This distribution provides an important description of how students understood fluid concepts. In general, students were able to recognize several relevant concepts, equations, and variables, but they were not always able to establish meaningful connections among them. Biggs and Collis (1982) described multistructural thinking as a stage in which learners can identify or use several relevant elements but have not yet integrated them into a coherent whole. Thus, students may demonstrate knowledge of different aspects of a topic while their overall conceptual structure remains fragmented.

The predominance of the multistructural level also indicates that students had developed considerable factual and procedural knowledge but had not consistently progressed toward relational reasoning. Kaharuddin and Hajeniati (2020) similarly found that students were often able to identify relevant information but experienced difficulties when they had to establish systematic relationships between concepts. Mulbar et al. (2017) likewise reported that learners may remain at intermediate SOLO levels when their learning is dominated by procedural approaches rather than conceptual integration. In the present study, students appeared to understand individual principles involving fluid pressure, buoyancy, and fluid equilibrium, but difficulties became more apparent when several principles had to be combined to explain a more complex physical situation.

The results at the relational and extended abstract levels nevertheless provide a positive indication. The difference between S4 (61%) and S5 (59%) was relatively small, suggesting that students who were able to establish relationships among concepts also showed potential to extend and generalize their understanding to new situations. Fernandez and Guzon (2025) emphasized that relational understanding can function as an important foundation for developing higher levels of abstraction because students who successfully integrate concepts are better positioned to transfer their knowledge across different contexts. Consistent with this perspective, Tian et al. (2024) found that progression from relational to extended abstract thinking is associated with the ability to construct generalized explanations rather than depending primarily on memorized procedures.

Despite the relatively favorable overall results, the decreasing achievement across the higher SOLO levels indicates that students continued to encounter difficulties with increasingly complex forms of thinking. Such a pattern is consistent with the hierarchical nature of the SOLO Taxonomy, in which each successive level requires greater cognitive complexity. Ahmad et al. (2024) explained that

movement between SOLO levels represents a progression from basic recognition and identification toward integrated and abstract reasoning. The findings of this study therefore suggest that students faced greater challenges when they were required to synthesize information, develop explanations, and apply their understanding beyond familiar situations.

These difficulties may also be related to the conceptual characteristics of fluid mechanics. Previous research has identified fluid concepts as a challenging area in physics learning. Hutasoit et al. (2025), for example, identified misconceptions involving hydrostatic pressure, Pascal's principle, buoyancy, and fluid equilibrium. Atmaja and Samsudin (2024) similarly reported that misconceptions may remain even after students have received formal instruction, potentially limiting their ability to establish meaningful relationships among concepts. As a result, students may be able to solve problems involving individual principles while struggling with tasks that require several principles to be considered simultaneously.

Similar conceptual difficulties have been reported in studies of both static and dynamic fluids. Kurniawan (2023) found that students could sometimes use appropriate formulas for static fluid problems despite having only a partial conceptual understanding of the underlying principles. Fatmawati et al. (2025) also reported a discrepancy between students' procedural performance and their deeper conceptual mastery. In a similar manner, Muki et al. (2025) observed that students' understanding of fluid concepts could remain superficial and fragmented, particularly when learning activities placed greater emphasis on mathematical procedures than on conceptual reasoning.

The challenges identified in this study are also consistent with findings from contexts outside Indonesia. Koto and Ilhami (2023) reported that secondary students continued to encounter conceptual difficulties in fluid dynamics following online learning during the pandemic. Putri et al. (2023) likewise identified substantial misconceptions in dynamic fluid concepts through the use of a four-tier diagnostic test. Taken together, these findings indicate that difficulties in understanding fluid concepts are not restricted to the students involved in the present study but represent a broader issue in physics education.

Differences between the upper and lower groups provide further evidence of variation in students' conceptual understanding. Students in the upper group demonstrated almost complete mastery of the concepts assessed, whereas the lower group obtained correct responses on fewer than one-quarter of the items. This considerable performance gap indicates that students did not develop conceptual understanding at the same level. Hasan (2017) stated that students functioning at higher SOLO levels tend to possess more interconnected cognitive structures, allowing them to interpret a problem from a more comprehensive perspective. Conversely, learners at lower SOLO levels tend to concentrate on individual pieces of information and have greater difficulty identifying relationships between concepts. The findings obtained in this study are consistent with this theoretical explanation.

The item discrimination results also contribute to the interpretation of students' conceptual profiles. Items with particularly high discrimination values were predominantly associated with questions that required conceptual reasoning and the integration of several ideas. This observation is consistent with Hartanti et al. (2021), who found that higher-order items based on the SOLO Taxonomy can more effectively differentiate students according to the complexity of their thinking. Claudia et al. (2020) similarly reported that tasks involving relational and extended abstract reasoning are more capable of revealing differences in conceptual understanding than items primarily requiring simple recall.

From an instructional perspective, the dominance of multistructural thinking indicates the need for greater attention to learning approaches that promote deeper conceptual understanding. Sabani et al. (2026) described deep learning as a process in which students actively establish relationships between concepts, critically evaluate information or evidence, and apply their knowledge flexibly in different contexts. The present findings suggest that many students had not yet fully achieved this form of learning because their knowledge remained relatively compartmentalized. Instruction should therefore move beyond the accumulation of factual information and provide greater opportunities for students to connect and integrate related concepts.

Inquiry-based learning is one approach that may support this development. Damopolii et al. (2020) found that inquiry-oriented learning can improve students' SOLO levels by encouraging them to investigate relationships between concepts and develop explanations independently. Such learning experiences can provide opportunities for students to engage in reasoning processes associated with the

relational and extended abstract levels. Accordingly, the incorporation of inquiry activities, problem-based tasks, conceptual discussions, and reflective questioning into physics instruction may help students develop more integrated and sophisticated understanding.

The findings are also relevant to the broader context of international science education. The OECD (2023) PISA report indicated that Indonesian students continue to face challenges in applying scientific knowledge to unfamiliar and complex situations. The dominance of multistructural responses in the present study reflects a related pattern, in which students demonstrate knowledge of individual concepts but experience greater difficulty when required to transfer and apply that knowledge in new contexts. These findings reinforce the importance of strengthening higher-order thinking through instructional practices that emphasize conceptual understanding and the application of knowledge.

Overall, the developed SOLO Taxonomy-based assessment instrument provides detailed information about students' conceptual understanding of fluid mechanics and the structure of their thinking. Rather than merely indicating whether students answered correctly, the instrument was able to reveal differences in the complexity of their understanding across SOLO levels. The predominance of the multistructural level indicates that students possessed a substantial foundation of conceptual knowledge but still required support to connect, integrate, and extend that knowledge. Future instructional efforts should therefore emphasize relational and extended abstract thinking through inquiry-based, conceptually oriented, and reflective learning activities that enable students to construct deeper and more transferable scientific understanding.

4. CONCLUSION

This development research successfully produced a SOLO Taxonomy-based diagnostic assessment framework on fluid mechanics for high school students using the revised Borg and Gall model. The development involved rigorous expert validation and field testing with 80 students. The results indicated that 14 items were evaluated, with 12 items (85.71%) declared valid and 2 items (14.29%) rejected based on validity, reliability, discrimination power, and difficulty level criteria. The final instrument demonstrated high reliability (0.887). Furthermore, most items exhibited 'very good' discrimination power (78.57%), complemented by a balanced difficulty distribution comprising easy (28.57%), moderate (57.14%), and difficult (14.29%) items.

REFERENCES

- Ahmad, F. A. K., Meshkat, M., & Taghipour, M. (2024). Using SOLO model for evaluation of students' physics learning. *Physics Journal Farhangian University*, 1(1), 55–65. <https://doi.org/10.48310/esip.2024.17397.1006>
- Atmaja, D. Y. S., & Samsudin, A. (2024). Are there misconceptions in my class? Misconception analysis with T3-SF (three tier-test on static fluid). *Tarbiyah: Jurnal Ilmiah Kependidikan*, 13(2), 155–170. <https://doi.org/10.18592/tarbiyah.v13i2.13518>
- Biggs, J. B & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of The Observed Learning Outcome)*. Academic Press.
- Biggs, J. (1987). *Student Approaches to Learning and Studying*. Australian Council for Educational Research.
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University (4th ed)*. McGraw-Hill Education.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD.
- Çelen, Ü., & Aybek, E. C. (2022). A novel approach for calculating the item discrimination for Likert type of scales. *International Journal of Assessment Tools in Education*, 9(3), 772–786. <https://doi.org/10.21449/ijate.1173356>
- Claudia, L. F., Kusmayadi, T. A., & Fitriana, L. (2020). The SOLO taxonomy: Classify students' responses in solving linear program problems. *Journal of Physics: Conference Series*, 1538(1), Article 012107. <https://doi.org/10.1088/1742-6596/1538/1/012107>
- Damopolii, I., Nunaki, J. H., Nusantara, E., & Kandowangko, N. Y. (2020). The effectiveness of inquiry-based learning to train students' thinking skill based on SOLO taxonomy. *Journal of Physics: Conference Series*, 1567(4), Article 042025. <https://doi.org/10.1088/1742-6596/1567/4/042025>

- Eryanto, D., & Prasetyono, H. (2025). Efektivitas pembelajaran mendalam dalam implementasi kurikulum merdeka di SD Negeri Cikampek Selatan I. *Jurnal Evaluasi Pendidikan*, 16(2). <https://doi.org/10.21009/jep.v16i2.60722>
- Fatmawati, W., Parno, & Diantoro, M. (2025). Identification of students' mastery level of physics concepts in static fluid material. *Jurnal Pendidikan Fisika dan Teknologi*, 11(1), 26–33. <https://doi.org/10.29303/jpft.v11i1.8360>
- Fernandez, P. J. M., & Guzon, A. F. H. (2025). A SOLO taxonomy-based rubric for assessing conceptual understanding in applied calculus. *Journal on Mathematics Education*, 16(2), 559–580. <https://doi.org/10.22342/jme.v16i2.pp559-580>
- Hartanti, D., Djudin, T., & Mursyid, S. (2021). Analisis tingkat berpikir dalam menyelesaikan soal gerak parabola menggunakan taksonomi Structure of the Observed Learning Outcomes (SOLO) siswa kelas X MIA SMAN 1 Sungai Raya. *Jurnal Inovasi Penelitian dan Pembelajaran Fisika*, 2(1), 1–9. <https://doi.org/10.26418/jippf.v2i1.44591>
- Hasan, B. (2017). Karakteristik respon peserta didik dalam menyelesaikan soal geometri berdasarkan taksonomi SOLO. *JINoP (Jurnal Inovasi Pembelajaran)*, 3(1), 449–458. <https://doi.org/10.22219/jinop.v3i1.4282>
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66–67. <https://doi.org/10.1136/eb-2015-102129>
- Hutasoit, Y., Nahor, D. B., Sianturi, E. S., Sati, L., & Simatupang, N. T. (2025). Studi literatur miskonsepsi siswa pada materi fluida dan tekanan zat cair. *Sindoro: Cendikia Pendidikan*, 18(1). <https://doi.org/10.9644/sindoro.v3i9.252>
- Indrasari, N. (2025). Analysis of senior high school students' science literacy in understanding static fluids. *Journal of Advances in Physics Education*, 1(1), 40–47. <https://doi.org/10.23960/jape.v1i1.376>
- Inggit, S. M., Liliawati, W., & Suryana, I. (2021). Identifikasi miskonsepsi dan penyebabnya menggunakan instrumen Five-Tier Fluid Static Test (5TFST) pada peserta didik kelas XI sekolah menengah atas. *Journal of Teaching and Learning Physics*, 6(1), 49–68. <https://doi.org/10.15575/jotalp.v6i1.11016>
- Kementerian Pendidikan Dasar dan Menengah. (2025). *Naskah akademik pembelajaran mendalam (deep learning) menuju pendidikan bermutu untuk semua*. Kemendikdasmen.
- Kaharuddin, A., & Hajeniati, N. (2020). An identification of students' responses based on SOLO taxonomy in mathematics learning toward learning activities and learning outcomes. *Al-Jabar: Jurnal Pendidikan Matematika*, 11(2), 191–200. <https://doi.org/10.24042/ajpm.v11i2.6270>
- Koto, I., & Ilhami, D. (2023). High-school students' conceptual understanding of fluid dynamics following online learning during the coronavirus pandemics. In *Proceedings Mathematics and Science Education International Seminar* (pp. 117–125). Universitas Negeri Semarang.
- Kurniawan, R. V. (2023). Analisis pemahaman konsep siswa pada topik fluida statis. *TEACHING: Jurnal Inovasi Keguruan dan Ilmu Pendidikan*, 3(1), 67–73. <https://doi.org/10.51878/teaching.v3i1.2165>
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- Mindayani, N., Hadi, W., & Ambarita, B. (2019). The development of SOLO taxonomy based assessment tool on text of observation reports of X grade students of senior high school 1 NA. IX-X. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, 2(4), 371–381. <https://doi.org/10.33258/birle.v2i4.527>
- Muki, B. G., Aviyanti, L., & Nawas, A. (2025). Investigation of students' conceptual understanding of fluids in Kupang City and Sabu Raijua District. *Momentum: Physics Education Journal*, 9(1), 144–162. <https://doi.org/10.21067/mpej.v9i1.11056>
- Mulbar, U., Rahman, A., & Ahmar, A. S. (2017). Analysis of the ability in mathematical problem-solving based on SOLO taxonomy and cognitive style. *World Transactions on Engineering and Technology Education*, 15(1), 68–73. <https://doi.org/10.26858/wtetev15i1y2017p6873>
- OECD. (2023). *PISA 2022 Results Factsheets Indonesia*. OECD Publishing
- Ole, F. C. B., & Gallos, M. R. (2023). Impact of formative assessment based on feedback loop model on high school students' conceptual understanding and engagement with physics. *Journal of Turkish Science Education*, 20(2), 333–355. <https://doi.org/10.36681/tused.2023.019>

- Putri, R. A., Permana, H., & Nasbey, H. (2023). Identifikasi miskonsepsi materi fluida dinamis dengan menggunakan tes diagnostik four-tier untuk siswa SMA kelas XI. *Joint Prosiding IPS dan Seminar Nasional Fisika*, 11(1), PF181–PF186. <https://doi.org/10.21009/03.1102.pf25>
- Sabani., Panggabean, D. D., & Ramadhani, I. (2026). *Asesmen Pembelajaran Mendalam pada Bidang Studi Fisika*. PT MAFY Media Literasi Indonesia.
- Sedhandoyo, C. (2024). *Pembelajaran Poeaw Berbantuan Simulasi Phet untuk Meningkatkan Pemahaman Konsep Siswa SMA pada Materi Fluida Dinamis*. Universitas Pendidikan Indonesia.
- Sugiyono. (2023). *Metode Penelitian Kuantitatif Kualitatif*. Alfabeta.
- Tian, P., Fan, Y., Sun, D., & Li, Y. (2024). Evaluating students' computation skills in learning amount of substance based on SOLO taxonomy in secondary schools. *International Journal of Science Education*, 46(15), 1578–1600. <https://doi.org/10.1080/09500693.2023.2291691>
- UNESCO. (2021). *Reimagining Our Futures Together: A New Social Contract for Education*. UNESCO.
- Yuberti, Y., Suryani, Y., & Kurniawati, I. (2020). Four-tier diagnostic test with certainty of response index to identify misconception in physics. *Indonesian Journal of Science and Mathematics Education*, 3(2), 245–253. <https://doi.org/10.24042/ijsme.v3i2.6061>

Vocational and Mathematical Knowledge: Designing Hybrid Literacy Tasks in the Context of Food Security

Ai Tusi Fatimah¹, Agus Yuniawan Isyanto², Euis Erlin³

¹Department of Mathematics Education, Universitas Galuh, Ciamis, Indonesia

²Department of Agribusiness, Universitas Galuh, Ciamis, Indonesia

³Department of Biology Education, Universitas Galuh, Ciamis, Indonesia

¹aitusifatimah@unigal.ac.id

ABSTRACT

Mathematical literacy in vocational contexts requires instructional tasks that effectively bridge academic content with job-specific knowledge. This case study examines the design and impact of hybrid literacy tasks—integrating national food policy narratives with exponential number problems—within a food security context relevant to agribusiness education. The study adopted a case study approach and collected data from teachers and students at an Indonesian agribusiness vocational school through Focus Group Discussions (FGDs), student worksheets, and pre-test and post-tests. FGDs established a functional framework connecting food security elements to mathematical content and essential work skills. The study found that the hybrid task design successfully facilitated students' reasoning by establishing the context as a source of analogy, which encouraged imitative reasoning (identifying similarities) to solve subsequent, more complex problems. This experience significantly improved students' computational fluency in representing large numbers using various power forms, a key vocational skill for food security analysis. The intervention resulted in a high average n-gain score in mathematics learning. The research demonstrates a practical model for bridging vocational and mathematical knowledge through transdisciplinary hybrid task design. The findings suggest that educational practices in vocational settings should leverage vocational contexts as functional mechanisms—not mere settings—to cultivate mathematical literacy and computational work skills.

Keyword: Exponential Numbers, Food Security Context, Hybrid Tasks, Imitative Reasoning, Vocational Mathematics



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Ai Tusi Fatimah,
Department of Matjematics Education,
Universitas Galuh,
Jalan R.E. Martadinata No 150 Ciamis 46274, Indonesia.
atusifatimah@unigal.ac.id

1. INTRODUCTION

Mathematical literacy is commonly understood as comprising contextual application, mathematical content knowledge, reasoning and problem-solving processes, and broader competencies required in contemporary life (OECD, 2019). As studies on mathematical literacy continue to develop, identifying relevant and meaningful contexts for learning has become increasingly important. Agriculture represents one such context, particularly within vocational education settings (Vallera & Bodzin, 2020). Evidence from FAO-based frameworks suggests that food insecurity is linked to poorer health outcomes and has become a significant global sustainability challenge (Beyene, 2023; Pérez-Escamilla, 2024). In Indonesia, agribusiness is a major vocational specialization designed to equip students with the competencies needed in the agricultural sector and the broader labor market (Menteri Pendidikan Kebudayaan Riset dan Teknologi, 2022). Consequently, mathematics learning in vocational high schools should support the development of vocational competencies, including students' dispositions, knowledge, and practical skills related to food security (Kemendikbudristek BSKAP, 2022). Therefore, employing food security as a learning context offers a meaningful way to strengthen mathematical literacy among agribusiness students (Fatimah et al., 2023, 2026).

Effective mathematics learning in vocational settings needs to reflect the mathematical demands embedded in occupational practice. This perspective is captured in the concept of Vocational Mathematical Knowledge, which emphasizes the relationship between mathematical understanding and workplace-specific skills and procedures (Bakker, 2014; Heusdens et al., 2016). Integrating vocational practice into mathematics teaching presents difficulties. A notable disparity exists between the theory of mathematical literacy and its instructional implementation (Yang et al., 2020), frequently leading to tasks that do not correspond with the principles of accessibility, challenge, and utility needed for vocational learners (Sawatzki, 2017). Additionally, educators require a blend of mathematical and

vocational knowledge to successfully create tasks that foster 21st-century skills such as transdisciplinarity and computational thinking (Hawkins & Neubauer, 2015). To address this challenge, this study adopts the hybrid task framework (Kuntze et al., 2017), in which a single task integrates two complementary learning goals. Previous studies suggest that hybrid tasks enhance conceptual understanding, critical thinking, and mathematical problem-solving through authentic learning experiences (Rahmawati et al., 2022; Smirnov et al., 2023). Accordingly, this study integrates vocational knowledge on food security with exponential functions, making the vocational context an essential part of students' mathematical reasoning rather than merely a contextual backdrop.

Earlier studies have examined the mathematical requirements of vocational high schools in agribusiness, recognizing arithmetic as a fundamental task in multiple industries (Fatimah & Solihah, 2020). Although current studies validate the necessity for contextual integration, they have yet to offer a solid and validated framework for creating hybrid tasks that systematically link national food policy narratives to specific mathematical concepts (e.g., exponential numbers) and examining the resulting impacts on students' reasoning.

Consequently, this study offers an in-depth examination of mathematics learning by specifically emphasizing the development and execution of hybrid literacy tasks related to food security. Its purpose is to emphasize how the elements of mathematical literacy (context, content, processes, and skills) are purposefully integrated in these designs. This study specifically contributes by: (1) creating a functional framework for designing hybrid tasks that merge policy context and exponential content, (2) examining how context aids students' imitative and analogical reasoning, and (3) assessing the influence of task-solving experiences on computational fluency and educational outcomes. In light of these goals, the subsequent research questions were addressed: (RQ1) What structure is essential for developing impactful hybrid mathematical literacy tasks that merge food security themes, exponential concepts, and vocational skill sets? (RQ2) In what way does the contextual framework in hybrid tasks enhance students' mathematical reasoning (e.g., analogical or imitative reasoning)? (RQ3) What effects do these hybrid tasks have on students' mathematical learning outcomes (n-gain) and their growth in computational fluency?

2. RESEARCH METHOD

This research employed a descriptive case study approach (Hancock & Algozzine, 2006) to thoroughly examine the planning, execution, and effects of a hybrid mathematical literacy task in a vocational environment. The case location was a public vocational high school focused on agribusiness and agrotechnology in Ciamis Regency, Indonesia. The study was conducted at a public vocational school offering a food crops and horticulture program. The school was selected based on accessibility considerations and its suitability to represent the agribusiness vocational context under investigation. Four teachers participated in the study, including two mathematics teachers and two agribusiness vocational teachers. Their involvement in a series of Focus Group Discussions (FGDs) was essential for refining and validating the proposed hybrid task framework presented in Table 1. The main participants consisted of 24 tenth-grade students (6 females, 18 males) focused on crop agribusiness. Students were split into seven cooperative teams (G1 to G7) to finish the assignment. This cooperative framework was essential for witnessing the fluid group reasoning process.

Table 1. Procedures for Designing and Investigating Hybrid Mathematical Literacy Tasks

Research Phase	Activities
Initial Exploration	Focus group discussions involving mathematics teachers and agribusiness teachers were conducted to: (a) examine vocational curriculum standards and learning processes in Indonesia; (b) identify the role of mathematics in agribusiness vocational education; and (c) explore mathematical literacy components related to food security as a basis for task development.
Task Development	Discussions with four mathematics teachers were carried out to: (a) formulate learning objectives; (b) develop mathematical literacy tasks situated in food security contexts; and (c) identify the student participants from the agribusiness food crops and horticulture program.

Research Design Selection	A descriptive case study approach was adopted to provide an in-depth account of the design process and classroom implementation of hybrid mathematical literacy tasks.
Classroom Observation	Students' activities, interactions, and strategies were observed while they engaged in the mathematical tasks.
Student Interviews	Semi-structured interviews were conducted with selected students to gain insight into their approaches to solving representative tasks.
Documentation Collection	Data sources included completed student worksheets as well as video and audio recordings collected during learning activities.
Data Analysis and Interpretation	Observation records, interview transcripts, and student work were analyzed to identify patterns related to mathematical understanding, reasoning, representation, and vocational work-skill knowledge.
Reporting of Results	Findings were synthesized and organized according to the research objectives and research questions.
Validation of Findings	The findings were compared with previous studies and discussed with mathematics and agribusiness teachers to strengthen the credibility of the interpretations.

The data gathering employed a multimodal strategy to guarantee triangulation. Initially, Focus Group Discussions (FGDs): Held with educators to outline the connection between the food security curriculum and mathematics material, which is essential for creating hybrid assignments (RQ1) (as illustrated in Table 2 and Figure 1). Second, Pre-tests and Post-tests: Administered to individual students to assess advancements in their understanding of exponential numbers. Learning outcomes were assessed using the normalized gain (n-gain) score method to evaluate the effectiveness of the intervention (RQ3). Third, Student Worksheets and Observation Notes were utilized to monitor the procedural steps followed by the groups, examine computational mistakes, and recognize the influence of context in problem-solving. Fourth, students were interviewed during standard problem-solving instances (specifically G1, G4, G6, G2, G7) to understand their thought processes and confirm imitative and analogical reasoning strategies (RQ2).

The analysis was conducted in two stages. First, learning improvement was assessed using normalized gain (n-gain) scores derived from students' pretest and posttest results. Secondly, qualitative data (worksheets, observation transcripts, and interviews) underwent thematic analysis. This analysis focused on students' cognitive processes to identify reasoning patterns, which were subsequently categorized according to Lithner's framework (Lithner, 2008, 2017), particularly instances of imitative reasoning and analogy-based problem-solving strategies influenced by the context of food security. This combined analysis provided a comprehensive understanding of how the hybrid task design supported mathematical processes and vocational computational skills.

3. RESULTS AND DISCUSSION

A. Results

The empirical findings are organized into three cohesive sections, aligned with the research inquiries: the development of the Hybrid Task Design Framework (RQ1), the identified process of student reasoning supported by the context (RQ2), and the measurable effect on computational fluency and educational results (RQ3). The central discovery in these three domains is that intentionally crafting tasks to fulfill a dual, hybrid role—combining vocational policy with mathematical methods—effectively converted the food security context into a practical cognitive tool, rather than just a setting. This practical context mainly grounded students' problem-solving in analogy, swiftly evolving into effective Imitative Reasoning techniques crucial for professional tasks.

1) Framework for Hybrid Task Design

The Focus Group Discussions (FGDs) involving mathematics and agribusiness educators developed the Hybrid Task Design Framework (Figure 1), validating the intentional connection between vocational context and mathematical content. The study's clear learning aim was stated as: "Students can represent numbers using powers," matching the task design with the "Number" component of the curriculum

(Kemendikbudristek BSKAP, 2022). The particular context, Food Availability (based on Law No. 18 of 2012) (Presiden Republik Indonesia, 2012), was chosen because assessments of grain production and stock fundamentally include substantial amounts that necessitate exponential notation for brief numerical expression.

Table 2. Mapping Food Security Dimensions to Learning Outcomes in the Basic Crop Cultivation Subject

Food Security Dimension	Description of the Dimension	Related Learning Outcomes
Food Availability	Food production, food reserves, diversification of food stocks, and supporting agricultural facilities and infrastructure.	Understanding fundamental technical processes in crop, fishery, livestock, and agricultural product processing activities; recognizing technological developments and contemporary issues in agribusiness industries.
Food Consumption	Food diversity and nutritional improvement.	Examining technological advancements and emerging global issues associated with agribusiness and agricultural product processing.
Food Accessibility	Distribution systems, marketing channels, trade mechanisms, food stability, and food assistance programs.	Understanding integrated business processes within the agribusiness sector.
Food Safety	Sanitation practices, regulation of food additives, genetically modified food products, irradiation processes, and packaging systems.	Understanding basic technical procedures related to agricultural production and agricultural product processing.

Table 2 above outlines the relationship between components of food security and aspects of vocational learning outcomes. The FGD analysis shows that Food Availability is the most substantial area for developing raw data computation skills and efficiency calculations, which significantly depend on grasping large quantities.

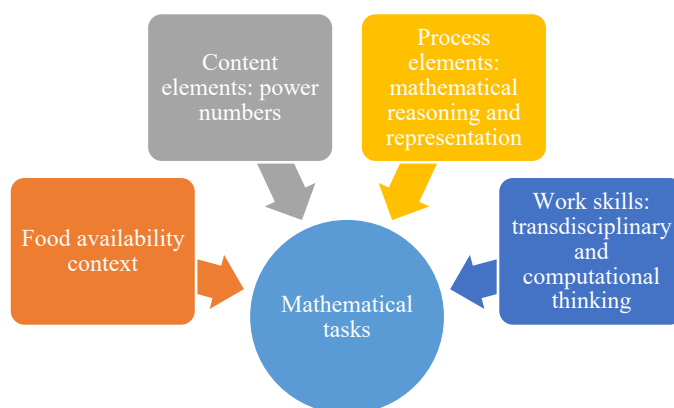


Fig 1. Conceptual Structure of Mathematical Literacy in Food Security-Based Learning

Figure 1 presents the conceptual structure of the proposed hybrid literacy task design. In this framework, quantitative reasoning serves as a transdisciplinary bridge connecting mathematical content with vocational knowledge. The mathematical component focuses on exponential-number concepts, while the vocational component emphasizes workplace-related calculations commonly encountered in

agribusiness activities. This integration informed the development of mathematical literacy tasks that linked food security issues, vocational practices, and mathematical reasoning, as summarized in Table 3.

Table 3. Design Characteristics of Food Security-Based Mathematical Literacy Tasks

Task	Food Security Dimension	Vocational Context	Mathematical Content	Literacy Process	Expected Work-Related Skill
1	Food Availability	Grain drying and moisture content	Exponential numbers	Interpreting quantitative information and determining water content	Estimating agricultural product quality
2	Food Production	Grain shrinkage during drying	Exponential numbers and percentages	Applying mathematical procedures to production data	Monitoring production efficiency
3	Food Reserves	Conversion from harvested grain to milled grain	Exponential numbers	Mathematical modelling and representation	Managing food reserve calculations
4	Food Consumption	Estimation of regional rice demand	Exponential numbers and multiplication	Quantitative reasoning and interpretation	Forecasting food demand

Table 3 presents the key characteristics of the food security-based mathematical literacy tasks. The tasks were organized to reflect increasing levels of contextual and mathematical complexity. Students were initially required to interpret quantitative information embedded in vocational situations and subsequently engage in more sophisticated reasoning, mathematical modelling, and calculations involving exponential-number concepts within agribusiness contexts.

2) Contextual Analogy and Imitative Reasoning

Examination of student performance and interview transcripts indicated that the context was essential in promoting problem-solving, quickly creating a system of Imitative Reasoning (Jonsson et al., 2014; Lithner, 2008, 2017) influenced by analogy.

Early observations validated the uniqueness of the tasks, although students acknowledged their job-related significance. A student from Group 5 (who initially felt uneasy with the new format) remarked: "This is the first occasion I've encountered a math problem that incorporates a law." "However, I now realize that the rice we are assessing is vital for the country's food reserves." This indicates that the policy context effectively gave functional importance to the mathematical issue.

The strongest proof for Imitative Reasoning was the procedural transfer noted between Task 1 and Task 2. A member of Group 2 (who transitioned successfully) described his thinking process: "We revisited the water percentage formula in the initial problem. Even though the equation varies, the reasoning remains identical: determining the disparity between the starting and ending weights. We applied a comparable method in the second issue. The rice scenario remains unchanged, so we are confident that the calculation method isn't very dissimilar. This verifies that the food security context acted as a cognitive anchor, allowing students to effectively apply procedural templates from one problem variation to another, supporting the context's function as a source of procedural analogy.

$$\frac{10^6 \times 13}{100} = 1.000.000 \times 13 = \frac{13.000.000}{100} = 130.000$$

Jadi, $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6$ dan dikalikan 13%, $13\% = \frac{13}{100}$

$$\begin{aligned} \text{Kenaikan} &= \frac{\text{berat GKP} - \text{berat GKG}}{\text{berat GKP}} \times 100\% \\ &= \frac{100.000 \text{ kg} - 88.000}{100.000} \times 100\% \\ &= \frac{12.000}{100.000} \times 100\% = \frac{1.200.000}{100.000} = 12\% \end{aligned}$$

$10^5 = 100.000 \text{ kg}$
 $1 \text{ ton} = 1000 \text{ kg}$
 $= 88.000$

(a)

$$\text{Beras Gkg} = \left[\frac{100\% - 25\% \text{ La Gkg}}{100\% - 14\% \text{ La Gkg}} \right] \times \text{berat Gkg}$$

$$= \frac{75\%}{86\%} \times 8000 = 6.976.7441$$

$$= 7 \times 10^3 \text{ kg}$$

(b)

$$\text{kebutuhan beras...} = 3 \times 10^2 \text{ kg} \times 1,5 \times 10^8$$

$$= 300 \times 150.000.000$$

$$= 450.000.000.000 \text{ kg} = 45 \text{ juta ton}$$

(c)

(d)

Fig 2. Examples of student responses to (a) Task 1, (b) Task 2, (c) Task 3, (d) Task 4

Figure 2 illustrates representative student responses to the four hybrid mathematical tasks. In Figure 2(b), the student correctly converted 88 tons into 88,000 kg, substituted the values into the shrinkage formula, and accurately calculated a 12% reduction from harvested dry grain to milled dry grain. Likewise, in Figure 2(d), the student correctly substituted the rice consumption and population values, applied the law of exponents, and obtained the annual rice demand of 4.5×10^{10} kg. In contrast, the responses in Figures 2(a) and 2(c) relied primarily on procedural calculations with limited contextual interpretation and less precise mathematical representation. These findings suggest that students were able to utilize appropriate solution strategies after being prompted by analogy triggers from earlier tasks, enabling them to solve increasingly complex mathematical expressions while maintaining connections to the vocational context.

3) Computational Fluency and Learning Outcomes

The intervention led to a notable improvement in personal mathematical performance, demonstrated by an average n-gain score categorized as high (Figure 3).

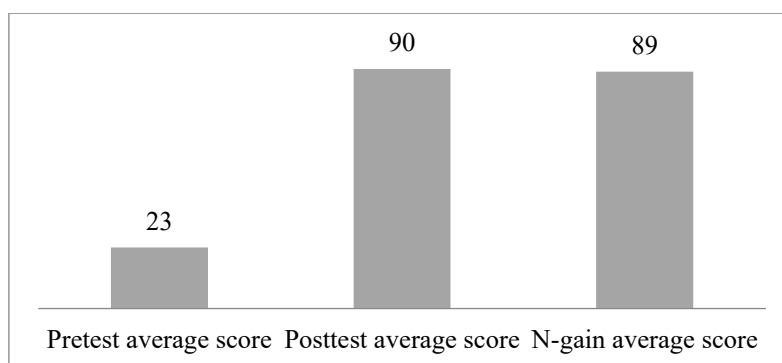


Fig 3. Average Pretest, Posttest, and N-Gain Scores

The elevated mean n-gain score (Figure 3) demonstrates that the hybrid task intervention effectively supported individual proficiency in exponential material. This enhancement was greatly aided by the group's experience in tackling complicated issues. Figure 4 illustrates the average n-gain scores for each group, highlighting the relationship between regular task completion and enhanced learning results.

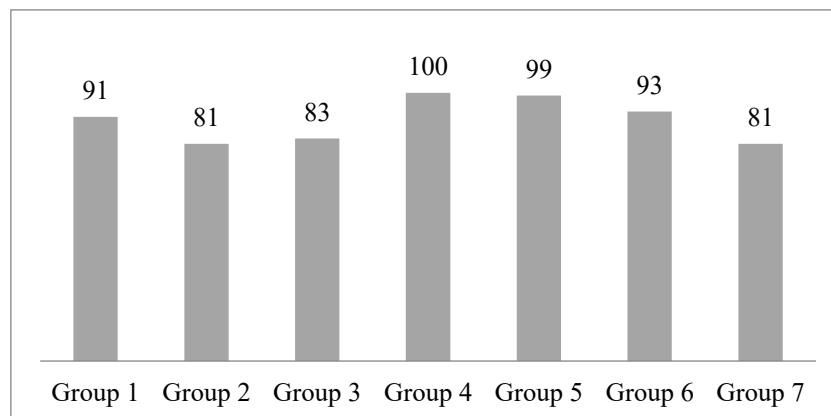


Fig 4. Average N-gain Score for Each Student Group

Groups that effectively completed all tasks (G1, G4, and G6) attained a notably higher average n-gain score (Figure 4), emphasizing the direct connection between continuous procedural success in hybrid tasks and learning success. Moreover, the concluding tasks distinctly showcased the growth of Computational Fluency as a fundamental work skill. Observations indicated that Group 4 students utilized unit conversion through practical vocational reasoning: "Since the per capita consumption figures in other questions are in kg. In agricultural business, it is necessary to change units to a uniform measurement prior to computing the overall inventory."

B. Discussion

This part analyzes the results by exploring the four interrelated elements of mathematical literacy—context, content, reasoning, and work skills—to illustrate the efficacy of the hybrid task design. The research determined that the intentional integration of national food security policy and the essence of exponential numbers constitutes the fundamental aspect of the hybrid tasks (Kuntze et al., 2017). This combined aim—enhancing food security awareness while developing mathematical comprehension—is essential, as activities must encourage coherent and interconnected goal areas (Thanheiser, 2017). The importance of this dual-domain strategy is being increasingly acknowledged in numerous complex and dynamic areas. For example, hybrid optimization methods are needed to manage intricate, multi-objective challenges in cloud task allocation (Sharma et al., 2025), while hybrid task scheduling is essential for reconciling human judgment with machine technology in Industry 5.0 (He et al., 2025). Similarly, recent studies in mathematics education have shown that hybrid instructional designs promote authentic learning experiences, student engagement, and mathematical understanding (Cosentino et al., 2025; Fang, 2026). Together, these findings reinforce the potential of hybrid task design as an effective pedagogical approach for preparing vocational students to apply mathematical knowledge in interdisciplinary workplace contexts (Triantafyllidis et al., 2023).

The concept of food security was not just a story; it acted as a mental aid to enhance students' problem-solving abilities. Results indicated that students, who were initially unaware of the policy narrative, swiftly employed the context as a comparison to identify solution approaches. This situation led to the reliable application of Imitative Reasoning (Lithner, 2003, 2008, 2017), particularly defined as "reasoning founded on recognizing similarities." This transfer process, in which strategy selection relies on recognizing surface features akin to those seen in past examples, was frequently noted as students replicated the effective procedural approach of one task to tackle the following one. This discovery holds great importance for Vocational Mathematics (Bakker, 2014). In a professional setting, efficiency is crucial, and effective practice frequently requires the precise application of recognized, standardized procedures (Yeo, 2017). The achievement of students in transferring procedures through Imitative Reasoning aligns seamlessly with the objective of vocational training: acquiring the practical skills required in the workplace (Fatimah et al., 2024). Earlier research additionally indicates that vocational settings greatly enhance student reasoning and problem-solving abilities (Fatimah et al., 2019; Fatimah & Prabawanto, 2020).

The effective transfer of procedural knowledge resulted in an elevated n-gain score and a distinct display of Computational Fluency. The design employed exponential figures to manage the vast scale of data involved in food security analysis, directly promoting Computational Thinking—the capacity

to convert extensive data into abstract ideas and grasp data-driven reasoning (Hawkins & Neubauer, 2015). This finding is consistent with recent studies showing that hybrid mathematics learning can foster computational thinking through technology-supported and integrated instructional designs (Helsa et al., 2023). The capability of students to execute precise unit conversions (e.g., tons to kilograms) and accurately replace variables in formulas signifies a proficiency in Transdisciplinarity—literacy and comprehension of cross-disciplinary ideas (Hawkins & Neubauer, 2015). The power-form numbers, when applied within a certain unit, shifted from theoretical mathematical understanding to practical knowledge vital for the agricultural industry. The procedural execution observed, while occasionally relying on experiential reasoning from prior education (Lithner, 2000), was effectively directed by the vocational context of the hybrid task to yield functionally accurate solutions, validating that the mathematics gained practical usefulness for the students. This practical orientation is also reflected in recent studies reporting that hybrid learning supported by problem-based scaffolding strengthens students' mathematical problem-solving in authentic contexts (Setiani et al., 2025).

However, the challenges noted with numerical estimation and the ultimate exponential format (Task 3) underscore a significant educational hurdle. Although the vocational setting effectively enhances procedural fluency, the connection between computational results and suitable mathematical representation (such as standard form or scientific notation) necessitates targeted instructional emphasis. The difficulties encountered by students in rounding and displaying the final outcome indicate that educators should offer targeted reflective prompts centered on their representational decisions to enhance the interaction between computational thinking and formal mathematical representation, as indicated by Kuntze et al. (2017). Moreover, the narrow focus of the context on national food security policies (Table 3) suggests that future studies should extend to global food challenges, where diverse resilience strategies (Yu et al., 2015) will require even more intricate mathematical modeling and interdisciplinary approaches.

4. CONCLUSION

This research effectively examined and validated the issues associated with incorporating mathematical literacy into vocational high school teaching, particularly via the creation and execution of hybrid tasks. The identified elements of mathematical literacy—integrating national food policy narratives (context) with exponential figures (content)—were validated as extremely pertinent. The study demonstrates that this particular design meets the requirement for vocational mathematics to have practical significance, as utilizing exponential forms directly aligns with the succinct depiction of substantial quantities found in food security assessment.

The examination of student feedback validated that the hybrid task structure effectively addressed the issue of connecting theoretical concepts with practical application. The context acted as a useful cognitive resource, assisting students in drawing analogies that supported procedural transfer. This mechanism has been shown to enhance Imitative Reasoning, proving that the design successfully generates the essential procedural templates for students to efficiently tackle intricate, later tasks.

Ultimately, the research verified that this intervention had a positive effect on learning results and job skills. The elevated n-gain scores validated personal learning accomplishments, while the students' steady fluency in unit conversion and formula substitution demonstrated the growth of vital transdisciplinary and computational skills. In conclusion, the results affirm that the strategic formulation of hybrid tasks is a proven and efficient model for converting mathematical knowledge into applicable working knowledge, effectively addressing the competency requirements of the agribusiness industry.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support provided by the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Education, Culture, Research, and Technology, through the Second-Year Higher Education Excellence Basic Research Grant 2023 under Contract No. 074/ES/PG.02.00.PL/2023.

REFERENCES

Bakker, A. (2014). Characterising and developing vocational mathematical knowledge. *Educational Studies in Mathematics*, 86(2), 151–156. <https://doi.org/10.1007/s10649-014-9560-4>

- Beyene, S. D. (2023). The impact of food insecurity on health outcomes: empirical evidence from sub-Saharan African countries. *BMC Public Health*, 23(1), 338. <https://doi.org/10.1186/s12889-023-15244-3>
- Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2025). Hybrid teaching intelligence: Lessons learned from an embodied mathematics learning experience. *British Journal of Educational Technology*, 56(2), 621–649. <https://doi.org/10.1111/bjet.13525>
- Fang, C. (2026). The Construction and Practical Path of the Online-Offline Integrated Hybrid Teaching System for College Mathematics. *International Journal of New Developments in Education*, 8(3), 26–32. <https://doi.org/10.25236/IJNDE.2026.080304>
- Fatimah, A. T., Isyanto, A. Y., Erlin, Euis, & Andriana, B. B. (2026). Critical Orientation in Bridging Mathematical and Food Security Literacy through Problem-Based STEAM-H. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, 10(3), 1428–1440. <https://doi.org/10.22437/jiituj.v10i3.53812>
- Fatimah, A. T., Isyanto, A. Y., & Toto. (2023). Science, technology, engineering, agriculture, mathematics, and health in agribusiness curriculum. *International Journal of Evaluation and Research in Education (IJERE)*, 12(4), 2316–2326. <https://doi.org/10.11591/ijere.v12i4.25665>
- Fatimah, A. T., & Prabawanto, S. (2020). Mathematical understanding and reasoning of vocational school students in agriculture-based mathematical tasks. *Journal for the Education of Gifted Young Scientists*, 8(2), 701–712. <https://doi.org/10.17478/JEGYS.702884>
- Fatimah, A. T., Pramuditya, S. A., & Wahyudin, W. (2019). Imitative and creative reasoning for mathematical problem solving (in context horticultural agribusiness). *Journal of Physics: Conference Series*, 1157(4). <https://doi.org/10.1088/1742-6596/1157/4/042092>
- Fatimah, A. T., & Solihah, S. (2020). Mathematics in the Subject of Vegetable Processing Production. *JUMLAHKU: Jurnal Matematika Ilmiah STKIP Muhammadiyah Kuningan*, 6(2), 176–187. <https://doi.org/10.33222/jumlahku.v6i2.1118>
- Fatimah, A. T., Yuniawan Isyanto, A., & Erlin, E. (2024). STEAM-H to Support Agribusiness Students' Work Skill. *Proceeding International Conference on Mathematics and Learning Research*.
- Hancock, D. R., & Algozzine, B. (2006). Doing Case Study Research: A Practical Guide for Beginning Researchers. In *Teachers College Press*. <https://doi.org/10.1353/csd.2007.0003>
- Hawkins, J. N., & Neubauer, D. E. (2015). Twenty-First Century Work Skills and Competencies. In *International and Development Education* (pp. 9–23). https://doi.org/10.1057/9781137491923_2
- He, Y., Dai, Z., Zhang, M., & Jian, C. (2025). A real-time hybrid task resilient collaborative scheduling strategy in the industry 5.0. *Cluster Computing*, 28(11), 702. <https://doi.org/10.1007/s10586-025-05456-0>
- Helsa, Y., Turmudi, & Juandi, D. (2023). Hybrid Learning for Conceptual Understanding Skills in Mathematics: A Meta-Analysis. *International Journal of Information and Education Technology*, 13(2), 355–363. <https://doi.org/10.18178/ijiet.2023.13.2.1814>
- Heusdens, W. T., Bakker, A., Baartman, L. K. J., & De Bruijn, E. (2016). Contextualising Vocational Knowledge: A Theoretical Framework and Illustrations From Culinary Education. *Vocations and Learning*, 9(2), 151–165. <https://doi.org/10.1007/s12186-015-9145-0>
- Jonsson, B., Norqvist, M., Liljekvist, Y., & Lithner, J. (2014). Learning mathematics through algorithmic and creative reasoning. *Journal of Mathematical Behavior*, 36, 20–32. <https://doi.org/10.1016/j.jmathb.2014.08.003>
- Kemendikbudristek BSKAP. (2022). Salinan Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 008/H/KR/2022 Tentang Capaian Pembelajaran Pada Pendidikan Anak Usia Dini Jenjang Pendidikan Dasar dan Jenjang Pendidid. In *Kemendikbudristek BSKAP RI* (Number 021).
- Kuntze, S., Aizikovitsh-Udi, E., & Clarke, D. (2017). Hybrid task design: connecting learning opportunities related to critical thinking and statistical thinking. *ZDM - Mathematics Education*, 49(6), 923–935. <https://doi.org/10.1007/s11858-017-0874-4>
- Lithner, J. (2000). Mathematical reasoning and familiar procedur. *International Journal of Mathematical Education in Science and Technology*, 31(1), 83–95. <https://doi.org/10.1080/0020739950260101>
- Lithner, J. (2003). Students' mathematical reasoning in university textbook exercises. *Educational Studies in Mathematics*, 52(1), 29–55. <https://doi.org/10.1023/A:1023683716659>
- Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3), 255–276. <https://doi.org/10.1007/s10649-007-9104-2>
- Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM - Mathematics Education*, 49(6), 937–949. <https://doi.org/10.1007/s11858-017-0867-3>

- Menteri Pendidikan Kebudayaan Riset dan Teknologi. (2022). Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 262 Tahun 2022 tentang Perubahan Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 56 tentang Pedoman Penerapan Kurikulum Pemulihan Pembelajaran. *Kementerian Pendidikan Kebudayaan Riset Dan Teknologi Republik Indonesia*. [https://ditpsd.kemdikbud.go.id/upload/filemanager/download/kurikulum-merdeka/Kepmen No 262 Perubahan 56 Pedoman Pelaksanaan Kurikulum Pemulihan Pembelajaran.pdf](https://ditpsd.kemdikbud.go.id/upload/filemanager/download/kurikulum-merdeka/Kepmen_No_262_Perubahan_56_Pedoman_Pelaksanaan_Kurikulum_Pemulihan_Pembelajaran.pdf)
- OECD. (2019). PISA 2018 Assessment and Analytical Framework. In *OECD Publishing*. <https://doi.org/10.1787/b25efab8-en>
- Pérez-Escamilla, R. (2024). Food and nutrition security definitions, constructs, frameworks, measurements, and applications: global lessons. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1340149>
- Presiden Republik Indonesia. (2012). Undang-undang Republik Indonesia Nomor 18 Tahun 2012 Tentang Pangan. In *Sekretariat Negara*.
- Rahmawati, P., Parta, I. N., & Permadi, H. (2022). Learning of Hybrid Methods to Improve Mathematics Critical Thinking Ability of Students Class XI MIPA. *Jurnal Inspirasi Pendidikan*, 12(1), 48–59. <https://doi.org/10.21067/jip.v12i1.6360>
- Sawatzki, C. (2017). Lessons in financial literacy task design: authentic, imaginable, useful. *Mathematics Education Research Journal*, 29(1), 25–43. <https://doi.org/10.1007/s13394-016-0184-0>
- Setiani, Y., Ruhiat, Y., & Hendrayana, A. (2025). Needs Analysis for Designing Hybrid Learning with Problem-Based Scaffolding to Enhance Mathematical Problem-Solving Skills. *Journal of Mathematics Instruction, Social Research and Opinion*, 4(2), 443–454. <https://doi.org/10.58421/misro.v4i2.459>
- Sharma, I., Gupta, R., & Singh, P. (2025). Task scheduling in cloud using multi-objective hybrid approach. *Cluster Computing*, 28(11), 692. <https://doi.org/10.1007/s10586-025-05363-4>
- Smirnov, E. I., Tikhomirov, S. A., & Zubova, E. A. (2023). Founding complexes of multi-stage mathematical and informational tasks in the hybrid intellectual environment of school mathematics. *Perspectives of Science and Education*, 64(4), 603–620. <https://doi.org/10.32744/pse.2023.4.37>
- Thanheiser, E. (2017). Commentary on mathematical tasks and the student: coherence and connectedness of mathematics, cycles of task design, and context of implementation. *ZDM - Mathematics Education*, 49(6), 965–969. <https://doi.org/10.1007/s11858-017-0895-z>
- Triantafyllidis, E., Acero, F., Liu, Z., & Li, Z. (2023). Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network ROMAN. *Nature Machine Intelligence*, 5(9), 991–1005. <https://doi.org/10.1038/s42256-023-00709-2>
- Vallera, F. L., & Bodzin, A. M. (2020). Integrating STEM with AgLIT (Agricultural Literacy Through Innovative Technology): The Efficacy of a Project-Based Curriculum for Upper-Primary Students. *International Journal of Science and Mathematics Education*, 18(3), 419–439. <https://doi.org/10.1007/s10763-019-09979-y>
- Yang, X., Kuo, L. J., & Jiang, L. (2020). Connecting Theory and Practice: a Systematic Review of K-5 Science and Math Literacy Instruction. *International Journal of Science and Mathematics Education*, 18(2), 203–219. <https://doi.org/10.1007/s10763-019-09957-4>
- Yeo, J. B. W. (2017). Development of a Framework to Characterise the Openness of Mathematical Tasks. *International Journal of Science and Mathematics Education*, 15(1), 175–191. <https://doi.org/10.1007/s10763-015-9675-9>
- Yu, W., Elleby, C., & Zobbe, H. (2015). Food security policies in India and China: implications for national and global food security. *Food Security*, 7(2), 405–414. <https://doi.org/10.1007/s12571-015-0432-2>

AI-Assisted Ethnomathematics Gamification through Traditional Stilt Games for Geometry Learning in Elementary Education

Tasya Fortuna Aurora¹, Beti Istanti Suwandayani²

^{1,2} Faculty of Teacher Training and Education, Universitas Muhammadiyah Malang, Malang, Indonesia
beti@umm.ac.id

ABSTRACT

This study developed and evaluated an artificial intelligence-assisted ethnomathematics learning model that integrates the traditional stilt game into elementary geometry instruction. Using a Research and Development (R&D) approach with the 4D model (Define, Design, Develop, and Disseminate), the study involved 38 elementary students and 7 mathematics teachers in Indonesia. The learning media combined kinesthetic activities, local cultural contexts, and Google Teachable Machine to support meaningful geometry learning. Data were collected through observations, expert validation, questionnaires, interviews, and pre-test–post-test assessments. Quantitative data were analyzed using descriptive statistics and normalized gain (N-Gain), while qualitative data were examined through thematic analysis. The developed media demonstrated high feasibility, obtaining a validity score of 89.4% and a practicality score of 91.2%. Students' geometry achievement improved markedly, with mean scores increasing from 61.8 to 83.5 after implementation. The integration of physical movement and AI-based pose recognition promoted active participation, collaboration, and deeper conceptual understanding, while the use of traditional games enhanced students' cultural awareness and learning engagement. These findings indicate that combining ethnomathematics, gamification, and AI technology provides an effective and culturally responsive approach to elementary mathematics instruction, offering a practical contribution to the development of innovative learning models that improve conceptual understanding, motivation, and classroom engagement. **Keywords:** gamification, Teachable Machine, geometry learning, elementary school, traditional games



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License .

Corresponding Author:

Beti Istanti Suwandayani
Primary School Teacher Education Department
Faculty of Teacher Training and Education
Muhammadiyah University of Malang, Indonesia
Tlogomas Highway 246 Malang
beti@umm.ac.id

1. INTRODUCTION

Learning mathematics directed to form ability think systematic, logical and creative basis, as well as develop skills collaborative student (Jannah et al., 2024). In context education school basic, learning mathematics own role important in practice ability thinking and solving problem participant educate.(Suaib & Sutriyani, 2024) show that performance Study mathematics in chapter geometry Still many have not reach Criteria Minimum Completion Criteria (KKM). This happen Because learning conventional dominated method lecture Still often applied in class, so that participant educate tend feel fed up Because lack of involvement active in learning (Yuniarti et al., 2023). Teachers' limited ability to design instructional media that meet students' needs remains one of the challenges in mathematics learning at the elementary school level (Nurfadilah et al., 2026). In addition, students also have difficulty understand draft base geometry, such as angles, lines, symmetry, planes, and space, as well as difficulty connect connection between get up geometry (Nu'man & Azka, 2023). Problems This the more complicated Because amount formula geometry that must be memorized student the more many, so that mathematics often considered as difficult and less lessons interesting (Rosilawati et al., 2024).

Use game traditional in teaching mathematics own role important in repair understanding concept, improve involvement active, add motivation learning, and strengthen interaction social among students (Nur, 2025). Study Hanan & Alim (2023) show that with use game simple in the learning process, students can more easily understand draft geometry and also easier remember various formula mathematics. In context this game traditional own potential big For utilized as a relevant learning medium with life students. In addition to helping understand concept, game traditional also has values like religious, feeling of love homeland, ability independent, work same, and honesty the one that can help formation attitude and character (Sholikin et al., 2022). Collaborative learning provides opportunities for students to participate actively, share responsibilities, and develop cooperation skills

through group activities (Intan Rahayu et al., 2026). This is appropriate with opinion that school base is stage important in form character child since beginning (Suyitno & Setiawan, 2021). Development theory Vygotsky's social explanation that interaction between fellow humans is very important in development thinking children, so that play together Can become method excellent learning (Smolucha & Smolucha, 2021).

Study previously Already try create learning media that uses game traditional. One of the for example is development of board game - based media One of the games for example is development of board game - based media game traditional stilts coconut shell, with design similar games with ludo, which is used as tool learning in class (Maryanti et al., 2021). Although the study demonstrated the potential of traditional games to support mathematics learning, the developed media was still limited to conventional game-based activities and did not integrate emerging digital technologies to enrich the learning process. As a result, opportunities to provide more interactive learning experiences and immediate feedback for students were still limited. The rapid development of artificial intelligence offers an opportunity to address these limitations by supporting more adaptive and interactive learning through automatic activity recognition and real-time feedback. Therefore, this study combines the traditional stilt game with artificial intelligence using Teachable Machine to enhance students' engagement and conceptual understanding while preserving the educational and cultural values of traditional games. Aliya et al. (2025) disclose that utilization technology intelligence artificial in learning capable increase efficiency, personalization, quality of the learning process, and intelligence artificial can help educator in understand characteristics Study participant educate as well as give bait more back right. The use of technology in learning can increase students' attention, motivation, and participation by providing more interactive learning experiences (Azizah et al., 2025). Although so, no studies have integrated traditional games with artificial intelligence in elementary mathematics learning. Traditional games provide meaningful learning experiences by promoting physical activity, social interaction, and the integration of local cultural values. However, their educational potential can be further enhanced through the use of artificial intelligence, which enables more interactive learning experiences and supports real-time feedback during learning activities. The integration of AI also allows students' movements during gameplay to be directly connected with mathematical concepts, making abstract geometric ideas more concrete and easier to understand. Therefore, combining traditional games with artificial intelligence not only preserves local cultural heritage but also enriches mathematics learning through a more interactive and contextual approach. Rapid development technology also has the potential cause fading culture local Because participant educate tend more interested in modern technology (Muarifin & Zainal, 2024). Conservation efforts culture area needs implanted start since child age Elementary (Oktaviasary & Sutini, 2024)

Study about use of Teachable Machine in learning mathematics at the level school base Still Not yet Lots (Kurz et al., 2024) On the other hand, progress technology intelligence artificial give chance new to innovate in method study. One of the technologies that can used is Teachable Machine, which is a platform that allows users training learning models machine without need Can write code. With Thus, people who do not own background behind programming is also possibly understand draft intelligence artificial in a way easy (Prasad et al., 2022). Technology This own potential to support more learning interactive as well as make it easier student in visualize concepts abstract mathematics. Research previously has developing media based on game traditional and technology intelligence artificial in a way separate. However, not yet Lots integrated research game traditional based culture local with artificial intelligence technology in learning geometry school basic. In addition, the use of Teachable Machine in context learning mathematics based kinesthetic still very limited

Apart from using results study previously, research this is also based on observation the beginning of what was done towards teachers and students at school basis. Survey results against 38 students' classes IV–VI show that 47.4% of student Not yet Once learn mathematics through game traditional. However, most student show interest in methods more learning innovative. As many as 76.3% of students say want to Study mathematics with method different from usually, while 63.1% of students interested in learning by using game traditional. In addition, 60.5% of student expressed great interest If learning use technology such as a cell phone or laptop. Answer open student show that difficulty main in Study mathematics related with memorize formula and understanding question story. Observation results against 7 school teachers base show that 57.1% of teachers said that student Still

experience difficulty in understand material geometry and measurement. In addition, all teachers who answered state that Spirit Study student to material the Still lacking and necessary creativity in combined learning game as well as technology. New study This lies in integration game traditional stilts with Teachable Machine technology based on pose recognition for create learning interactive, contextual, and object - based geometry activity physique.

Based on condition mentioned, there are difference between learning mathematics as it should be nature interactive, contextual, and utilize technology with condition real in the field which is still dominated by methods conventional. Therefore, it is needed innovation in a combined learning process culture local with digital technology so that the concept complicated mathematics Can understood with more clear and real. Innovation in study This is development of learning prototypes which use gamification, namely combine game traditional stilts with Teachable Machine Learning technology. This research combines approach ethnomathematics local with machine learning technology to create experience learn more interactive and contextual in the material geometry in school basic.

2. RESEARCH METHOD

This study employed a Research and Development (R&D) approach using the 4D development model proposed by Thiagarajan, Semmel, and Semmel, which consists of four stages: Define, Design, Develop, and Design.

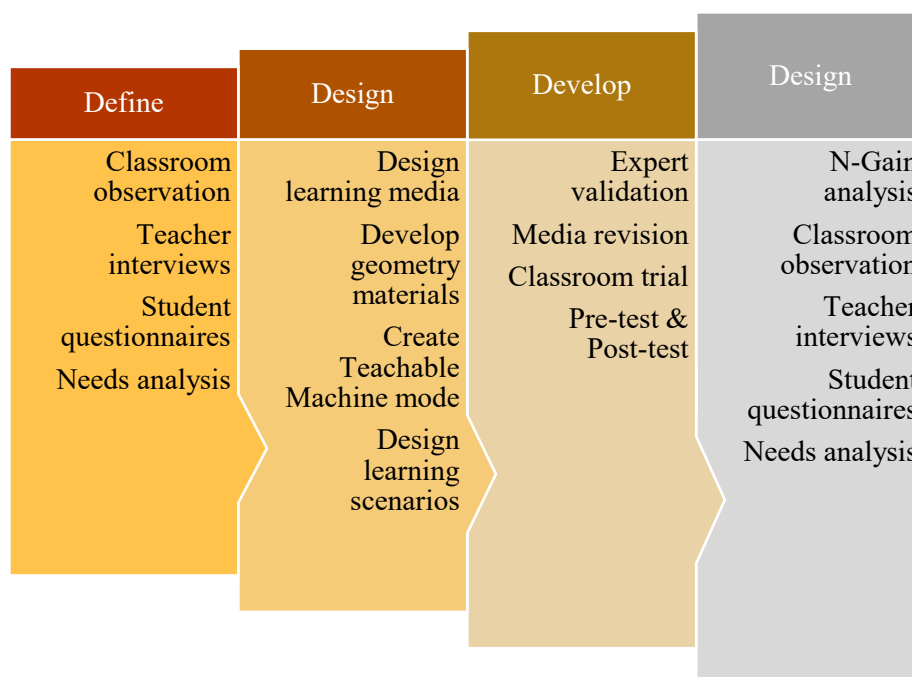


Fig 1. Research procedure based on the 4D development model

The research aimed to develop a gamification-based learning media integrating traditional stilt games with Teachable Machine technology for geometry learning in elementary schools. The study involved 38 elementary school students from grades IV–V and 7 elementary school teachers. The participants were selected using purposive sampling based on the consideration that students had previously studied basic geometry concepts and experienced difficulties in understanding abstract geometric materials. The teachers involved had experience teaching mathematics at the elementary school level. The define stage aims to identify learning problems and analyze students' needs related to geometry learning. Data collection was conducted through classroom observations, interviews with teachers, and questionnaires distributed to students. The analysis focused on identifying students' difficulties in understanding geometric concepts, learning motivation, and the need for interactive learning media integrating physical activities and technology.

The results of the preliminary analysis indicated that geometry learning was still dominated by conventional lecture methods, causing students to experience difficulties in understanding abstract

concepts such as angles, lines, and geometric shapes. Students also showed high interest in learning activities involving games and digital technology. At the Design stage, the prototype of gamified learning media was designed by integrating traditional stilt games with Teachable Machine technology. The learning activities were developed using kinesthetic approaches through body movements and hand poses representing geometry concepts.

Students' hand poses were used as input data for Teachable Machine training. Each pose represented specific geometric concepts, such as angles and geometric patterns. The collected visual datasets were categorized and trained using the Teachable Machine platform to develop an image recognition model capable of detecting students' movements in real time. The media design also included learning scenarios, instructional flow, game rules, and visual feedback mechanisms generated by the system to support interactive learning experiences.

At the Develop stage, the prototype was validated by two experts consisting of a mathematics education expert and an instructional media expert. The validation process aims to assess the feasibility of the developed media in terms of content suitability, instructional quality, usability, visual communication, and interactivity. The validation instrument used a Likert scale ranging from 1 to 5. The validation score was calculated using the following formula:

$$\text{Validity Percentage} = \frac{\text{Obtained Score}}{\text{Maximum Score}} \times 100\%$$

After revision based on expert suggestions, the media prototype was implemented through limited trials involving elementary school students. The trial aimed to evaluate practicality, student responses, and learning effectiveness. The Disseminate stage was conducted on a limited scale by introducing the developed media to teachers and students. At this stage, questionnaires were distributed to evaluate students' responses toward the implementation of the learning media. The dissemination process also aims to identify the strengths and limitations of the developed prototype as an initial evaluation for further development.

Quantitative data obtained from validation sheets, questionnaires, and learning outcome tests were analyzed descriptively using percentages and normalized gain (N-Gain) analysis. To evaluate the effectiveness of the developed learning media, the normalized gain (N-Gain) analysis was conducted based on student's pre-test and post-test scores. The N-Gain analysis was used to determine the magnitude of students learning improvement after the implementation of the developed learning media. The N-Gain formula used was:

$$g = \frac{\text{Posttest Score} - \text{Pretest Score}}{\text{Maximum Score} - \text{Pretest Score}}$$

Qualitative data obtained from observations and interviews were analyzed using the interactive model of Miles and Huberman consisting of data collection, data reduction, data display, and conclusion drawing. The combination of quantitative and qualitative analyzes was intended to provide comprehensive findings regarding the feasibility and effectiveness of the developed learning media.

3. RESULTS AND DISCUSSION

A. Define

Data collection at the stage analysis need done through questionnaires given to student For get description about experience learning, level understanding to material geometry, as well as interest they to use method more varied. Nursalam & Djaha (2023) learning questionnaire is instrument main in research data collection that produces numerical data For analyzed in a way statistics. The goal is get relevant, valid, and reliable information. Sugiyono (2023) researchers compile questionnaire as instrument data collection. Before used, questionnaire moreover formerly tested For ensure validity and reliability. Questionnaire arranged with a number of indicators that include habit Study students, difficulties faced in understand draft geometry, as well as interest to learning based activities and technology. The results obtained show that learning mathematics in progress Still tend dominated by methods lecture, so that involvement student in the learning process Still limited. Condition This impact

on low understanding draft as well as lack of interest Study students, especially on material that is abstract like geometry In addition, some student state that they experience difficulty in understand connection between get up geometry and remembering the formula used. Students also show more interest tall to learning that involves activity direct and use technology. This is show existence need to more learning interactive, contextual, and capable connect draft mathematics with experience real student. Fatimah & Sari (2025) approach contextual in learning mathematics emphasize relatedness draft with life everyday so that students more easy understand and apply it. Activities filling questionnaire by students at the stage This can seen in Figure 2. Data obtained used as base in designing and developing more learning media in accordance with characteristics as well as need student school base .



Fig 2. Filling questionnaire by students

B. Design

Design learning This utilizing hand poses as a medium for represent draft angle to be more easy understood by students. In activity learning, students requested lift left hand with amount finger certain as symbol big angle, namely One finger For corner one, two fingers For angles two and three finger For corner three. Approach This in line with style Study kinesthetic, namely method learning that emphasizes use activity physique as means main For understand material optimally. Therefore Supit et al. (2023) that, the pose used designed in a way simple to easy imitated, not confusing, and appropriate with characteristics student school the basis that is still is at the stage Study concrete.

Every pose done by the student recorded as visual data later utilized as material training on the Teachable Machine syste. At this stage this, poses are grouped based on amount finger shown for the system can learn difference characteristics evey movement in a way systematic. Training process This aim For increase ability system in recognize poses right, though there is variation position or difference between students. After the model is successful formed, system can identify the pose shown student in a way directly and provide bait come back in accordance with category the angle that has been determined. Feed return given functioning as means evaluation for student For evaluate suitability movements performed with the concepts studied. With existence response direct from system, students pushed For try return until get correct results, so that the learning process become more active and recurring. Ariani et al. (2026)in evaluation learning, assessment own role important as means giving bait come back for teachers. Assessment formative no only used for measure results learning, but also becomes a medium for reflection in improve the learning process. Example hand pose representation along with results detection can seen in Figure 3.

To further evaluate the effectiveness of the developed learning media, students' learning outcomes were analyzed using pre-test and post-test assessments involving 38 elementary school students. The mean pre-test score was 61.8, indicating that students' initial understanding of geometry concepts was still limited before implementation. After participating in the kinesthetic and AI-assisted learning activities, the average post-test score increased significantly to 83.5, demonstrating a substantial improvement in conceptual mastery. To quantify the magnitude of this improvement, a

normalized gain (N-Gain) analysis was conducted. The N-Gain score obtained was 0.68, which falls into the medium category ($0.30 < g < 0.70$). This medium N-Gain score indicates that integrating traditional stilt games with Teachable Machine technology effectively bridges abstract geometric concepts with concrete physical movements, fostering real-time self-correction, high engagement, and meaningful conceptual understanding among students.

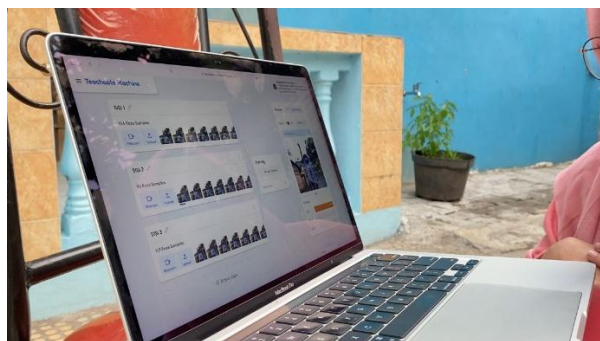


Fig 3. Design of Teachable Machine's pose detection

C. Develop

At the Develop stage, the prototype learning that has been designed Then implemented through trials limited to student grades IV–V of school basic in the environment around researchers. Trial done For see implementation learning, response students, as well as level understanding draft after media use. Activities learning designed in form activities that combine game traditional stilts with Teachable Machine technology so student can Study through experience directly. At the start of the activity, students introduced return with draft geometry that will studied, then to be continued with practice hand poses that have been designed at the stage previously. The pose used as representation draft corner, where students lift hand with amount finger certain in accordance categories that have been trained in system. Each pose displayed student Then detected by the Teachable Machine and generates bait visual feedback directly. This process help student understand relatedness between movements performed with draft geometry in question. Representation of student poses and results detection can seen in Figure 4.



Fig 4. Student pose detection using Teachable Machine

Next, students do activity play stilts For form pattern geometry through movement. Students directed For walk follow track certain so that forming lines and shapes flat like triangle. Activity This done in a way in groups, so that student can each other interact, work the same, and discuss concepts learned. Trial activities learning based stilts can seen in Figure 5. Via activity this, students No only understand form visually, but also experientially direct How draft geometry formed from movement. Satriyo et al. (2026) the use of audio-visual media contributes significant in increase quality learning, because capable interesting attention students, clarify material that is abstract, and increase motivation and results study. Trial results show that student more active and enthusiastic during the learning

process ongoing. Students seen more easy understand draft geometry Because involved direct in connecting activities movement with draft mathematics. In addition, the use of Teachable Machine provides experience learn more interesting Because existence bait helpful visual feedback student correcting understanding they in a way directly. In overall, stage development This show that prototype designed learning capable increase involvement students and help they understand draft geometry in a way more concrete and meaningful. Manihuruk & Sutabri (2024) gamification present experience learn more interactive and dynamic. In addition, the approach This participate increase flexibility as well as independence student in the learning process. The use of multimedia also helps participant educate in understand material with more easy.



Fig 5. Trial learning-based stilts

The expert validation results indicated that the developed learning media obtained an overall validity score of 89.4%, which is categorized as highly valid. This result demonstrates that the developed media meets the required criteria in terms of content, instructional design, usability, and visual presentation, making it appropriate for implementation in elementary geometry learning. Furthermore, the practicality assessment conducted during the limited classroom trial resulted in a score of 91.2%, indicating excellent usability. The findings suggest that the developed learning media is easy to use and supports interactive learning activities by integrating traditional stilt games with Teachable Machine technology.

D. Disseminate

Disseminate stage is carried out in a way limited as form evaluation beginning to prototype learning that has been developed. At this stage this is the test result learning delivered return to student for get response related use of media based game combined stilts with Teachable Machine technology. Data collection was carried out through questionnaires given after activity learning finished. Evaluation in education functioning for obtain data regarding level achievement participant educate to objective learning, at the same time evaluate effectiveness methods, strategies, and activities implemented in the learning process (Febriana, 2021). With Thus, the stage This no only focus on response students, but also become base for evaluate success beginning from implementation of the developed media. Based on the data obtained, students show responses that tend to positive to learning carried out. They feel that activity Study become more interesting, no monotonous, and capable increase interest in follow the learning process. In addition, students also convey that draft geometry more easy understood when served through activities that involve movement body and use technology in a way simultaneously. Merger activity No solely focus on aspects form, but also encourage involvement physique in a way active (Vriegde & Pudyaningtyas, 2025).

This matter show that approach integrated learning element kinesthetic and technology can help student in connect experience Study with concepts learned (Fatimah & Sari, 2025) student respond positive use of interactive media Because help they understand material with easier and more interesting. Teachers use various strategies for adapt learning with style Study students, including visual and kinesthetic approaches. Involvement active during activity ongoing also has an impact on increasing focus, participation, and trust self-student in try and fix error in a way independent. Further, use questionnaire as instrument evaluation allows researchers get a clearer picture systematic about

experience Study students. The results obtained show that part big student feel learning become more fun and helpful they understand material in a way more concrete. This indicates that integration game traditional with technology no only increase motivation learn, but also give experience learn more meaning that. Husna & Supriyadi (2023) modern learning media is seen own role important in increase motivation Study student as well as support achieving an effective and efficient learning process. Findings research also shows that use of learning media give significant influence in grow and improve motivation Study students. Although Thus, the implementation stage This Still limited in number relative subject small, so that required testing advanced with more coverage wide for get more results representative. In overall, prototype learning This own potential for developed as alternative innovation capable learning create atmosphere learn more interactive, contextual, and appropriate with characteristics student school base.

To further evaluate the effectiveness of the developed learning media, students' learning outcomes were analyzed using pre-test and post-test assessments. The average pre-test score was 61.8, indicating that students' initial understanding of geometry concepts was still limited before the implementation of the learning media. After participating in the learning activities, the average post-test score increased to 83.5, demonstrating an improvement in students' understanding of geometry concepts. Furthermore, the normalized gain (N-Gain) analysis yielded a score of 0.68, which falls into the medium category. These findings indicate that the developed learning media was effective in improving students' conceptual understanding of geometry by integrating traditional stilt games with Teachable Machine technology into the learning process.

4. CONCLUSION

Study This produce prototype integrated learning game traditional stilts with Teachable Machine technology on materials geometry in school base through Research and Development approach. Prototype designed for connect activity physique with draft mathematics so that student can understand material in a way more concrete. Implementation results show that combined learning movement body and technology capable create experience learn more interactive as well as help student in linking draft abstract with experience real. Approach this also provides variation in the previous learning process tend nature conventional. In addition, the results evaluation through questionnaire show that student give response positive to learning carried out. The questionnaire results showed a practicality score of 91.2%, indicating excellent usability. This finding demonstrates that students responded positively to the developed learning media and considered it easy to use, engaging, and supportive of geometry learning activities. They show improvement interest, involvement, and convenience in understand draft geometry through activities carried out in a way direct. This positive response was also supported by the improvement in students' learning outcomes, with the average pre-test score increasing from 61.8 to 83.5 in the post-test. Furthermore, the normalized gain (N-Gain) score of 0.68 indicates a medium level of improvement in students' conceptual understanding of geometry. Integration of elements culture local and technology participate enrich context learning so that more relevant with life students. Although study This Still done in scale limited, the findings obtained show that prototype learning This own potential For developed more carry on as alternative innovation capable learning increase quality learning mathematics at school base .

REFERENCES

- Aliya, F., Anhar, M., Syamsiah, & Marhamah. (2025). Metode Sorogan Pembelajaran di Pesantren Memanfaatkan Teknologi Kecerdasan Buatan. *SAP (Susunan Artikel Pendidikan)*, 9(3).
- Ariani, S. Y., Oktrifanty, E., Agustin, M., & Maharani, N. S. (2026). Peran Tes Formatif Sebagai Alat Umpan Balik dalam Proses Evaluasi Pembelajaran. *Jurnal Transformasi Pendidikan Modern*, 7(1), 17–27. <https://ejournals.com/ojs/index.php/jtpm>
- Azizah, N., Ariefian, N., Febriyanti, F., Karomah, N., Amelia, I. J., & Suwandayani, B. I. (2025). *Pemanfaatan Smart TV untuk Mengoptimalkan Keterampilan Menyimak pada Pembelajaran Bahasa Indonesia Siswa Kelas 3 di SD Riyadul Arkham*. 9(2). <http://studentjournal.umpo.ac.id/index.php/edupedia>
- Fatimah, F., & Sari, Y. I. (2025). Merekonstruksi Pembelajaran IPS: Media Interaktif, Gaya Belajar Siswa, dan Kreativitas Guru dalam Menumbuhkan Pemahaman Kontekstual. *Dinamika Sosial: Jurnal Pendidikan Ilmu Pengetahuan Sosial*, 4(2), 150–162. <https://doi.org/10.18860/dsjpips.v4i2.17921>

- Febriana, R. (2021). *Evaluasi Pembelajaran* (B. S. Fatmawati, Ed.). Bumi Aksara.
- Hanan, M., & Alim, J. (2023). Analisis Kesulitan Belajar Matematika Siswa Kelas VI Sekolah Dasar pada Materi Geometri. *Journal of Mathematics Education*, 2.
- Husna, K., & Supriyadi. (2023). Peranan Manajemen Media Pembelajaran Untuk Meningkatkan Motivasi Belajar Siswa. *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, 4(1), 981–990. <https://doi.org/10.37680/almikraj.v4i1.4273>
- Intan Rahayu, A., Putri Amanda, A., Nur Azizah, H., Diah fatmawati, L., Hafidz Nur Abdillah, C., & Istanti Suwandayani, B. (2026). *Peran Guru dalam Mendorong Aktivitas Gotong Royong melalui Lingkungan Belajar Kolaboratif*. 7(1), 50–56.
- Jannah, R., Soraya, R. A., Suriansyah, A., & Cinantya, C. (2024). Kemampuan Berpikir Kritis dalam Pembelajaran Matematika Terhadap Hasil Belajar Di Sekolah Dasar. *MARAS: Jurnal Penelitian Multidisiplin*, 2(4), 1991–1998. <https://doi.org/10.60126/maras.v2i4.550>
- Kurz, T. L., Jayasuriya, S., Swisher, K., Mativo, J., Pidaparti, R., & Robinson, D. T. (2024). The Impact of Teachable Machine on Middle School Teachers' Perceptions of Science Lessons after Professional Development. *Education Sciences*, 14(4). <https://doi.org/10.3390/educsci14040417>
- Manihuruk, D., & Sutabri, T. (2024). Perancangan Bahan Ajar Berbasis Multimedia untuk Membangun Pengalaman Belajar yang Menarik dan Interaktif. *Journal of Creative Student Research*, 2(6), 48–53. <https://doi.org/10.55606/jcsr-politama.v2i6.4553>
- Maryanti, E., Ekok, A. S., & Febriandi, R. (2021). Pengembangan Media Board Games Berbasis Permainan Tradisional Egrang Batok untuk Siswa Sekolah Dasar. *Jurnal Basicedu*, 5(5), 4212–4226. <https://doi.org/10.31004/basicedu.v5i5.1486>
- Muarifin, & Zainal. (2024). Luntunya Moralitas Pendidikan Di Era Artificial Intelligence. *Jurnal Creativity*, 2(2), 221–234. <http://creativity.masmubata-bata.com/index.php/creativity>
- Nu'man, M., & Azka, R. (2023). Kesulitan Siswa dalam Menyelesaikan Masalah Geometri. *Journal in Mathematics Education*, 3(2). <https://doi.org/10.14421/polynom.2023.32.49-55>
- Nur, M. A. (2025). Efektifitas Permainan Tradisional Terhadap Hasil Belajar Matematika Siswa Sekolah Dasar: Literature Review. *Jurnal Ilmiah Matematika (JIMAT)*, 6(1), 165–182. <https://doi.org/10.63976/jimat.v6i1.808>
- Nurfadilah, A., Salsabila, Z., Dwi, N., Nayla, W., Rahmanda, F., & Suwandayani, B. I. (2026). Eksplorasi Tantangan Integrasi Media Pembelajaran Matematika di Kelas V Sekolah Dasar: Pendekatan Studi Kasus di SDN 2 Ngepeh. *Jurnal Pendidikan Dan Pembelajaran Matematika*, 5(1).
- Nursalam, & Djaha, A. (2023). Pelatihan Pembuatan Kuesioner Penelitian Bagi Mahasiswa Prodi Administrasi Fisip Universitas Nusa Cendana. *Jurnal Pengabdian Inovasi Dan Teknologi Kepada Masyarakat*, 3(1), 25–30.
- Oktaviasary, A., & Sutini, A. (2024). Gempuran Budaya Modern terhadap Budaya Lokal Generasi Alpha: Tinjauan Literatur Review. *Jurnal Onoma: Pendidikan, Bahasa Dan Sastra*, 10(4). <https://e-journal.my.id/onoma>
- Prasad, P. Y., Prasad, Dr. D., Malleswari, D. N., Shetty, M. N., & Gupta, Dr. N. (2022). Implementation of Machine Learning Based Google Teachable Machine in Early Childhood Education. *International Journal of Early Childhood Special Education (INT-JECSE)*, 14(03), 2022.
- Rosilawati, Sulistiani, S., Putri, S., Indriani, Y., Witono, H., & Ibrahim. (2024). Analisis Kesulitan Belajar Matematika Materi Bangun Datar. *Jurnal Teknologi Pendidikan Dasar*.
- Satriyo, M., Syari, P., Pratiwi, D., & Episiasi. (2026). Studi Literatur: Analisis Penggunaan Media Audio Visual pada Pembelajaran IPS di Sekolah Dasar. *Tarbiyatul Ilmu: Jurnal Kajian Pendidikan*, 3(3), 349–359.
- Sholikin, M., Fajrie, N., & Ismaya, E. A. (2022). Nilai Karakter Anak Pada Permainan Tradisional Gobak Sodor dan Egrang. *Jurnal Educatio FKIP UNMA*, 8(3), 1111–1121. <https://doi.org/10.31949/educatio.v8i3.3035>
- Smolucha, L., & Smolucha, F. (2021). Vygotsky's Theory in-Play: Early Childhood Education. *Early Child Development and Care*, 191(7–8), 1041–1055. <https://doi.org/10.1080/03004430.2020.1843451>
- Suaib, M., & Sutriyani, W. (2024). The Effectiveness of Using Assemblr Edu Application Media Based on AR (Augmented Reality) on Understanding the Concept of Elementary School Geometry 3D. *Indonesian Journal of Education & Mathematical Science*, 5(3), 190–196. <https://doi.org/10.30596/ijems.v5i3.21018>
- Sugiyono. (2023). *Metode Penelitian Kuantitatif Kualitatif dan R&D* (2nd ed.). Alfabeta. www.cvalfabeta.com
- Supit, D., Meiske Maythy Lasut, E., Jerry Tumbel, N., Klabat, U., Airmadidi Bawah, J., & Utara, S. (2023). Gaya Belajar Visual, Auditori, Kinestetik terhadap Hasil Belajar Siswa. *Journal on Education*, 05(03), 6994–7003.

- Suyitno, & Setiawan, F. (2021). Penguatan Pendidikan Karakter melalui Permainan Tradisional di Sekolah Dasar Muhammadiyah Se-Kecamatan Depok. *EDUKASI: Jurnal Penelitian & Artikel Pendidikan*, 13(1), 13–22. <http://journal.ummgl.ac.id/nju/index.php/edukasi>
- Vriegde, M. I., & Pudyaningtyas, A. R. (2025). Peningkatan Kemampuan Mengenal Bentuk Geometri Anak Usia 5–6 Tahun Melalui Permainan Estafet. *Early Childhood Education and Development Journal*, 12(1), 147–157. <https://doi.org/10.20961/ecedj.v%vi%i.108622>
- Yuniarti, A., Safarini, F., Rahmadia, I., & Putri, S. (2023). *Media Konvensional dan Media Digital Dalam Pembelajaran*. 4(2), 2023

Development of a Flipbook-Based E-Module for Junior High School Geometry Learning: Its Effectiveness in Enhancing Student Engagement and Learning Motivation

Sitti Zuhaerah Thalhah¹, Dwy Risky Arifanti², Sumardin Raupu³, Irma T.⁴

^{1,2,3,4}Faculty of Tarbiyah and Education, Universitas Islam Negeri Palopo, South Sulawesi, Indonesia

¹hera@uinpalopo.ac.id, ²dwi_risky_arifanti@uinpalopo.ac.id, ³sumardin_aldhy@uinpalopo.ac.id, ⁴irma@uinpalopo.ac.id

ABSTRACT

Low student engagement and learning motivation remain challenges in geometry instruction in junior high schools because abstract geometric concepts are often presented through less interactive teaching materials. This study aims to develop a flipbook-based e-module for geometry using the ADDIE model and test its validity, practicality, and effectiveness in improving student engagement and learning motivation. The approach used was research and development with mixed methods. The study subjects included 125 seventh-grade students from a public junior high school in Palopo. Data were collected using expert validation sheets, interviews, practicality questionnaires, student engagement observations, and learning motivation questionnaires. The results showed that the e-module achieved 95.87% validity from media experts and 85.46% from material experts, categorizing it as very valid. Its practicality reached 83.38%, categorizing it as very practical. The effectiveness test showed an N-Gain of 70.79% for student engagement and 87.00% for learning motivation, both in the high category. These findings confirm that the flipbook e-module is a valid, practical, and effective learning medium for increasing student engagement and learning motivation in geometry learning.

Keywords: E-Module, Flipbook, Student Engagement, and Learning Motivation



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Sitti Zuhaerah Thalhah,
Departement of Mathematics Education,
Universitas Islam Negeri Palopo,
Jalan Agatis, Kelurahan Balandai, Kecamatan Bara, Kota Palopo, Indonesia.
hera@uinpalopo.ac.id

1. INTRODUCTION

Student engagement in mathematical pedagogy is evident in many areas of modern education. Immersion is defined by behavioral, emotional, and cognitive participation by learners (Sattar et al., 2022). It has been suggested that learners who demonstrate high levels of emotional, behavioral, and cognitive engagement are more likely to participate in lessons and persevere and maintain focus through the completion of learning tasks (Khan et al., 2023). In contrast, diminished engagement is characterized by low levels of participation in learning activities and focus, which, in turn, leads to low levels of achievement (Mizani et al., 2022; Santi et al., 2022). In the context of pedagogy specific to learning and teaching geometry, it is important to use resources that engagingly represent abstract concepts of mathematics. This is necessary because learners tend to experience mental exhaustion and cognitive and conceptual struggles (Mulyana et al., 2023).

Along with student engagement, academic drive is a core component of successful mathematical instruction (C. Wang et al., 2022). The drive to learn propels students to participate, break down the barriers of the learning process, and to develop a long-term love of learning. Unfortunately, the drive that lower-secondary students have for learning mathematics is very low, and is often worsened by lesson delivery methods, such as pedagogy that is monotone, boring, and devoid of modern instructional technologies (Cohen et al., 2023). Traditional instructional methods and materials that are overly text-based often fail to engage students and lead to the achievement of shallow and short-lived learning outcomes (Bryant, 2023). As a result, there is an urgent need to design instructional technologies that deliver teaching and learning content while capturing and sustaining students' interest, energy, and engagement at all levels of the learning process.

Understanding the design of flipbook-infused modules of digital mathematics within the context of digitally based learning systems provides an opportunity to conceptualize how the integration of technology can disrupt and improve the learning (Kurnianto et al., 2022). An e-module is an electronically formulated unit that encourages individualized learning through the integrated and purposeful use of various forms of media, including text, images, audio, animations, and videos (Cavanagh & Kiersch, 2023). The use of these e-resources provides an enhanced level of ease of use, as they can be used from anywhere and at any time through the possession of portable, digital technology (Mesra et al., 2025). In addition, digital modules provide a more engaging and interactive format of the e-content compared to the traditional printed formats, thereby encouraging students to be more motivated to learn (Manggala et al., 2024).

Multimedia Learning Theory suggests that contemporary media comprising prose, graphics, and animations can help learners achieve higher levels of mastery with respect to a particular concept. This occurs because learners process information through the text and the graphics via their verbal and visual cognitive channels, respectively. According to Mayer, the most effective means of information transfer occurs when learners are presented with visual and verbal materials that work in tandem (Mayer & Moreno, 2003). This is especially true for the visual-based media of the mathematics discipline, namely, the spatial areas of the discipline, such as geometry, which requires a high level of spatial reasoning and the ability to visualize concepts internally. Therefore, the placement of interactive multimedia in electronic instructional units is relevant for the purpose of transforming the abstract nature of mathematical proofs and theorems into concrete and comprehensible representations that are sound and logical (Adiastuty et al., 2025).

Digital flipbooks are an innovative tool for electronic educational media that provide a near-real-world interface for content delivery that includes the sounds and motions of physically turning a page of a book (Elfina et al., 2024). In this framework, the medium itself provides higher levels of user interactivity, visualization, and navigation (Putri et al., 2025). Historically, research shows that flipbooks enhance student motivation to learn due to the combination of design and delivery that is visually stimulating and engaging (Artal-Sevil et al., 2020; Cevikbas & Kaiser, 2022; Kim et al., 2007). Additionally, digital flipbooks offer the opportunity to add instructional videos, fluid on-screen movement, and self-check assessments that promote user engagement (Reza et al., 2025; Tani et al., 2022). In the area of mathematics, flipbooks have been shown to deepen student interest by describing a measurement in a more contextualized and visually stimulating presentation (Amiruddin & Imama, 2025).

Geometric concepts found at the lower secondary benchmark involve high degrees of abstraction and require a significant degree of internal visual imagination. This area challenges students to go beyond rote memorization of surface area and volume formulas. Instead, they must parse the complexity of the relational syntactic structures of the physical world (Mucti & Nurmala R, 2025). Attempting to learn geometry is associated with insufficient spatial visualization and is compounded by the use of pedagogical tools that provide insufficient concretization of the ideas (Hidayah et al., 2025). (Hidayah et al., 2025) Therefore, the use of an e-module with a flip-book structure that contains geometric simulations, illustrations, and animations is essential to enable learners to build an understanding of spatial systems (Supratman & Tundreng, 2025; Žakelj & Klančar, 2022).

This study is similarly founded on the idea of student engagement as a key indication of learning achievement. Student engagement encompasses three main dimensions, namely behavioral engagement, emotional engagement, and cognitive engagement (M.-T. Wang & Fredricks, 2014). Behavioural engagement refers to students' active participation in learning activities; emotional engagement represents students' interest and affective responses to learning; and cognitive engagement refers to students' efforts to comprehend the information (Martinjak, 2024). High levels of student involvement have been linked to increased motivation, learning persistence, and academic accomplishment (Torres & Statti, 2023).

Academic drive refers to a combination of internal motivators and external stimuli that influence the degree to which a student expends effort on academic tasks. Following the tenets of John M. Keller, this drive can be further developed by structuring educational environments that meaningfully and immediately connect to students, help them develop a sense of self-efficacy, and provide satisfaction (Bayounes Bayouth Saâdi, I., & Kinsuk, 2022). Each of the educational tasks contained in flipbook-

based e-modules is designed to be aesthetically pleasing, easy to use, and contain a wide variety of learning tasks. Therefore, it is reasonable to assume that flipbook technology can be used to improve student motivation and engagement in the field of mathematics to a high level (Asmorowati & Pratama, 2025).

Other research on the use of flipbooks in mathematics learning reported an increase in learning motivation after using the media (Astuti & Andriani, 2025). Unayah et al., (2024) also highlighted that flipbooks are effective in increasing student engagement, as interactive features such as multimedia elements and simulations support various learning styles, which in turn increases student engagement. Although various studies have reported that the use of flipbook-based learning media still focuses on students' cognitive aspects and has not given adequate attention to simultaneously increasing student engagement and learning motivation. The implementation of flipbooks in geometry learning is generally still limited to presenting material in digital form without integrating various multimedia features that can encourage active student involvement during the learning process. In fact, the characteristics of geometry material require learning media that not only present information but also can visualize geometric objects dynamically, thereby building student engagement and motivation.

Based on these gaps, this study offers novelty through the development of a flipbook-based e-module that integrates various multimedia features, such as learning videos that can increase student's emotional engagement, animations to facilitate understanding of geometric concepts or cognitive engagement, interactive Wordwall-based exercises to increase behavioral engagement through direct feedback and interactive exercises, flexible digital navigation to increase learning motivation, and geometric visualization to strengthen cognitive engagement. In contrast to previous studies that mostly evaluated learning outcomes, this study analyzed the use of e-modules on student engagement that includes behavioral, emotional, and cognitive dimensions, as well as learning motivation. In addition, the developed product was evaluated based on aspects of validity, practicality, and effectiveness using the ADDIE development model, thus providing a more comprehensive picture of the quality of the resulting learning media.

Reflecting on the aforementioned theoretical and empirical currents, the engineering of a flipbook-infused e-module rests upon a solid foundational bedrock for deployment in junior high geometry classes. The orchestration of interactive multimedia inside this electronic capsule is projected to demystify complex geometric theorems through explicit visual modeling, while concurrently sparking a renaissance in student engagement and academic drive throughout the pedagogical lifecycle. This project aims to fill a significant gap by developing and validating the efficacy of a flipbook-based e-module as an innovative instructional tool designed to meet the changing needs of mathematics education in the age of digital technology.

2. RESEARCH METHOD

This study established an intensive mixed-methods framework that seamlessly integrated research and development (R&D) stages along with rigorous analysis of product effectiveness. The focus of the engineering pathway was on the construction of a digital module for teaching geometry that was interactive and integrated with flip books. This digital tool was then placed in an active teaching and learning environment to evaluate its effectiveness in increasing student engagement and motivation. The implementation of both of these strategies enabled the researcher to go beyond the simple construction of a teaching tool to evaluate its impact on the students' educational experience and academic performance.

For this study, the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model was used. This framework was chosen for its systematic nature and for the fact that each of the stages can be revisited and/or adapted. These characteristics mean that this model lends itself easily to the development of digital learning modules. In addition, the ADDIE framework was designed in such a way that there can be continuous evaluations at each of the stages, thus ensuring that the learning module could be constructed and design improved in response to the findings and feedback collected throughout the study (Suwanto et al., 2023).

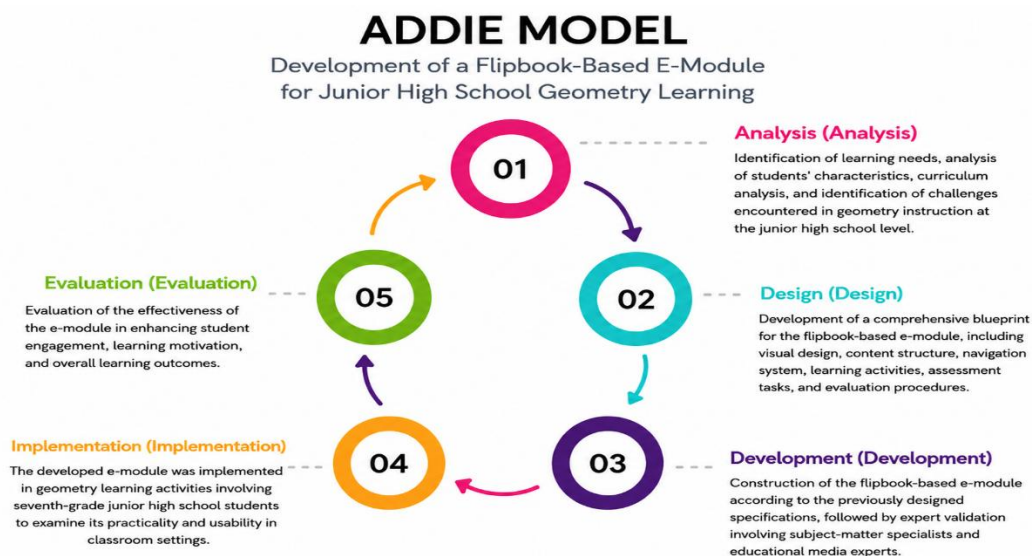


Fig. 1. ADDIE Model Procedure Design

This study's sample included 125 seventh graders at a public secondary school on A. Kambo Street in Salekoe Village, East Wara District, Palopo City, South Sulawesi Province, Indonesia. Focusing on this seventh-grade group was based on the belief that students in this age group (12-14) are in the process of moving from the concrete to the formal operational stages of thinking and reasoning based on Jean Piaget's theory of the stages of cognitive development.

The data gathered in this study included the following four categories: (1) qualitative interviews conducted to identify key product features and form, (2) expert credibility and usability evaluations conducted by interface and content specialists to assess the systemic integration of the flipbook-based digital learning resource, (3) participatory observations of students conducted before and after the geometry activity to assess the behavior, feelings, and thoughts level of engagement, and (4) participants' intrinsic engagement surveys conducted at the start and end of the geometry unit to evaluate the level of engagement in the learning activities.

The mean values of the validation and usability criteria were determined by calculating the percentage of scores assigned to each of the evaluation categories within the expert assessment and feedback forms. The baseline criteria of the validation and usability dimensions are presented in Table 1.

Table 1. Validity and Practicality Criteria

No	Criteria	Percentage (%)
1	Very Valid/Practical	81 – 100
2	Valid/Practical	61 – 80
3	Moderately Valid/Practical	41-60
4	Less Valid/Practical	21 – 40
5	Invalid/Impractical	0 – 20

Empirical metrics reflecting student engagement levels were collected using a quad-point assessment matrix. In the matrix, points were assigned as follows: 1 for no engagement, 2 for some engagement, 3 for considerable engagement, and 4 for outstanding engagement. The average percentage totals of these observational logs were reviewed as analytics of student engagement with the geometry curriculum. Table 2 describes the primary classification levels for the assessment of this particular type of multi-faceted learner engagement

Table 2. Student Engagement Criteria

No	Criteria	Percentage (%)
1	Very High	75 – 100

2	High	50 – 74,99
3	Moderate	25 – 49,99
4	Low	0 – 24,99

Scholastic drive was quantified by computing the mean percentage metrics aggregated from the participants' self-reported feedback on the motivational inventory. The resulting numerical aggregates were subsequently categorized in alignment with established benchmarks to determine the precise intensity of the students' academic enthusiasm. The structural boundaries governing these learning motivation classifications are explicitly delineated in Table 3.

Table 3. Learning Motivation Criteria

No	Kriteria	Persentase (%)
1	Very High	85 – 100
2	High	69 – 84
3	Moderate	53 – 68
4	Low	37 – 52
5	Very Low	20 – 36

The instructional potency of the flipbook-infused digital module in amplifying student immersion and academic drive was scrutinized utilizing the normalized gain (N-Gain) metric formula. The standardized N-Gain interpretive benchmarks harnessed within this analytical framework are explicitly detailed in Table 4.

Table 4. Learning Motivation Effectiveness Criteria

No	Criteria	Skor (%)
1	High	$g > 70$
2	Moderate	$30 \leq g \leq 70$
3	Low	$g < 30$

3. RESULTS AND DISCUSSION

This research and development endeavor culminated in the synthesis of a flipbook-infused electronic module designed as an innovative pedagogical vehicle for lower-secondary geometry instruction. The empirical insights harvested across each successive milestone of the ADDIE developmental framework are systematically detailed below:

A. Analysis Stage

The first step in the research and development process for the flipbook-based electronic module is the diagnostic stage. As part of this first step, two major activities were conducted: a needs assessment and a curriculum review. The needs assessment was mainly based on data obtained from observations of a state secondary school in Palopo City. The data collected showed that the school had a very limited collection of digital instructional media, resulting in the lack of many digital learning resources. From the observations, it was found that the school had a sufficient ICT infrastructure and a policy to permit the use of smartphones for educational purposes, thus enabling the use of digital resources in the school. Given the lack of digital resources, the use of an electronic module to create a digital learning resource was a valid option. The existing materials for geometric instruction at the school were very limited, so the provision of a flip book-based e-module would be the very first resource of this type for the 7th grade at this public junior high school in Palopo City. There were no interactive multimedia resources or audio-visual materials available for instruction in plane geometry.

According to findings from this investigation, the institution appears to have some satisfactory infrastructural scaffolding that includes access to smartphones, with internet connectivity that may be used during instruction. However, these resources remain largely unutilized for foundation-driven technology in the realm of mathematics education. As a result, the provision of a flipbook-type e-

module is a highly warranted intervention in the construction of a more interactive and dynamic learning environment that is in demand.

Qualitative interviews with seventh-grade math teachers in Palopo City public junior high schools were conducted to set boundaries for the curricular audit. From the interviews, it was found that the school implemented the Merdeka Curriculum with a high level of implementation. However, there are still some problems with the realities of the psychology of students who consider some of the math subjects trivial, and some even dislike math. From the baseline audit of student participation in learning geometry, the level of participation of students was found to be moderate. This condition is not permanent and varies with the learning media used to sustain student interest and motivation to learn. Students show relatively high levels of interest, participation, and focus in learning geometry, proportional to the learning media used by the teachers. This condition is a big challenge for teachers to provide geometry learning media to overcome the challenge of low student engagement and low student motivation to learn geometry. Overall, the findings show that the level of student engagement in learning is highly dependent on the learning media used by the teachers. If the learning media used are attractive and highly interactive, students are highly engaged and motivated to learn geometry.

Recent discoveries provide empirical evidence where studies show that electronic modules enhance students' immersion because they integrate elements of design, clarity, and frameworks that foster communication incredibly. Further, digital learning media provide an opportunity for students to take charge of their learning in a more fulfilling manner and, therefore, are completely in line with the learning styles of the digital age (Anggitarachmawati, 2025). In the teaching of geometry, the use of visual and interactive tools is a must, since the teaching and learning of plane geometry theorems requires clear visual images to aid the learners in understanding the relations of concepts and the forms of the geometric figure (Žakelj & Klančar, 2022).

Given the reported findings, the development of a flipbook-based e-module for the teaching of geometry is a great step forward for the school in meeting the needs of the learners, the requirements of the school as a whole, and the demands of contemporary teaching for the 21st century as it relates to the use of technology. For this reason, the information collected in this functional study represents the basis for the development of a teaching tool that is effective, relevant, and highly catered to the needs of the target users.

B. Design Stage

As part of this architectural milestone, the design of the pedagogical content and the digital tools was aligned with the seventh-grade learners' cognitive profile at the State Secondary Schools in Palopo. The design of the digital learning platform was developed through a number of strategic initiatives, including the following: (1) alignment of the structural elements of the teaching and learning modules; (2) articulating the expected learning outcomes and the relevant curriculum components to be embedded in the digital learning platform; (3) sourcing credible literature to support the content of the digital learning platform; (4) creating a Canva graphic design platform and a secured workspace; (5) creating an author account on the Heyzine Flipbook; (6) selecting an appropriate interface template; (7) designing using the graphic elements in the Canva workspace; (8) adding animations in Canva that aligned with the design theme and content with the focus on the target audience for the user testing of the design; (9) adding content headings, navigation instructions, learning outcomes, core content, solved examples, placeholders for supplementary media related to the practical exercises, and graphic placeholders for the Wordwall games; and adding a bibliography; (10) after completing the design in Canva, the layout was exported so that the Heyzine Flipbook was created with no effort once the first layouts were saved; (11) once the Heyzine Flipbook was created, the animated prototype was finalized by layering games and activities on the graphics and the instructional videos; (12) embedded the instructional videos. YouTube hyperlinks would integrate into the live digital presentation instantaneously. Additionally, Wordwall diagnostic quizzes would embed via Wordwall's HTML code sequencing. (13) After all multimedia elements are incorporated, the e-module will be finalized and ready for distribution as a stand-alone digital application, a web link, or a QR code.

The prototype of the electronic module was developed using the combined collaborative capabilities of Canva and Heyzine Flipbook. Both tools facilitate the creation of highly engaging, flexible educational software. The thoughtful placement of design and motion elements within the e-module aims to engage learners and stimulate interest in learning. This design philosophy supports

historical research that shows how design and aesthetic elements in instructional materials can capture student interest and improve engagement at every level of the educational system (Reza et al., 2025; Tani et al., 2022). In the teaching of mathematics and especially in the teaching of spatial concepts of plane geometry and figures, one must internalize and visualize concepts. This helps students interpret the structural and abstract propositions of geometry and supports them in developing more tangible and more integrated mental constructions (Žakelj & Klančar, 2022).

This segment of the research aims to explore the efficacy of a flipbook-style e-module in enhancing a learner's ability to construct and deconstruct knowledge. Flipbook-style e-modules containing video vignettes, together with clickable links to PlayPosit, provide a dual-channel learning experience. Gamified Wordwall tasks encourage the active participation of learners during self-assessments. The combination of video, animation, and gamified diagnostics within an e-module is especially powerful for reinforcing the understanding of core concepts and motivating learners. Additionally, this study sought to enhance the usability and navigational ease of the flipbook-style e-modules. This approach is in line with the consensus among instructional design experts that the e-modules and media used must adhere to principles of design and usability to maximize the efficacy of the instructional content (Carroll & Kop, 2016; Cavanagh & Kiersch, 2023).

C. Development Stage

During the advancement milestone, an authentication procedure was carried out to preserve the structural and educational integrity of the built module. This procedure was carried out mainly through two methods: instrument calibration and the authentication of the final product. From the calibration of the instruments, it was evidenced that the qualitative interview schedule reached a benchmark of 83.93%, the student immersion observation matrix was rated valid at 85.72%, and the learning motivation inventory was rated valid at 90.65%. With the same line of thought, the subject matter expert rubric, the interface design evaluation, and the operational usability evaluation were all ranked within the top tier of validity.

Continuing on systematically, the authentication of the product was sought to furnish a valid pedagogical tool. To ensure the validity of the design, an evaluation panel of two experts was engaged: an interface design specialist and a subject matter expert. The validation of the interface design specialist stated that the digital resource was valid with an authenticity index of 95.87%, and the critique from the subject matter expert confirmed a valid status at 85.46%. Overall, these evaluations and critiques reinforced that the newly designed flipbook-integrated e-module has passed the stringent tests of both interface design and educational content. These findings are aligned with previous studies showing that academically validated flipbook e-resources provide a great enhancement to the quality of teaching (Amiruddin & Imama, 2025).

The authentication of peer experts has potential to elevate the quality of modern education. While subject-matter validation is critical to ensure technical accuracy and relevance of educational content, validation of the interface design is essential to guarantee that the digital medium is visually appealing and provides a user-friendly ergonomic design. Notably, the interface designer made some strategic suggestions and recommendations for corrections to refine the geometry flipbook e-module, which demanded some thorough structural revisions. Although the mathematical assessment criteria classified the prototype as having high validity, these qualitative recommendations were recognized as an essential component of the product design development. The responsive structural critiques and design revisions to the geometry flipbook e-module based on the recommendations of the interface designer are presented systematically in Table 5.

Table 5. Revisions Based on Media Expert Suggestions

No.	Suggestions	Before Revision	After Revision
1	Replace the object image to better suit the age of the e-module users.	The e-module cover used an illustration that did not adequately represent the characteristics and age of the target users, making it less suitable for junior high school students.	The cover illustration was replaced with an image that better reflects the characteristics of junior high school students, making the e-module more relevant, attractive, and appropriate for the target users.

2	Revise the content of the user instructions.	The user instructions consisted of relatively long and less systematic sentences, which could potentially make it difficult for students to understand the steps for using the e-module.	The user instructions were revised using simpler, clearer, and more systematic language, making them easier for students to understand and follow.
3	Increase the font size of the material by one or two levels.	The font size used in the learning material was relatively small, causing some parts of the text to be less clear and less comfortable for students to read.	The font size of the learning material was increased by two levels, making the text clearer, more comfortable to read, and improving readability for students.

The study's outstanding authenticity metrics indicate that the design process was performed with careful precision and in accordance with the standard operating procedures of research and development. Further improvements to the design, as a result of the evaluators' qualitative feedback, increased the structural integrity of the educational tool. This finding is consistent with the opinion of the experts that in the process of developmental research, the purpose of making iterative changes to the research is to refine a prototype, and as a result, it becomes more functional, expressive, and customized to the needs of the users (Rosman & Hidayati, 2023).

D. Implementing Stage

The subsequent operational milestone involved the classroom deployment of the flipbook-infused electronic module during geometry instruction for seventh-grade pupils at a state secondary academy in Palopo City. Throughout the instructional cycle, structured observations of student immersion were systematically executed. The empirical diagnostics emanating from this student engagement efficacy trial are explicitly detailed in Table 6.

Table 6. Results of the Student Engagement Effectiveness Test

Table of Results of the Student Engagement Effectiveness Test				
No	Assessment Aspect	Total Score Obtained		N-Gain (%)
		Before Implementation	After Implementation	
	Behavioral Engagement Indicators			
1	Students attend class on time and are ready to participate in learning activities	49	85	70
2	Students actively participate in learning activities	39	86	77
3	Students focus on the teacher's explanation and learning media	42	85	74
4	Students actively use the flipbook-based e-module	40	95	92
5	Students complete the assigned exercises/tasks	60	93	82
6	Students ask or answer questions during the learning process	52	83	64
	Emotional Engagement Indicators			
7	Students show enthusiasm during the learning process	42	89	81
8	Students are interested in using the flipbook-based e-module	30	85	78
9	Students demonstrate positive expressions while learning	33	79	69
10	Students appear to enjoy the learning process	43	80	65

11	Students show confidence when completing tasks or participating in discussions	40	87	78
Cognitive Engagement Indicators				
12	Students make efforts to understand geometry concepts	38	90	84
13	Students actively seek solutions to the given problems	38	79	66
14	Students learn independently using the e-module	36	65	45
15	Students continue trying to complete tasks despite experiencing difficulties	33	70	55
16	Students use specific strategies to understand the material	40	73	55
Average		40,94	82,75	70,79
Category		Moderate	Very High	High

Analyzing the quantitative data of student immersion matrices, the baseline mean score in the absence of flipbook-infused e-modules was just 40.94%. Following the use of e-modules in instructional design, that mean jumped to 82.75%. The normalized gain (N-Gain) empirical measure positioned that particular instructional design element in its high range at 70.79%. These findings indicate that the use of the flipbook-based digital module successfully created a high degree of learner involvement at all levels of the instructional design process. A high degree of learner engagement at multiple levels indicates that the digital educational module was extremely effective in capturing learners' attention, eliciting active behavioral participation, and committing learners affectively and cognitively to the instructional design process in the area of geometry. These findings support the earlier work on the positive impact of flipbook-based e-resources on learner engagement by providing an instructional resource that is visually appealing and engaging and eliminates boredom (Nafiah et al., 2023).

The increase in student engagement in this study was not solely influenced by the use of digital media, but was integrated through various learning features in the flipbook-based e-module, such as learning videos, animations, interactive wordwall-based exercises, flexible navigation, and geometry visualizations. The integration of these features encouraged behavioral, emotional, and cognitive student engagement, as proposed by Erawati et al., (2024). These findings suggest that increased student engagement is influenced by learning designs that accommodate learning needs more interactively than conventional media.

The presentation of learning videos and animations plays a role in increasing emotional engagement through the visual and verbal delivery of information. According to the Cognitive Theory of Multimedia Learning, the combination of images and narrative facilitates the process of forming mental representations, while animation helps reduce extraneous cognitive load, making it easier for students to understand abstract geometry concepts (Mayer, 2024). Furthermore, interactive wordwall-based exercises provide immediate feedback and gamification elements that encourage active student participation, while interactive navigation allows students to revisit the material as needed. This aligns with Self-Determination Theory, which states that self-directed learning contributes to increased student engagement and motivation (Saeed & Zyngier, 2012).

The results of this study extend previous research findings, which generally only report improvements in learning outcomes or conceptual understanding through the use of flipbooks. This research demonstrates that the integration of various multimedia features within a single e-module supports the understanding of geometry concepts through more concrete spatial visualizations, but also increases overall student engagement across behavioral, emotional, and cognitive dimensions. Therefore, this study's contribution lies in demonstrating the effectiveness of flipbooks, driven by the digitization of teaching materials and the application of multimedia learning principles, which can provide a more active, meaningful, and student-centered learning experience.

After studying geometry with the flipbook-based digital module, seventh graders completed an operational feedback survey to assess the module's design and usability. The survey included multiple dimensions of design and usability including the logic of the pedagogical module and the degree of understanding the module facilitated. The survey of usability and design of the module received a score of 83.38%, thus landing it in the high range of evaluative tools. This score indicates that the module facilitated smooth instruction and understanding of the concepts of geometry. An absence of cognitive and mechanical friction in the design of the digital learning interface confirms that the interface was best utilized as intended, throughout the education timeline - even after the design and development teams had exited the project (Nurhayati et al., 2025).

E. Evaluation Stage

The efficacy testing of the newly engineered pedagogical tool represented the terminal phase within the systematic ADDIE developmental model. Within this empirical framework, to systematically appraise the instructional potency of the developed learning medium, a student scholastic drive inventory was administered to map the relative growth in the participants' academic enthusiasm prior to and following the digital intervention. The empirical diagnostics emanating from this learning motivation effectiveness trial are explicitly detailed in Table 7.

Table 7. Results of the Learning Motivation Effectiveness Test

No	Indicator	Items	Total Score Obtained		N-Gain (%)
			Pratest	Post-test	
1	Desire and willingness to achieve success	2	78	92	64
2	Learning drive and learning needs	1	77	96	83
3	Future aspirations and goals	3	82	99	94
4	Appreciation in learning	1	79	99	95
5	Engaging learning activities	1	80	100	1
6	A conducive learning environment	1	80	98	90
Total		9	476	584	
Average			79,00	97,00	87,00
Criteria			High	Very High	High

The motivation index for students expressed a 79.00% baseline metric before the integration of the flipbook-infused electronic modules, compared to the 97.00% observed after the integration. The High concern level for the motivation index was further substantiated by a positive normalized gain (N-Gain) value of 87.00%. This finding demonstrates that e-modules function as a medium for delivering material and a tool capable of building student's interest, self-confidence, and willingness to actively engage in mathematics learning. This increased motivation is supported by the integration of various multimedia features, such as text, images, instructional videos, animations, interactive wordwall-based exercises, and geometric visualizations, making learning more engaging, interactive, and understandable. According to the Cognitive Theory of Multimedia Learning, presenting information through a combination of visual and verbal representations can increase attention and facilitate knowledge construction (Mayer, 2024). Thus, students gain a more positive learning experience and are motivated to continue participating in the learning process.

The interactive design of flipbooks also contributes to increased learning motivation through a more enjoyable learning experience. Appealing visuals, easy navigation, and easy access to materials through various devices help maintain student attention, as explained in the ARCS Motivation model (Fuadi & Aulia, 2025). Furthermore, interactive Wordwall-based exercises that provide immediate feedback boost student's confidence in their abilities, while the flexibility to revisit the material independently provides a more independent learning experience. This aligns with Self-Determination

Theory, which emphasizes that independent learning is a crucial factor in enhancing students' intrinsic motivation (Saeed & Zyngier, 2012; Yengkopiong, 2025).

The findings of this study reinforce previous research showing that flipbook-based learning media can enhance learning motivation through more engaging and interactive presentations. However, this study makes a broader contribution by demonstrating that increased motivation is influenced not only by the digitalization of teaching materials but also by the integration of various multimedia features designed based on modern learning principles. Specifically, in geometry, rich visualizations help reduce the abstractness of concepts, making it easier for students to understand the relationships between shapes and fostering a positive perception of mathematics learning. Therefore, flipbook-based e-modules have a positive impact on engagement and learning motivation, which are key factors in learning success, because students who have high engagement and learning motivation will more easily overcome challenges in learning and can obtain better results (Asmorowati & Pratama, 2025; Chairunnisa et al., 2025; Ghifari & Linggowati, 2024). In addition to adding to the collection of teaching and learning resources of the institution, the design of this flipbook-based digital resource for teaching geometry may also have the potential to form a positive disposition towards mathematics among younger students.

Although this study shows that flipbook-based e-modules are effective in increasing student engagement and learning motivation, the results need to be interpreted with several limitations in mind. This study was conducted only in one public junior high school with subjects limited to seventh-grade students, so generalization of the findings to other educational levels still requires further verification. Furthermore, the effectiveness of the e-modules was only evaluated based on increases in student engagement and learning motivation using Normalized Gain (N-Gain) analysis, thus not providing a picture of the magnitude of the influence or causal relationships between variables. This study also did not examine the impact of the e-modules on other mathematical abilities relevant to geometry learning, such as spatial ability, conceptual understanding, critical thinking, and problem-solving. According to Design-Based Research, the validity of a learning innovation will be stronger if it is tested repeatedly in various contexts, student characteristics, and different implementation cycles Design-Based Research Collective (Ghosh & Kumar, 2025). Therefore, further research is recommended to implement e-modules in a more diverse population, using more comprehensive analyses such as effect size, ANCOVA, or Structural Equation Modeling (SEM), and integrating technologies such as Augmented Reality (AR), Artificial Intelligence (AI), and learning analytics to evaluate their effects on students' engagement, motivation, and various mathematical abilities more comprehensively.

The final phase of the development of this product entailed the successive modification of the flip-book style electronic module in accordance with the qualitative comments and recommendations provided by the expert evaluation panel during the validation process, and the feedback obtained from end-users during the live teaching process. This final phase resulted in the production of a complete geometry learning e-module for seventh-grade students, which is accessible by scanning the QR code or by clicking on the link provided:



Fig. 2. E-Module QR Code

<https://heyzine.com/flip-book/5a8ab63247.html>

(E-module presented in Indonesian.)

The revised version of the flipbook-based e-module based on the improvement suggestions is presented in the figure below

Fig. 3. Cover Animation Display



Fig. 4. Display of Instructions

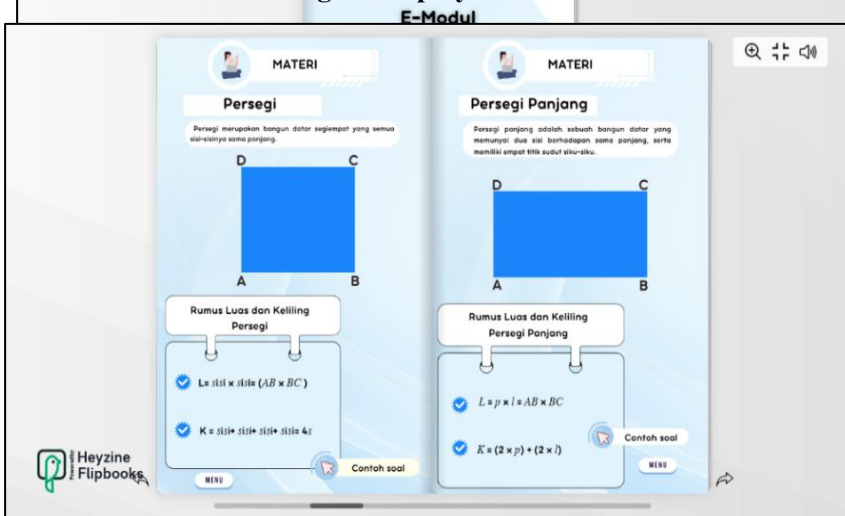


Fig. 5. Font Size Display

4. CONCLUSION

This study utilized the ADDIE instructional design framework and is the first to develop a flipbook-based electronic module for lower-secondary geometry that integrates flipbook technology as an example of modular design. The use of the flipbook-based module is warranted within the secondary mathematics curriculum due to its positive user experience and functionality, as well as its empirically substantiated potential for increasing student focus and engagement, as evidenced by the student-centered approach of the module. The findings suggest that the integration of technology within a modular design framework provides a means of promoting student engagement in the learning process and encourages students to take more responsibility for their learning.

The use of flipbook technology in the design of the electronic module positively influenced participants' perceptions of mathematics as a motivating factor in their learning. The findings indicate that a flipbook-based electronic module can be used as a motivator for secondary students to learn difficult concepts within geometry. The findings support the premise that secondary mathematics teachers may utilize a flipbook-based electronic module to promote a self-directed and engaged learning

environment. The module may also be used to address unmotivated learning behaviors within the geometry classroom.

Future research should design electronic modules that use flipbooks and are focused on different branches of mathematics. These will incorporate interactive elements that include augmented reality and sophisticated pedagogical game designs. In addition, future research may examine the potential of these digital tools for teaching children of different ages and/or conduct more in-depth empirical investigations of their possible effects on students' ability to think critically, creatively solve problems, and their overall success in school, that is, their academic performance.

REFERENCES

- Adiastuty, N., Sumarni, S., Nurhayati, N., & Rukoyah, S. S. (2025). Development of an interactive flipbook-based e-module to enhance students' ability to understand mathematical concepts in arithmetic sequences and series. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 14(4), 1733–1744. <https://doi.org/10.24127/ajpm.v14i4.11429>
- Amiruddin, A., & Imama, K. (2025). E-flipbook innovation through problem-based learning: Enhancing mathematical concept mastery. *Interdisciplinary Journal of Pedagogy and Research in Media Technology*, 1(1), 37–46. <https://doi.org/10.64268/inspire.v1i1.40>
- Anggitarachmawati. (2025). Innovative Digital Learning Media to Enhance Student Engagement and Learning Outcomes in Elementary Schools. *Schola*, 3(1), 22–27. <https://doi.org/10.26877/schola.v3i1.2328>
- Artal-Sevil, J. S., Castel, A. F. G., & Gracia, M. S. V. (2020). Flipped teaching and interactive tools. A multidisciplinary innovation experience in higher education. *International Conference on Higher Education Advances, 2020-June*, 103–112. <https://doi.org/10.4995/HEAd20.2020.10990>
- Asmorowati, D. A., & Pratama, F. W. (2025). Pengembangan Flipbook pada Materi Bentuk Aljabar untuk Meningkatkan Motivasi dan Pemahaman Konsep. *Math Didactic: Jurnal Pendidikan Matematika*, 11(2), 187–198. <https://doi.org/10.33654/math.v11i2.161>
- Astuti, I. W., & Andriani, A. (2025). Analysis of Students' Learning Motivation Using Flipbook Media in Elementary School. *Proceedings Series on Social Sciences & Humanities*, 25, 175–177. <https://doi.org/10.30595/pssh.v25i.1685>
- Bayounes, W., Saâdi, I. B., & Kinshuk. (2022). Adaptive learning: toward an intentional model for learning process guidance based on learner's motivation. *Smart Learning Environments*, 9(1), 33. <https://doi.org/10.1186/s40561-022-00215-9>
- Bryant, P. (2023). Student experience and digital storytelling: Integrating the authentic interaction of students work, life, play and learning into the co-design of university teaching practices. *Education and Information Technologies*, 28(11), 14051–14069. <https://doi.org/10.1007/s10639-022-11566-8>
- Carroll, F., & Kop, R. (2016). Colouring the gaps in learning design: Aesthetics and the visual in learning. *International Journal of Distance Education Technologies*, 14(1), 92–103. <https://doi.org/10.4018/IJDET.2016010106>
- Cavanagh, T. M., & Kiersch, C. (2023). Using commonly-available technologies to create online multimedia lessons through the application of the Cognitive Theory of Multimedia Learning. *Educational Technology Research and Development*, 71(3), 1033–1053. <https://doi.org/10.1007/s11423-022-10181-1>
- Cevikbas, M., & Kaiser, G. (2022). Student Engagement in a Flipped Secondary Mathematics Classroom. *International Journal of Science and Mathematics Education*, 20(7), 1455–1480. <https://doi.org/10.1007/s10763-021-10213-x>
- Chairunnisa, A., Usodo, B., & Yamtinah, Y. (2025). Media Flipbook sebagai Inovasi Digital untuk Pembelajaran Matematika di Madrasah Ibtidaiyah: Studi Kuasi-Eksperimen. *Didaktika: Jurnal Kependidikan*, 14(3), 5219–5230. <https://doi.org/10.58230/27454312.2863>
- Cohen, R., Katz, I., Aelterman, N., & Vansteenkiste, M. (2023). Understanding shifts in students' academic motivation across a school year: the role of teachers' motivating styles and need-based experiences. *European Journal of Psychology of Education*, 38(3), 963–988. <https://doi.org/10.1007/s10212-022-00635-8>
- Elfina, E., Waskito, W., Darni, R., Maksum, H., & Novaliendry, D. (2024). Revolutionary Flipbook-Based Digital Comic: Changing Student Engagement in Learning. *JTP - Jurnal Teknologi Pendidikan*, 25(3), 514–527. <https://doi.org/10.21009/jtp.v25i3.44809>
- Fuadi, A. P., & Aulia, D. (2025). The Influence of Interactive Media Based on Information Technology on the Learning Effectiveness of Students. *Jurnal Pendidikan Ekonomi Indonesia*, 7(1), 74–90. <https://doi.org/10.17509/jpei.v7i1.78136>
- Ghifari, M. Z., & Linggowati, T. (2024). Application-Based Flipbook Increases Student Learning Motivation. *Indonesian Journal of Education Methods Development*, 19(3). <https://doi.org/10.21070/ijemd.v19i3.846>

- Ghosh, O., & Kumar, B. (2025). *An Introduction to Design-Based Research* (pp. 1–30). <https://doi.org/10.4018/979-8-3373-2086-1.ch001>
- Hidayah, H., Riyadi, R., & Matsuri, M. (2025). The Role of Spatial Visual Intelligence in the Problem Solving Process of Building Spatial Geometry. *Social, Humanities, and Educational Studies (SHES): Conference Series*, 8(4), 13. <https://doi.org/10.20961/shes.v8i4.109327>
- Khan, H., Gul, R., & Zeb, M. (2023). The Effect of Students' Cognitive and Emotional Engagement on Students' Academic Success and Academic Productivity. *Journal of Social Sciences Review*, 3(1), 322–334. <https://doi.org/10.54183/jssr.v3i1.141>
- Kim, S., Yoon, M., Whang, S. M., Tversky, B., & Morrison, J. B. (2007). The effect of animation on comprehension and interest. *Journal of Computer Assisted Learning*, 23(3), 260–270. <https://doi.org/10.1111/j.1365-2729.2006.00219.x>
- Kurnianto, D., Sudiansyan, Heriyanto, Yani T., A., & T., H. (2022). Development of Mathematics E-Modules Through the Professional Flip PDF Application Assisted React Strategy to Improve Problem Solving Ability of Vocational Middle School Students Concentration of Accounting Expertise. *International Journal of Science and Society*, 4(3), 499–512. <https://doi.org/10.54783/ijssoc.v4i3.544>
- Manggala, M. A., Nyanasuryanadi, P., & Suherman, H. (2024). Innovative Learning Using E-Modules. *JETISH: Journal of Education Technology Information Social Sciences and Health*, 3(1), 550–557. <https://doi.org/10.57235/jetish.v3i1.1983>
- Martinjak, N. Z. (2024). Student Engagement in Distance Learning. *Didactica Slovenica - Pedagoska Obzorja*, 39(1), 83–97. <https://doi.org/10.55707/ds-po.v39i1.124>
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8. <https://doi.org/10.1007/s10648-023-09842-1>
- Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6
- Mesra, E. R., Hakim, R., Darmansyah, & Zuwirna. (2025). Development of Interactive E-Modules Using Canva and Heyzine Flipbook to Improve the Learning of Grade VII Junior High School Students. *Jurnal Penelitian Pendidikan IPA*, 11(9), 370–377. <https://doi.org/10.29303/jppipa.v11i9.12364>
- Mizani, H., Cahyadi, A., Hendryadi, H., Salamah, S., & Retno Sari, S. (2022). Loneliness, student engagement, and academic achievement during emergency remote teaching during COVID-19: the role of the God locus of control. *Humanities and Social Sciences Communications*, 9(1). <https://doi.org/10.1057/s41599-022-01328-9>
- Mucti, A., & Nurmala R. (2025). Development of Computer-Based Role-Playing Game Mathematics Learning Media: Spatial Geometry Material in Junior High School. *ETDC: Indonesian Journal of Research and Educational Review*, 4(4), 1618–1633. <https://doi.org/10.51574/ijrer.v4i4.3956>
- Mulyana, E. H., Lidinillah, D. A. M., & Latifah, F. (2023). Gebox Media to Facilitate the Cognitive Development of Children Aged 5-6 Years on Geometry Concepts. *AURELIA: Jurnal Penelitian Dan Pengabdian Masyarakat Indonesia*, 2(1), 227–233. <https://doi.org/10.57235/aurelia.v2i1.298>
- Nafiah, N., Syamsul Ghufro, Sri Hartatik, Pance Mariati, & Afib Rulyansah. (2023). The effect of flipbook-based digital books on elementary school students' interest in learning. *World Journal on Educational Technology: Current Issues*, 15(4), 342–354. <https://doi.org/10.18844/wjet.v15i4.7833>
- Nonik Erawati, K., Dita Ardiani, N. N., & Santiago, G. A. S. (2024). E-Module Interaktif Berbasis Flipbook Pada Matakuliah Machine Learning Untuk Meningkatkan Kreatifitas Mahasiswa. *Jurnal Penjaminan Mutu*, 10(01), 45–51. <https://doi.org/10.25078/jpm.v10i01.3603>
- Nurhayati, N., Nursyafitri, E. A., & Adiastuty, N. (2025). Development of a Desmos-Assisted Flipbook E-Module to Improve Mathematical Problem-Solving Ability. *Kognitif: Jurnal Riset HOTS Pendidikan Matematika*, 5(4). <https://doi.org/10.51574/kognitif.v5i4.3546>
- Putri, D. W. E., Syabiya, N. F., Sa'diyah, Z., Kunaifi, A., Mufidah, Z. R., & Istiq'faroh, N. (2025). FLITARA: Culture-Based Interactive Flipbook Innovation to Improve Elementary Schools' Literacy in the Era of Society 5.0. *Profesi Pendidikan Dasar*, 12(2), 199–220. <https://doi.org/10.23917/ppd.v12i2.10465>
- Reza, B. R. A., Efrizon, E., Dewi, I. P., & Anori, S. (2025). Perancangan dan Pengembangan Media Pembelajaran Interaktif Berbasis FlipBook pada Mata Pelajaran Teknik Elektronika Industri untuk Materi Elemen Sistem Kontrol Industri. *TSAQOFAH*, 5(6), 6349–6360. <https://doi.org/10.58578/tsaqofah.v5i6.7573>
- Rosman, Z., & Hidayati, A. (2023). Development Of Learning Multimedia In Mathematics Subjects For Grade III Elementary School Students. *KOLOKIU Jurnal Pendidikan Luar Sekolah*, 11(1), 14–21. <https://doi.org/10.24036/kolokium.v11i1.563>
- Saeed, S., & Zyngier, D. (2012). How Motivation Influences Student Engagement: A Qualitative Case Study. *Journal of Education and Learning*, 1(2). <https://doi.org/10.5539/jel.v1n2p252>
- Santi, E. A., Gorghiu, G., & Pribeanu, C. (2022). Students' Engagement and Active Participation During the Pandemic. *Informatica Economica*, 26(1/2022), 5–15. <https://doi.org/10.24818/issn14531305/26.1.2022.01>

- Sattar, T., Bashir, F., & Arshad Rana, F. (2022). Underlying Causes of Student Academic Engagement in High Schools of South Punjab, Pakistan. *IUB Journal of Social Sciences*, 4(2), 111–124. <https://doi.org/10.52461/ijoss.v4i2.1433>
- Supratman, & Tundreng, S. (2025). Utilizing Visual Media to Improve Conceptual Understanding of Geometry Among Junior High School Students. *Journal of Research in Science and Mathematics Education (J-RSME)*, 4(2), 95–107. <https://doi.org/10.56855/jrsme.v4i2.1570>
- Suwanto, F. R., Wijaya, A., Aprisal, Fauzi, A., & Tobondo, Y. V. (2023). Realistic Mathematics Textbook Toward The Students' Mathematical Generalization Skills. *Mosharafa: Jurnal Pendidikan Matematika*, 12(2), 291–300. <https://doi.org/10.31980/mosharafa.v12i2.784>
- Tani, M., Manuguerra, M., & Khan, S. (2022). Can videos affect learning outcomes? Evidence from an actual learning environment. *Educational Technology Research and Development*, 70(5), 1675–1693. <https://doi.org/10.1007/s11423-022-10147-3>
- Torres, K. M., & Statti, A. (2023). What Can Data Tell Us? *International Journal of Curriculum Development and Learning Measurement*, 4(1), 1–13. <https://doi.org/10.4018/IJCDLM.320219>
- Unayah, H., Rasyid, M., Suyanta, S., Wilujeng, I., & Rahmawati, L. (2024). A Recent Study on Flipbook as Media Implementation in Science Education in Digital Age: A Systematic Literature Review. *Unnes Science Education Journal*, 13(3), 248–258. <https://doi.org/10.15294/usej.v13i3.17200>
- Wang, C., Cho, H. J., Wiles, B., Moss, J. D., Bonem, E. M., Li, Q., Lu, Y., & Levesque-Bristol, C. (2022). Competence and autonomous motivation as motivational predictors of college students' mathematics achievement: from the perspective of self-determination theory. *International Journal of STEM Education*, 9(1). <https://doi.org/10.1186/s40594-022-00359-7>
- Wang, M.-T., & Fredricks, J. A. (2014). The Reciprocal Links Between School Engagement, Youth Problem Behaviors, and School Dropout During Adolescence. *Child Development*, 85(2), 722–737. <https://doi.org/10.1111/cdev.12138>
- Yengkopiong, J. P. (2025). The Way Forward for Secondary School Students: The Role of Self-Determination Theory and Intrinsic Motivation. *East African Journal of Education Studies*, 8(1), 288–299. <https://doi.org/10.37284/eajes.8.1.2643>
- Žakelj, A., & Klančar, A. (2022). The Role of Visual Representations in Geometry Learning. *European Journal of Educational Research*, 11(3), 1393–1411. <https://doi.org/10.12973/eu-jer.11.3.1393>

Spiritual Economic Dynamics: A Nonlinear ODE Model of Giving, Faith, and Wealth Growth with Threshold Transition from Give to Take

Sutanto¹, Deni Setiawan²

^{1,2}Department of Mathematics, Universitas Sebelas Maret, Surakarta, Indonesia

¹sutanto71@staff.uns.ac.id, ²denys3009@student.uns.ac.id

ABSTRACT

Giving is commonly viewed in conventional economics as a reduction of personal resources, whereas spiritual perspectives associate it with long-term prosperity and broader social value. This study develops a nonlinear ordinary differential equation (ODE) model to examine the dynamic interaction among giving, faith, and wealth and the conditions under which giving contributes to sustained wealth accumulation. The model is analyzed through equilibrium analysis, Jacobian-based local stability analysis, and fourth-order Runge–Kutta numerical simulation. The results show that the proposed system admits a non-trivial positive equilibrium and satisfies the local stability condition $\gamma\eta > \beta\delta$, ensuring that the equilibrium is locally asymptotically stable. The model also identifies the threshold condition $\alpha SI > \mu W$, which governs the transition from the Give regime to the Take regime when the productive effects of giving exceed economic friction. Numerical simulations further demonstrate that faith reinforcement and economic friction strongly influence long-term system trajectories. These findings suggest that giving is not only a moral practice but also a productive mechanism that can support sustainable wealth growth under appropriate structural and behavioral conditions.

Keyword: Giving Behavior, Faith, Wealth Growth, Ordinary Differential Equations, Nonlinear Dynamics



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Deni Setiawan,

Department of Mathematics,

Universitas Sebelas Maret,

Jalan Ir. Sutami No. 36A, Kentingan, Jebres, Kota Surakarta, Jawa Tengah, Indonesia.

denys3009@student.uns.ac.id

1. INTRODUCTION

The principle of giving before receiving appears across a wide range of human activities, spanning economic behavior, social interaction, and spiritual practice. In many real-world settings, individuals and institutions allocate time, effort, capital, and other resources without immediate certainty of return. Firms invest in innovation long before profits are realized, while individuals often devote resources to education, social relationships, or charitable action with the expectation that long-term benefits will eventually emerge. These patterns suggest that giving should not always be reduced to a simple loss of wealth in the present moment. Under certain conditions, it may instead operate as part of a longer process through which economic and social returns are gradually formed (Rufaedah & Patih, 2025; Feinman et al., 2025).

Within conventional economics, however, giving is generally interpreted as a reduction in individual utility or wealth because resources are transferred to other parties (Assari & Zare, 2024; Majer et al., 2022). Andreoni's warm-glow model refined this view by showing that individuals may derive intrinsic satisfaction from giving, yet the act itself is still treated as a cost borne by the giver (Andreoni, 1990). Related perspectives, including fairness and inequity aversion, likewise frame giving primarily as a mechanism for restoring social balance or reducing inequality rather than as a direct source of wealth formation (Sen et al., 2026; Sohn et al., 2024). As a result, standard economic formulations have been more successful in explaining why individuals give than in clarifying how giving might contribute to long-term wealth growth.

A broader interpretation is offered by faith-based and spiritual economics. In this perspective, giving is not necessarily viewed as a pure depletion of resources. Rather, it is embedded in a wider structure of trust, moral obligation, solidarity, and collective welfare. In Islamic economic thought, for instance, zakat and sadaqah are understood not only as redistributive instruments but also as practices that may strengthen economic resilience and social cohesion (Azizah et al., 2025; Jaenudin and Hamdan, 2022; Lestari, 2025; Muqhniiy et al., 2026). Empirical studies have shown that such forms of giving can contribute to poverty reduction and improvements in household welfare, while broader charitable

practices have been associated with stronger cooperation, higher social trust, and the expansion of productive opportunities within communities (Beik and Arsyianti, 2016; Sugianto and Mohammad, 2024; Afrimaigus and Mohammed, 2026). This distinction matters because it opens the possibility that giving may function not only as a transfer of wealth, but also as part of a mechanism through which wealth is reproduced and sustained over time.

The conceptual foundation of this study is supported by three complementary theoretical perspectives. First, Behavioral Economics explains that individuals do not always make economic decisions solely to maximize immediate utility. Instead, giving behavior is also driven by intrinsic motivation, altruism, fairness, reciprocity, and moral preferences (Andreoni, 1990; Fehr & Schmidt, 1999). Second, Social Exchange Theory argues that social interactions involve reciprocal relationships in which present sacrifices may generate future benefits through trust, cooperation, and social capital (Blau, 1964; Cropanzano & Mitchell, 2005). From this perspective, giving should not be interpreted merely as a transfer of resources but also as an investment capable of producing long-term economic and social returns. Third, Religious Commitment Theory suggests that faith shapes individual attitudes and behaviors by strengthening moral commitment, consistency, and willingness to engage in prosocial actions despite the absence of immediate material rewards (Worthington et al., 2003).

Research on this issue has developed across several strands, although these strands do not always speak to one another. Behavioral economics has examined the motivational foundations of giving, emphasizing altruism, reciprocity, prestige, empathy, and social norms. Bekkers and Wiepking (2010), for example, show that awareness of need, reputational concerns, and perceived effectiveness shape charitable behavior. More recent studies also suggest that inequality and broader economic conditions affect the willingness to give (Sohn et al., 2024). Yet this literature remains centered on the decision to give itself. It says much less about what happens after giving becomes part of a repeated economic process or how it may influence wealth accumulation over time.

A different line of inquiry comes from wealth dynamics and econophysics, where the focus shifts from motivation to accumulation. Here, wealth is analyzed as a variable that evolves through stochastic processes, heterogeneous returns, and nonlinear interactions among agents. Benhabib et al. (2011) demonstrate that inequality may emerge endogenously through differences in capital returns and accumulation mechanisms, while econophysics models describe wealth and income distributions as outcomes of complex interactions within the system (Dragulescu and Yakovenko, 2009). From a modeling perspective, these studies offer powerful tools for understanding how wealth changes over time. Surprisingly, however, giving is rarely incorporated as an endogenous mechanism capable of shaping those dynamics.

Another perspective comes from Islamic social finance, which regards zakat, almsgiving, and waqf as important instruments for improving social welfare. Beik and Arsyianti (2016) showed that zakat helps reduce poverty while improving household welfare. These findings provide empirical support for the social and economic benefits of giving. Despite that, much of the existing literature focuses on measuring observed outcomes rather than explaining how wealth evolves through repeated acts of giving. A dynamic mathematical framework that captures this long-term process is still largely absent.

In the context of this study, faith is operationally defined as the degree of spiritual commitment that consistently encourages giving behavior despite the absence of immediate material rewards. Meanwhile, the Take phase refers to the stage at which the productive effects of giving exceed economic friction, resulting in sustainable wealth accumulation rather than merely receiving resources from others.

The gap becomes clearer when viewed through the lens of nonlinear systems. Although previous studies have examined giving behavior, wealth dynamics, and spiritual economics from different perspectives, there is still no formal mathematical model that integrates giving, faith, and wealth within a unified nonlinear dynamical systems framework. Existing studies explain the motivations for giving or the mechanisms of wealth accumulation separately, but none explicitly models the threshold transition from the Give phase to the Take phase. This research addresses that gap by proposing a nonlinear ordinary differential equation model that integrates these three variables into a single dynamic framework.

Based on these problems, this study proposes a nonlinear spiritual-economic dynamics model based on ordinary differential equations (ODEs) that integrates wealth, giving, and faith as endogenous

variables. The model is developed using a physics-inspired framework approach, in which economic growth is viewed as the result of interaction between productive force and friction. This study aims to develop a nonlinear mathematical model capable of explaining the dynamic relationship between giving, faith, and wealth growth, as well as identifying the transition mechanism from the Give phase toward the Take phase within the framework of quantitative spiritual economics.

2. RESEARCH METHOD

This study employs a theoretical quantitative approach using a nonlinear dynamical system framework to model the dynamic relationship between giving, faith, and wealth growth. The theoretical approach was chosen because the study focuses on developing a conceptual mathematical model that explains how giving may transform from a short-term sacrifice into a mechanism of long-term wealth growth. The model is developed using a physics-inspired framework approach, namely an approach that adopts the analogy of force and friction in explaining the behavior of dynamic economic systems (Hakim et al., 2025). Similar approaches have been widely used in studies of econophysics and nonlinear economics to explain wealth distribution, economic growth, and regime changes in complex systems (Dragulescu & Yakovenko, 2009; Benhabib et al., 2011).

The data sources in this study are theoretical in nature and derived from relevant literature studies regarding behavioral economics, wealth dynamics, Islamic social finance, econophysics, and nonlinear dynamics. Data collection techniques were conducted through library research by reviewing journal articles, scientific books, and previous studies related to giving behavior, wealth distribution, nonlinear transition theory, and spiritual economics. Literature regarding warm-glow giving was used to understand giving behavior within conventional economics (Andreoni, 1990), while theories of fairness and inequity aversion were used to explain resource transfers within a social context (Fehr & Schmidt, 1999; Hughes et al., 2018). In addition, studies regarding wealth distribution and econophysics were used as the basis for constructing the nonlinear model (Dragulescu & Yakovenko, 2009; Dolfín & Lachowicz, 2014). Critical transition and nonlinear dynamics approaches were also employed to explain the possibility of regime transitions from Give to Take (Scheffer, 2009; Mascareño, 2022).

The research model was constructed using a nonlinear system of ordinary differential equations (ODEs) consisting of three endogenous variables, namely wealth $W(t)$, giving $S(t)$, and faith $I(t)$. The relationship among these three variables is represented through the following system of equations:

$$\frac{dW}{dt} = \alpha SI - \mu W \quad (1)$$

$$\frac{dS}{dt} = \beta I - \gamma S \quad (2)$$

$$\frac{dI}{dt} = \delta S - \eta I \quad (3)$$

where α represents the effectiveness of giving toward wealth growth, μ represents economic friction, β represents the influence of faith on giving, γ represents the rate of giving decline, δ represents the reinforcement of faith resulting from giving activities, and η represents the rate of faith decline. The system forms a feedback loop:

$$S \rightarrow I \rightarrow W$$

which indicates that giving strengthens faith, faith increases the effectiveness of giving, and the interaction between both variables drives sustainable wealth growth. The model also allows the emergence of a threshold transition, namely a condition in which the productive force generated by the interaction between giving and faith successfully surpasses economic friction, causing the system to move from the Give regime toward the Take regime.

The first stage was determining the equilibrium point by setting all time derivatives equal to zero. The second stage involved local stability analysis using the Jacobian matrix and eigenvalues to determine whether the equilibrium point is stable, unstable, or undergoes a regime transition. The

Jacobian–eigenvalue approach is a standard method in nonlinear system analysis used to evaluate the local behavior of dynamic systems (Escot & Sandubete, 2023). Numerical experiments were then performed using the fourth-order Runge–Kutta algorithm to observe system behavior under different conditions.

The RK4 method was selected because it has a high level of accuracy in solving nonlinear ordinary differential equation systems (Agustin et al., 2024; Saragih et al., 2024). The simulation outcomes were interpreted from the perspective of quantitative spiritual economics, with particular attention to how moral behavior and faith-related factors may influence long-term wealth dynamics.

3. RESULTS AND DISCUSSION

A. Characteristics of the Equilibrium Point

The equilibrium point of the system is obtained by setting all time derivatives equal to zero:

$$\frac{dW}{dt} = 0, \frac{dS}{dt} = 0, \frac{dI}{dt} = 0 \quad (4)$$

Based on the proposed system, the following equations are obtained:

$$\alpha SI - \mu W = 0 \quad (5)$$

$$\beta I - \gamma S = 0 \quad (6)$$

$$\delta S - \eta I = 0 \quad (7)$$

The second equation gives a direct relationship between the equilibrium values of giving and faith:

$$S^* = \frac{\beta}{\gamma} I^* \quad (8)$$

Meanwhile, from the third equation:

$$I^* = \frac{\delta}{\eta} S^* \quad (9)$$

After replacing the variables in the first equation, the equilibrium expression for wealth is obtained:

$$W^* = \frac{\alpha S^* I^*}{\mu} \quad (10)$$

This result indicates that long-term wealth depends on the interaction between giving and faith, while economic friction reduces the equilibrium level. The finding is consistent with Behavioral Economics, which argues that prosocial behavior is not solely motivated by immediate utility maximization but also by intrinsic motivation and reciprocal preferences (Andreoni, 1990; Fehr & Schmidt, 1999). It also supports the perspective of Social Exchange Theory, which suggests that present sacrifices may generate future economic benefits through the accumulation of trust, cooperation, and social capital (Cropanzano & Mitchell, 2005). Accordingly, increasing levels of giving and faith strengthen the productive interaction within the model, thereby creating more favorable conditions for sustainable wealth growth.

From a conceptual perspective, the positive equilibrium suggests that wealth, giving, and faith can develop simultaneously over time. This result is consistent with studies in Islamic social finance, which emphasize that giving may generate long-term socioeconomic benefits through strengthened solidarity, social trust, and community welfare rather than merely reducing individual resources (Beik & Arsyianti, 2016; Sugianto & Mohammad, 2024). The equilibrium obtained in this study provides a mathematical representation of this theoretical argument by demonstrating that giving may become a productive mechanism supporting sustainable wealth accumulation under appropriate parameter conditions.

B. System Stability Analysis

Local stability is examined using the Jacobian matrix shown below:

$$J(W, S, I) = \begin{bmatrix} -\mu & \alpha I & \alpha S \\ 0 & -\gamma & \beta \\ 0 & \delta & -\eta \end{bmatrix} \quad (11)$$

The characteristic determinant of the system is obtained from:

$$\det(J - \lambda I) = 0 \quad (12)$$

which produces the characteristic equation:

$$(-\mu - \lambda)[(-\gamma - \lambda)(-\eta - \lambda) - \beta\delta] = 0 \quad (13)$$

From the equation, one eigenvalue is obtained as:

$$\lambda_1 = -\mu \quad (14)$$

The remaining two eigenvalues are determined by solving the quadratic equation below:

$$\lambda^2 + (\gamma + \eta)\lambda + (\gamma\eta - \beta\delta) = 0 \quad (15)$$

Local stability requires every eigenvalue to have a negative real part. If this condition is satisfied, the trajectories gradually approach the equilibrium point.

Accordingly, stability is guaranteed under the following:

$$\gamma\eta > \beta\delta$$

The inequality highlights that stability depends on the balance between the reinforcing effect of faith and the mechanisms that reduce system intensity. Neither process should dominate excessively if a stable equilibrium is expected.

If the interaction between giving and faith becomes too strong, the system may become unstable. When the parameters remain well balanced, the trajectories gradually move toward the equilibrium state. As time progresses, the variables approach stable values instead of diverging.

In the economic context, the friction parameter (μ) represents obstacles such as inflation, economic uncertainty, inefficiency, and social pressure. The greater the friction value, the greater the barriers to wealth growth.

C. Threshold Transition from Give to Take

An interesting feature of the proposed model is the presence of a threshold that separates two different system regimes, referred to here as the Give regime and the Take regime. This transition is represented by the condition:

$$\alpha SI > \mu W$$

Within the proposed model, the inequality compares the productive influence generated by the interaction between giving and faith with the economic friction that constrains wealth growth. This interpretation is consistent with Social Exchange Theory, which suggests that present sacrifices may generate future reciprocal benefits through the accumulation of trust, cooperation, and social capital (Cropanzano & Mitchell, 2005). Likewise, Behavioral Economics recognizes that giving behavior is often motivated not only by immediate utility but also by intrinsic motivation and long-term expectations of reciprocal outcomes (Andreoni, 1990; Fehr & Schmidt, 1999). Therefore, the Give

regime represents a stage in which the productive effects of giving have not yet become sufficiently strong to overcome existing economic constraints.

As the levels of giving and faith increase, the productive contribution represented by the term αSI also becomes larger. When this contribution exceeds the economic friction represented by μW , the model predicts a transition toward a more favorable growth regime. This threshold phenomenon is consistent with Critical Transition Theory, which explains that nonlinear systems may experience abrupt regime shifts once a critical threshold has been surpassed (Scheffer, 2009). Similar nonlinear transitions have also been discussed in nonlinear dynamical systems, where feedback mechanisms determine whether a system remains in its current state or shifts to a new equilibrium (Strogatz, 2018).

This threshold mechanism provides a mathematical explanation for why the economic benefits of giving are not always immediately observable. Previous studies in Islamic social finance have demonstrated that charitable activities such as zakat and sadaqah contribute to poverty reduction and household welfare, although these benefits generally emerge over time rather than instantaneously (Beik & Arsyianti, 2016). Similarly, studies by Sugianto and Mohammad (2024) suggest that productive charitable practices strengthen long-term economic resilience through social cooperation and resource redistribution. The proposed nonlinear model complements these empirical findings by providing a mathematical mechanism that explains how the cumulative interaction between giving and faith may eventually generate sustainable wealth growth after a critical threshold has been reached.

From the perspective of nonlinear dynamics, the identified threshold highlights the importance of feedback interactions among giving, faith, and wealth. Once the critical threshold has been exceeded, positive feedback strengthens the interaction between these variables, allowing the system to converge toward a more sustainable growth trajectory. Although the proposed model remains theoretical, it extends previous studies by integrating concepts from Behavioral Economics, Social Exchange Theory, and nonlinear dynamical systems into a unified mathematical framework capable of describing the transition from the Give regime to the Take regime.

D. Interpretation of Numerical Simulation

Numerical simulations using the fourth-order Runge–Kutta method were carried out to examine how the proposed system evolves over time under different parameter settings. The simulation results indicate that the trajectories of wealth, giving, and faith are sensitive to the balance between productive effects and economic friction.

When faith-related parameters are relatively weak and economic friction is comparatively high, wealth growth tends to be limited. Under these conditions, $\alpha SI < \mu W$, indicating that the productive contribution generated by the interaction between giving and faith is insufficient to compensate for economic friction. This result is consistent with Behavioral Economics, which argues that prosocial behavior alone does not necessarily generate favorable economic outcomes unless supported by conducive behavioral and institutional conditions (Andreoni, 1990; Fehr & Schmidt, 1999). It also agrees with empirical findings in Islamic social finance showing that the effectiveness of charitable activities depends on supporting socioeconomic conditions rather than on the act of giving alone (Beik & Arsyianti, 2016).

A different pattern emerges when the reinforcing effects of faith and giving become stronger while economic friction remains relatively low. Under these conditions, $\alpha SI > \mu W$, allowing the productive component of the system to dominate. The resulting trajectories converge toward a positive equilibrium, indicating that wealth, giving, and faith can coexist in a stable long-run state. This finding supports Religious Commitment Theory, which emphasizes that strong spiritual commitment reinforces prosocial behavior and encourages persistence in morally motivated economic actions despite delayed material rewards (Worthington et al., 2003). It is also consistent with previous studies suggesting that charitable activities reinforced by trust and spiritual values contribute to sustainable socioeconomic welfare (Sugianto & Mohammad, 2024; Afrimaigus & Mohammed, 2026).

The simulations also demonstrate the nonlinear sensitivity of the proposed model to parameter changes. Small variations in the effectiveness of giving, faith reinforcement, or economic friction produce substantially different long-term trajectories. This behavior is consistent with nonlinear dynamical systems theory, which states that complex systems may undergo significant qualitative changes following relatively small parameter perturbations due to feedback interactions and threshold

effects (Strogatz, 2018; Scheffer, 2009). Therefore, the proposed model exhibits the characteristic nonlinear behavior commonly observed in many complex adaptive systems.

Taken together, the simulation results suggest that giving may contribute to long-term wealth development when supported by sufficient spiritual reinforcement and when economic friction remains within manageable levels. These findings complement previous empirical studies in Islamic social finance, which report that charitable practices contribute to poverty reduction, social welfare, and economic resilience (Beik & Arsyianti, 2016; Sugianto & Mohammad, 2024). Unlike previous empirical studies, however, the present research provides a nonlinear mathematical framework capable of explaining the dynamic mechanism through which the interaction among giving, faith, and economic conditions generates sustainable wealth growth. Consequently, this study extends the existing literature by integrating behavioral, spiritual, and nonlinear dynamical perspectives into a single quantitative model.

4. CONCLUSION

This study developed a nonlinear system of ordinary differential equations to examine how giving, faith, and wealth may evolve within a single dynamic framework. Rather than treating these elements separately, the model places them in continuous interaction. In that setting, economic and spiritual factors are allowed to influence one another over time, not only through direct effects but also through feedback processes within the system.

Several findings emerge from the analysis. First, the system admits a positive equilibrium under certain parameter conditions, which indicates that wealth, giving, and faith can coexist in a stable configuration. Second, the model identifies a threshold that separates two different regimes of system behavior. When the productive contribution generated by the interaction of giving and faith exceeds the level of economic friction, the trajectory of the system shifts toward a more favorable pattern of growth. When this reinforcing effect weakens, however, the system is more likely to remain in a less productive state.

The numerical simulations lead to a similar reading. Variations in faith reinforcement, giving effectiveness, and economic friction produce markedly different long-run outcomes. Some parameter combinations support convergence toward a positive equilibrium, whereas others limit the system's growth potential. These results do not suggest that giving mechanically generates prosperity. That would be too strong a claim. What they do suggest is that giving may operate as more than a simple reduction of resources when it is embedded in a broader structure of faith and economic interaction.

This point matters because the conventional treatment of giving often stops at the idea of transfer and sacrifice. The present model offers a different possibility. Under the assumptions adopted here, giving can also be understood as part of a dynamic process that shapes future wealth through cumulative feedback involving belief, behavior, and economic activity. The relationship, therefore, is not linear, and it should not be reduced to a single-direction loss mechanism.

At the same time, the study remains theoretical and rests on a number of simplifying assumptions. The model has not yet been tested against empirical data, and many socioeconomic influences that matter in real settings are still absent from the framework. For that reason, the results should be read as a conceptual and mathematical exploration rather than as a direct empirical description of actual giving behavior.

Further work is therefore needed. Empirical calibration would be an obvious next step, but it is not the only one. The framework could also be extended by incorporating additional socioeconomic variables, alternative behavioral responses, or more complex forms of interaction between spiritual and economic factors. Such developments would make it possible to assess whether the dynamics proposed in this study can account for observable patterns of wealth formation and charitable behavior in real communities.

ACKNOWLEDGEMENTS

The authors sincerely acknowledge the financial and institutional support provided by the Institute for Research and Community Service (LPPM), Universitas Sebelas Maret (UNS). The authors also thank everyone who contributed to this research through constructive feedback, valuable discussions, and support during the preparation and improvement of this manuscript.

REFERENCES

- Afrimaigus, R., & Mohammed, A. S. (2026). Wellness through Sharing : The Trend of Routine Almsgiving in the Mental Health of Modern Muslims. *Asian Journal of Muslim Philanthropy and Citizen Engagement*, 2(1), 99–123. <https://doi.org/10.63919/ajmpce.v2i1.66>
- Agustin, F., Panjaitan, T., Khairunisa, Y., & Fadilah, P. (2024). Modifikasi Persamaan Diferensial Biasa dengan Metode Runge Kutta Orde Lima. *Katalis Pendidikan: Jurnal Ilmu Pendidikan Dan Matematika*, 1(4), 45–52. <https://doi.org/10.62383/katalis.v1i4.855>
- Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economics Journal*, 100(401), 464–477. <https://doi.org/10.2307/2234133>
- Assari, S., & Zare, H. (2024). Handing Money to the Poor Is Never Enough: The Impact of Marginalization-Related Diminished Returns. *Global Journal of Epidemiology and Infectious Disease*, 4(1), 34–43. <https://doi.org/10.31586/gjeid.2024.1026>
- Azizah, D., Aprilia, D., & El'Aliy, H. (2025). Teori Perilaku Konsumen Dalam Ekonomi Islam: Antara Kebutuhan dan Kemaslahatan. *Jurnal Ilmiah Ekonomi Dan Manajemen*, 3(11), 155–161. <https://doi.org/10.61722/jiem.v3i11.6970>
- Beik, I. S., & Arsyianti, L. D. (2016). Measuring Zakat Impact on Poverty and Welfare Using CIBEST Model. *Journal of Islam Monetary Economics and Finance*, 1(2), 141–160. <https://doi.org/10.21098/jimf.v1i2.524>
- Bekkers, R., & Wiepking, P. (2010). A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms That Drive Charitable Giving. *Nonprofit and Voluntary Sector Quarterly*, 20(10), 1–50. <https://doi.org/10.1177/0899764010380927>
- Benhabib, J., Bisin, A., & Zhu, S. (2011). The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents. *Econometrica*, 79(1), 123–157. <https://doi.org/10.3982/ECTA8416>
- Dolfin, M., & Lachowicz, M. (2014). Modeling Altruism and Selfishness in Welfare Dynamics: The Role of Nonlinear Interactions. *Mathematical Models and Methods in Applied Sciences*, 24(12), 2361–2381. <https://doi.org/10.1142/S0218202514500237>
- Dragulescu, A., & Yakovenko, V. (2009). Statistical Mechanics of Money, Income, and Wealth. *APS Journal*, 81(4), 1703–1725. <https://doi.org/10.1103/RevModPhys.81.1703>
- Escot, L., & Sandubete, J. E. (2023). Estimating Lyapunov Exponents on a Noisy Environment by Global and Local Jacobian Indirect Algorithms. *Applied Mathematics and Computation*, 436, 1–17. <https://doi.org/10.1016/j.amc.2022.127498>
- Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 817–868. <https://doi.org/10.1162/003355399556151>
- Feinman, G. M., Cervantes Quequezana, G., Green, A., Lawrence, D., Munson, J., Ortman, S., Petrie, C., Thompson, A., & Nicholas, L. M. (2025). Assessing Grand Narratives of Economic Inequality Across Time. *Proceedings of the National Academy of Sciences*, 122(16), 1–8. <https://doi.org/10.1073/pnas.2400698121>
- Hakim, M. T., Salsabillah, F., Zikri, M., & Nasution, A. O. (2025). Penerapan Konsep Fisika Dalam Pengembangan Algoritma Kompresi Data. *Jurnal Studi Multidisipliner*, 9(5), 493–500.
- Hughes, E., Leibo, J., Phillips, M., Tuyls, K., Duenez-Guzman, E., Castañeda, A., Dunning, I., Zhu, T., McKee, K., Koster, R., Roff, H., & Graepel, T. (2018). Inequity Aversion Resolves Intertemporal Social Dilemmas. *Neural and Evolutionary Computing*. <https://doi.org/10.48550/arXiv.1803.08884>
- Jaenudin, M., & Hamdan, A. (2022). Penilaian Dampak Zakat, Infak, Sedekah Terhadap Kemiskinan Spiritual Dan Material Penerima Manfaat Laznas LMI: Pendekatan CIBEST. *Jurnal Ekonomi Syariah Teori Dan Terapan*, 9(3), 362–378. <https://doi.org/10.20473/vol9iss20223pp362-378>
- Lestari, S. F. (2025). Peran Zakat dalam Meningkatkan Kesejahteraan Perekonomian Masyarakat Berdasarkan Perspektif Manajemen Bisnis Syariah. *SANTRI: Jurnal Ekonomi Dan Keuangan Islam*, 3(3), 137–148. <https://doi.org/10.61132/santri.v3i3.1511>
- Majer, J. M., Zhang, K., Zhang, H., Hühne, B. P., & Trötschel, R. (2022). Give and Take Frames in Shared-Resource Negotiations. *Journal of Economic Psychology*, 90, 1–20. <https://doi.org/10.1016/j.joep.2022.102492>
- Mascareño, A. (2022). Critical Transitions in Ecosystems and Society . The Contribution of Sociological Systems Theory to the Analysis of Socio-Environmental Transformations. *Frontiers in Sociology*, 6, 1–14. <https://doi.org/10.3389/fsoc.2021.763453>
- Muqhnay, C. M., Aulia, S., Labib, M., & Sayuti, W. (2026). Peran Zakat , Infaq, Shadaqah, dan Wakaf dalam Memberdayakan Ekonomi Ummat. *Ikhlas: Jurnal Ilmiah Pendidikan Islam*, 3(1), 142–154. <https://doi.org/10.61132/ikhlas.v3i1.1796>
- Rufaedah, D., & Patih, H. (2025). Islamic Social Finance in Indonesia: Literature Review. *Journal of Industrial and Syariah Economics*, 2(2), 76–95. <https://doi.org/10.63321/jise.v2i2.80>
- Saragih, V. Y., Subakti, G. B., Munthe, N. E. B., & Wardani, S. A. (2024). Implementasi Metode Runge-Kutta dalam Simulasi Lintasan Peluru pada Medan Gravitasi Bumi. *Bilangan: Jurnal Ilmiah Matematika, Kebumihan Dan Angkasa*, 2(5), 41–50. <https://doi.org/10.62383/bilangan.v2i5.272>
- Scheffer, M. (2009). *Critical Transitions in Nature and Society*. Princeton University Press. <https://doi.org/10.5406/amerjpsyc.124.3.0365>
- Sen, S., Levi, E., Ramalingam, A., & Chaudhuri, A. (2026). Economic Inequality, Social Capital and The Role of Merit and Luck: An Experimental Study. *Journal of Economic Behavior & Organization*, 246. <https://doi.org/10.1016/j.jebo.2026.107588>
- Sohn, S. H., Yoon, J. M., & Jeon, S. (2024). Inequality as Determinant of Donation: A Theoretical Modeling and Empirical Analysis of Korea. *PLoS ONE*, 19(9), 1–12. <https://doi.org/10.1371/journal.pone.0306370>
- Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos* (2nd ed.). CRC Press. <https://doi.org/10.1201/9780429492563>

- Sugianto, & Mohammad, F. (2024). Waqaf Produktif : Menggerakkan Perekonomian Rakyat Menuju Ketahanan Ekonomi Berkelanjutan. *OPPORTUNITY: Jurnal Penelitian Dan Pengabdian Masyarakat*, 2(1), 136–145. <https://doi.org/10.55352/opportunity>
- Wiepking, P., & Bekkers, R. (2012). Who Gives? A Literature Review of Predictors of Charitable Giving. Part Two: Gender, Marital Status, Income, and Wealth. *Voluntary Sector Review*, 3. <https://doi.org/10.1332/204080512X649379>

A Data-Driven Travel-Time Scenario Framework for the Preliminary Analysis of School Meal Distribution as Education Support

Muliawan Firdaus¹, Marah Doly Nasution²

¹Department of Mathematics, Universitas Negeri Medan, Medan, Indonesia

²Department of Mathematics Education, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

muliawanfirdaus@unimed.ac.id marahdoly@umsu.ac.id

ABSTRACT

School meal distribution supports education, yet variable travel time makes delivery timing uncertain and can erode service reliability. This study builds a data-driven travel-time scenario framework for the preliminary screening of school meal distribution in Medan, Indonesia. The framework integrates school-depot nodes, OpenStreetMap road-network distance matrices processed with OSMnx, pilot TomTom traffic observations, and historical Open-Meteo weather profiles into a scenario travel-time matrix. Its focus is scenario-based travel-time risk screening rather than vehicle-route optimization or operational validation of official MBG/SPPG services, and the link between timeliness and educational outcomes is treated as conceptual motivation rather than a tested relationship. Results use a primary depot-to-school metric and a secondary all-OD network metric. Under the normal scenario, the mean depot-to-school travel time is 7.766 minutes; under a controlled extreme scenario it rises to 13.908 minutes, and the P90 increases from 11.044 to 19.779 minutes. The extreme scenario (S₅) is read as a conservative upper bound rather than a point estimate, because combining the traffic and weather factors risks double-counting. A sensitivity analysis shows a unit-elastic, inverse relationship between baseline speed and travel time. The framework offers an early computational representation of travel-time exposure in education-support school meal distribution. **Keywords:** education-support logistics, pilot traffic observation, road-network distance, school meal distribution, travel-time scenario



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Muliawan Firdaus,
Department of Mathematics,
Universitas Negeri Medan,
Jalan Willem Iskandar Pasar V Medan Estate Medan 20221, Indonesia.
muliawanfirdaus@unimed.ac.id

5. INTRODUCTION

School meal programs are a form of education support, since attending to nutrition can shape learning readiness, attendance, and student wellbeing. A systematic review and meta-analysis by Wang *et al.* (2021) links school feeding in low- and middle-income countries to children's educational and health outcomes, yet finds those effects mixed across outcomes and stresses implementation quality and the need for stronger studies; we therefore treat the relationship as motivating background rather than a settled causal effect. Timing within that window matters as well: Cohen *et al.* (2016) found that the minutes children actually have to eat changed both what they chose and how much they finished. Seen this way, distribution is not simply moving food from place to place; it is one link in a wider education-support system whose payoff depends on governance, supply chains, and delivery that arrives when it should (Drake *et al.*, 2016; Bundy *et al.*, 2017; Bliznashka *et al.*, 2024).

The path from timely distribution to learning runs through several steps: when travel time varies, meal arrival becomes uncertain, an uncertain arrival unsettles the meal schedule, and a disrupted schedule can then bear on intake and on how dependable the nutrition service is. This study engages only the first link. Measuring nutrient intake, concentration, attendance, or learning outcomes lies outside its scope, and the link to educational outcomes is kept conceptual rather than tested directly.

The Indonesian setting makes such analysis timely. The Free Nutritious Meal Program (*Program Makan Bergizi Gratis*, MBG) and its Nutrition Fulfillment Service Units (*Satuan Pelayanan Pemenuhan Gizi*, SPPG) are governed by technical guidelines from the National Nutrition Agency (*Badan Gizi Nasional*, BGN) under Decree No. 401.1 of 2025 (Badan Gizi Nasional, 2025a), with food-safety certification for SPPG set out in Decree No. 52.2 of 2025 (Badan Gizi Nasional, 2025b), together covering menu provision, governance, compliance, and food safety. In a release dated 31 May 2026, BGN reported that between 6 January 2025 and 29 May 2026, of 27,208 SPPG in operation, 8,182 had

been temporarily suspended at some point and 2,213 remained suspended (Badan Gizi Nasional, 2026). We use these figures as policy context, not as analytical data.

Urban planning needs travel-time estimates that go beyond straight-line distance, because vehicles move through a road network shaped by connectivity, intersections, obstructions, and conditions that change over time; the time-dependent vehicle-routing literature shows that both traffic and departure time alter travel time (Malandraki & Daskin, 1992; Ichoua *et al.*, 2003; Gendreau *et al.*, 2015; Blauth *et al.*, 2022). Gajpal *et al.* (2019), studying mid-day meal distribution in India, treat the task as route optimization with delivery ahead of the eating schedule, whereas our aim is earlier: to estimate travel-time exposure from the road network, pilot traffic, and historical weather before any optimization. Methodologically we take a middle path between simple distance-based estimation and full real-time prediction. OSMnx supports reproducible OpenStreetMap road networks (Boeing, 2017, 2022, 2025); related OpenStreetMap work has computed reproducible origin-destination distances and times through a different engine, OSRM (Huber & Rust, 2016), which we cite as context rather than conflate with our pipeline. Since a static network carries no traffic, we add pilot TomTom observations as a scenario traffic-correction factor and Open-Meteo historical weather (Acharya *et al.*, 2024; TomTom, 2025; Open-Meteo, 2026), following the practice of estimating travel time from limited open data when fleet-operational data are unavailable (Boeing & Zhou, 2026; Ludwig *et al.*, 2023), whose authors also warn that a naïve maximum- or constant-speed assumption tends to under-predict travel time, a caution that bears on the constant-speed model used here.

The gap is narrow and deliberate: no simple, open, and reproducible framework yet merges road-network distance, pilot traffic, and historical weather into a preliminary-screening scenario travel-time matrix for school meal distribution in a developing city where fleet data are unavailable. Each component is included because it captures a distinct source of travel-time variation: road-network distance sets the spatial structure of feasible movement but assumes free-flow conditions, pilot TomTom observations add an empirical correction for temporal congestion, and historical Open-Meteo profiles represent weather conditions that can further lengthen travel time. Combining the three is necessary because distance-only estimation tends to underestimate travel-time exposure, while full fleet telemetry is not feasible where operational fleet data are unavailable. Both Gajpal *et al.* (2019) and the WFP Route the Meals initiative (World Food Programme Innovation Accelerator, 2026) carry out real route optimization, the latter reportedly cutting lead time, distance, and emissions in the field; our framework sits one stage earlier and does not compete at that level, so we cite Route the Meals for positioning rather than as a quantitative benchmark. The contribution is a data-driven travel-time scenario framework comprising a node representation of depot and schools, a road-network distance matrix as the spatial basis, a traffic factor from pilot TomTom data, a weather factor from Open-Meteo, a set of normal, pilot-traffic, rainy, severe-congestion, and controlled-extreme scenarios with a non-composite comparator, and a sensitivity analysis on the baseline-speed parameter. We position it as preliminary travel-time risk screening, not an operational digital twin, an operational validation of MBG/SPPG, a food-safety model, or a final route-optimization model.

6. RESEARCH METHOD

A. Research Design

The study uses a computational design based on secondary data and API observations. This design is transparent, reproducible, and relatively economical, since it builds a consistent origin-destination matrix from open data and API observations without requiring proprietary fleet telemetry. It can also be updated as new observations become available and lets several scenarios be tested without intervening in actual school-meal operations. The TomTom observations are used only as a pilot traffic correction rather than a full multi-day profile. The method proceeds through node definition, construction of the road-network distance matrix, formulation of base travel time, construction of the pilot traffic factor, construction of the weather factor, scenario design, construction of the scenario travel-time matrix, sensitivity analysis, and a reproducibility statement.

B. Data Sources and Analytical Roles

The data are positioned as simulative analytical components for building the travel-time scenario framework. They support a reading of travel-time risk exposure in the school meal distribution system

while remaining distinct from official MBG/SPPG operational records. Table 1 sets out the data sources and their analytical functions.

C. Nodes, Road Network, and Base Travel Time

C. Nodes, Road Network, and Base Travel Time

Let $\mathcal{V}$ be the full node set, comprising the depot and the schools, with $\mathcal{N}$ denote the set of schools and the depot fixed as node 0.

$$\mathcal{V} = \{0\} \cup \mathcal{N}, \quad \mathcal{N} = \{1, 2, \dots, n\} \quad (1)$$

Table 1. Data Sources and Analytical Roles in the Travel-Time Scenario Framework

Data component	Data type	Main variables	Function in the framework	Limitation
School and depot nodes	Coordinate data	District, status, student proxy, latitude, longitude	Define service nodes	Simulative, not official MBG/SPPG data
Road-network distance matrix	OD matrix	L_{ij}	Basis for estimating T_{ij}^0	Static; does not yet carry traffic
TomTom observations	Pilot traffic observations	$CS_p, FFS_p, CTT_p, FFTT_p, \alpha_p$	Form the scenario traffic factor	Not a multi-day profile
Open-Meteo weather data	Historical weather	Precipitation, rain class, β_t	Form the scenario weather factor	Not food/cabin temperature
Scenario travel-time matrix	Derived data	T_{ij}^h, θ_h	Compare scenarios S_1 - S_5	No route optimization yet

Distances come from the OSMnx road-network distance matrix rather than from Euclidean distance, where L_{ij} (km) is the road-network distance from node i to node j . Base travel time then follows from a baseline speed v_0 in km/h (Ludwig *et al.*, 2023):

$$T_{ij}^0 = \frac{L_{ij}}{v_0} \times 60 \quad (2)$$

with T_{ij}^0 in minutes. We set $v_0 = 30$ km/h as a uniform baseline assumption for an initial estimate under normal conditions, not as a validated field speed. Structurally this is a “naive” constant-speed model in the sense of Boeing & Zhou (2026), who find that naïve maximum-speed prediction tends to under-predict true travel time by more than three minutes on average in their case study. The normal-scenario (S_1) estimates may therefore run optimistically low, so the absolute S_1 values are best read as an indicative lower bound; the scenario multiplier θ_h offsets this only in part and does not replace field calibration. Since v_0 governs the results so strongly, we test the model’s sensitivity to its variation.

We use OSM as the working basis for distance because it lends itself to graph-based modeling, though we have not checked its completeness for Medan against official local road records. That no OD pair came back disconnected tells us only that the graph is connected, not that the network is complete. A firmer adequacy check is left for later work, looking at edge density, the spread of segment lengths, and node degree, and spot-checking sample routes against an independent mapping service.

D. Pilot TomTom Traffic Factor

Let $\mathcal{P}$ be the set of TomTom monitoring points. At each point $p \in \mathcal{P}$ we form the traffic factor as the ratio of the observed travel time to the free-flow travel time at that same point.

$$\alpha_p = \frac{CTT_p}{FFTT_p} \quad (3)$$

Here CTT_p is the current travel time and $FFTT_p$ the free-flow travel time at p , in the sense the TomTom Flow Segment Data API uses these terms; for completeness the results tables also list the current speed CS_p and free-flow speed FFS_p . We work with a time ratio rather than a speed ratio

because travel time is exactly what the framework outputs, and such a ratio is bounded below by one whenever traffic runs from normal to congested ($\alpha_p \geq 1$).

Taken together these ratios give a single pilot snapshot across the ten monitoring points rather than a continuous record, and we log the retrieval timestamp so the same pull can be repeated. Two caveats apply: the local weather state at retrieval was not checked against field weather sensors, so α_p may already absorb part of the weather effect, and each value reflects the road segments around a monitoring point rather than the full travel time of an OD route. We therefore read α_p as a pilot scenario factor, not a multi-day empirical traffic profile or a complete representation of traffic in Medan. The median of the TomTom factors drives the pilot-traffic scenario and the 90th percentile the severe-congestion scenario:

$$\alpha_{S_2} = \text{median}(\alpha_p), \quad \alpha_{S_4} = P_{90}(\alpha_p) \quad (4)$$

E. Weather Factor

The weather factor β_t classifies precipitation over the distribution window; its rain thresholds act as operational scenario cut-offs, not as an official climatological classification:

$$\beta_t = \begin{cases} 1.00, & \text{if dry or no rain,} \\ 1.05, & \text{if light rain } (0 < p \leq 1 \text{ mm/h}), \\ 1.10, & \text{if moderate rain } (1 < p \leq 4 \text{ mm/h}), \\ 1.20, & \text{if heavy rain or storm } (p > 4 \text{ mm/h}). \end{cases} \quad (5)$$

Rain is known to alter travel time, speed, and urban traffic intensity (Maze *et al.*, 2006; Cools *et al.*, 2010; Tsapakis *et al.*, 2013), which supports the ordinal form used here, the multiplier grows with rainfall intensity. That literature does not, however, fix the specific values 1.05/1.10/1.20, so we state plainly that the β_t are assumed parameters rather than the result of local calibration for Medan; their influence should be judged alongside the sensitivity analysis and the non-composite comparator, with field calibration left to future work. The scenario weather factor is taken as the maximum over the distribution window $\mathcal{T}_d$:

$$\beta_h = \max_{t \in \mathcal{T}_d} \beta_t \quad (6)$$

F. Scenario-Multiplier Design and the Double-Counting Risk

Scenario travel time applies a total multiplier θ_h to the base time:

$$T_{ij}^h = T_{ij}^0 \times \theta_h \quad (7)$$

For the main scenarios the multiplier is composite:

$$\theta_h = \alpha_h \beta_h \quad (8)$$

Eq. (8) acts as a controlled scenario multiplier, not as a causal model establishing independence between traffic and weather. Because the weather state at the time of the TomTom observations was not independently verified, an α_h that already embeds a rain effect would make the product $\alpha_h \beta_h$ count rain twice (double-counting). For that reason the composite scenarios, and S_5 in particular, are labeled as conservative upper-bound screening scenarios rather than point estimates or daily empirical predictions.

Where the weather-traffic interaction cannot be verified, a non-composite alternative serves as a sensitivity comparator:

$$\tilde{\theta}_h = \max(\alpha_h, \beta_h) \quad (9)$$

Eq. (9) does not combine the two effects multiplicatively, so it is more conservative against double-counting. Our headline numbers keep the composite form of Eq. (8) and should be read as no more than an early scenario reading; for S_5 we set the non-composite value beside it so the two can be compared directly. The five scenarios are:

$$S = \{S_1, S_2, S_3, S_4, S_5\} \quad (10)$$

with traffic factors:

$$\alpha_h = \begin{cases} 1.0000, & h = S_1, \\ 1.2927, & h = S_2, \\ 1.0000, & h = S_3, \\ 1.3567, & h = S_4, \\ 1.20 \times 1.3567 = 1.6281, & h = S_5, \end{cases} \quad (11)$$

and weather factor:

$$\beta_h = \begin{cases} 1.0000, & h \in \{S_1, S_2, S_4\}, \\ 1.1000, & h \in \{S_3, S_5\}. \end{cases} \quad (12)$$

The value $\alpha_{S_5} = 1.6281$ uses the full-precision $P90(\alpha_p) \approx 1.35673$, so that $1.20 \times 1.35673 \approx 1.6281$; the four-decimal value 1.3567 would instead give 1.6280. Every later computation carries full precision, so this ± 0.0001 gap leaves the rounded entries in Table 5 untouched. The controlled-extreme multiplier for S_5 is therefore:

$$\theta_{S_5} = \alpha_{S_5} \beta_{S_5} = (1.20 \times 1.3567) \times 1.1000 = 1.7909 \quad (13)$$

We express each scenario's effect as its percentage change against the normal baseline:

$$\Delta_h = \frac{\bar{T}^h - \bar{T}^{S_1}}{\bar{T}^{S_1}} \times 100\% \quad (14)$$

G. Analysis Metrics

We keep two travel-time statistics apart throughout. The depot-to-school statistic is the primary one, since the service we model runs from a single depot out to the schools. The all-OD-pairs statistic is secondary: it describes the general travel-time structure among the nodes rather than the service itself. So the two are never confused, every figure in the Results carries an explicit label, “depot-to-school” or “all-OD.” The depot-to-school pairs are:

$$\mathcal{D} = \{(0, i) \mid i \in \mathcal{N}\} \quad (15)$$

The mean depot-to-school travel time under scenario h is:

$$\bar{T}_{0\mathcal{N}}^h = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} T_{0i}^h \quad (16)$$

For the all-valid-OD statistic, let $\mathcal{E}$ collect every valid OD pair, leaving out the diagonal and any pair flagged as disconnected. The mean travel time across these pairs is:

$$\bar{T}_{\mathcal{E}}^h = \frac{1}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} T_{ij}^h \quad (17)$$

We attach a student-count proxy to each node only to convey the scale of service it represents. Since the scenarios deal in travel time and not in portion counts, vehicle capacities, or capacitated routes, this proxy never enters the travel-time model as a demand quantity.

H. Sensitivity Analysis

For any scenario with a fixed multiplier θ_h , the mean travel time takes the closed form:

$$\bar{T}^h = \frac{60 \cdot \bar{L} \cdot \theta_h}{v_0} \quad (18)$$

Differentiating with respect to the baseline speed v_0 gives:

$$\frac{\partial \bar{T}^h}{\partial v_0} = -\frac{60 \cdot \bar{L} \cdot \theta_h}{v_0^2} = -\frac{\bar{T}^h}{v_0} \quad (19)$$

From this, the elasticity in v_0 works out to:

$$\varepsilon_{v_0} = \frac{\partial \bar{T}^h / \bar{T}^h}{\partial v_0 / v_0} = -1 \quad (20)$$

If a scenario follows the multiplicative branch $\theta_h = \alpha_h \beta_h$, the elasticities with respect to α_h and β_h are likewise unit-elastic:

$$\varepsilon_{\alpha} = +1, \quad \varepsilon_{\beta} = +1 \quad (21)$$

Eq. (20) and (21) hold for the scalar model used here and nowhere beyond it. The unit elasticity $\varepsilon_{v_0} = -1$ is a property of that algebra, not a law of urban traffic: once volume approaches a road's critical capacity, the ties among speed, volume, queueing, and travel time can turn nonlinear in ways the scalar form does not capture. We therefore read the sensitivity results as a guide to how parameter uncertainty travels through an early framework, not as a replacement for a fuller dynamic traffic model.

I. Reproducibility

The run writes out its analysis tables, figures, processed data, the scenario travel-time matrix, the configuration file, and procedure logs, and these together act as replication material for the manuscript. The same record holds the TomTom retrieval metadata (timestamp and access date), the baseline-speed setting, the scenario factors, the decimal precision behind each multiplier, and notes on how the data were processed.

7. RESULTS AND DISCUSSION

A. Characteristics of the School Service Nodes

The node set is one simulative depot together with 99 schools. For each school we hold its district, status, coordinates, a student-count proxy, and its initial distance to the depot. These nodes are simulative; they stand in for service points and are not drawn from official MBG recipient lists or live SPPG operations. Because the proxy is built by spreading per-district reference figures by weight, some schools carry non-integer student counts. To keep the article compact, Table 2 reports the nodes aggregated to district level.

Table 2. Summary of School Service Nodes by District

District	No. of schools	Total student proxy	Mean student proxy	Mean distance to depot (km)	Max. distance to depot (km)
Medan Amplas	9	3,645	405.000	3.636	5.550
Medan Barat	4	1,767	441.750	4.997	5.460
Medan Baru	8	1,596.035	199.504	5.386	5.650
Medan Denai	15	4,138.035	275.869	3.937	5.280
Medan Kota	32	6,556.017	204.876	2.779	5.510
Medan Maimun	11	3,567	324.273	3.837	5.810
Medan Petisah	4	321.051	80.263	4.895	5.480
Medan Polonia	10	1,791	179.100	4.542	5.800
Medan Timur	6	1,950	325.000	5.263	5.790

As Table 2 and Fig 1 show, the service nodes spread unevenly across nine districts: Medan Kota carries the most, followed by Medan Denai. That imbalance matters for distribution, since the busier districts are likely to concentrate demand, even though we do not yet optimise the order in which vehicles visit them.

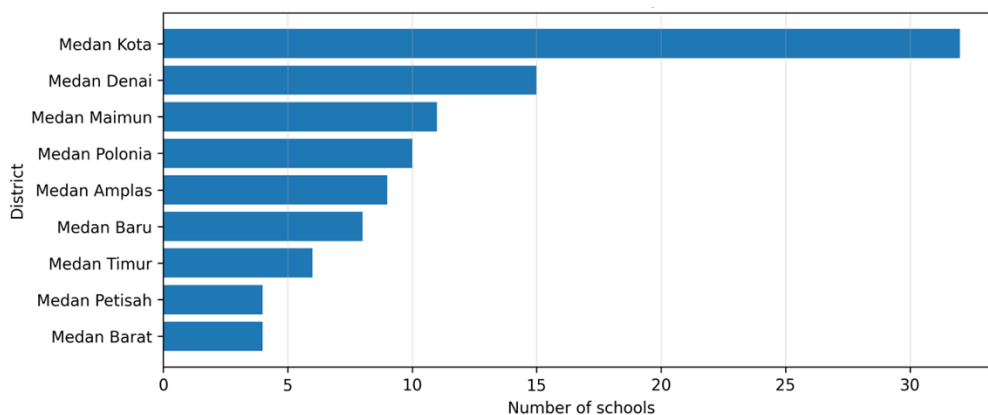


Fig 1. Distribution of School Service Nodes by District

B. Road-Network Distance Statistics

The inter-node distance matrix is used to compute road-network distances. Where it appears, a value of 9999.0 marks a disconnected origin-destination pair and is excluded from the valid statistics. The value 9999.0 is not a distance threshold taken from the literature but a sentinel code assigned within the data-processing procedure. It is written whenever the shortest-path search returns no finite route between an origin and a destination, that is, when the two nodes are not connected on the routable road network. Because it does not represent an observed distance, it is excluded from all valid-distance and travel-time statistics. Table 3 reports the distance-matrix validation.

Four OD pairs with $L_{ij} = 0.000$ km are excluded from the scenario analysis because two nodes share the same nearest OSMnx network point. Removing these four pairs corresponds to about 0.04% of the total OD pairs (4/9900) and does not materially change the distribution. The off-diagonal count of 9,900 is consistent with 100×99 , and the effective count is $9,896 = 9,900 - 4$.

Table 3. Road-Network Distance Statistics

Indicator	Value
Number of row nodes	100
Number of column nodes	100
Off-diagonal OD pairs	9,900
Topologically valid OD pairs	9,900
Disconnected OD pairs 9999.0	0
OD pairs effective for scenarios	9,896
Mean distance, all valid OD (km)	4.966
Median distance, all valid OD (km)	4.927
P90 distance, all valid OD (km)	7.881
Maximum distance, all valid OD (km)	11.571
Mean depot-to-school distance (km)	3.883
Median depot-to-school distance (km)	4.131
P90 depot-to-school distance (km)	5.522
Maximum depot-to-school distance (km)	6.219

Table 3 and the distance distribution in Fig 2 bring out an important difference between the all-OD statistic and the depot-to-school statistic. The mean distance across all valid OD pairs reaches 4.966 km, whereas the mean depot-to-school distance is 3.883 km. For that reason, the main distribution-service results are read through the depot-to-school statistic, with the all-OD statistic serving as a secondary network picture.

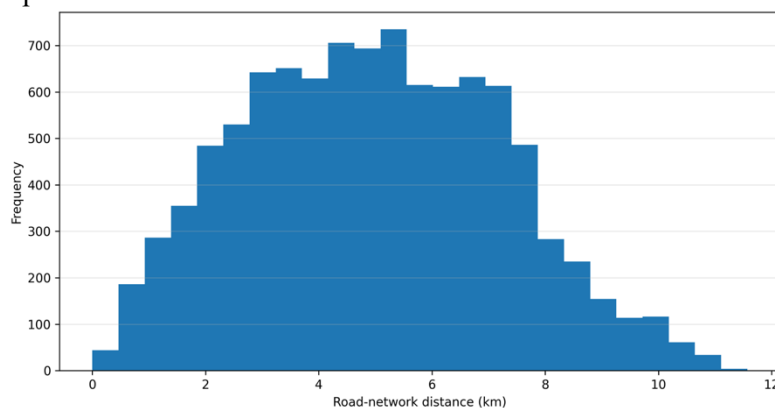


Fig 2. Distribution of Road-Network Distances

C. External Scenario Parameters: TomTom and Open-Meteo

To keep the presentation compact, we combine the summary of TomTom observations and the weather profiles into a single set of external scenario parameters. TomTom forms the traffic factor and Open-Meteo the weather factor. Table 4 sets out the scenario parameters used in the travel-time computation.

We note that the TomTom observations are a pilot snapshot across 10 monitoring points. The execution log records the TomTom retrieval timestamp and provides reproducibility information. The weather state at the time of TomTom retrieval has not been independently verified. The composite scenario S_5 is therefore labeled an upper-bound screening scenario (a conservative upper bound), not the primary empirical prediction.

Table 4. External Scenario Parameters and Travel-Time Multipliers

Scenario	Scenario name	Traffic-factor basis	α_h	Weather-factor basis	β_h	θ_h	Interpretation
S_1	Normal baseline	No traffic penalty	1.0000	Dry conditions	1.0000	1.0000	Base condition
S_2	Pilot TomTom traffic	Median α_p over 10 TomTom points	1.2927	Dry conditions	1.0000	1.2927	Pilot-traffic scenario
S_3	Rainy weather	No traffic penalty	1.0000	Max. rain factor over the distribution window	1.1000	1.1000	Rainy scenario
S_4	Severe congestion	$P_{90}(\alpha_p)$ over 10 TomTom points	1.3567	Dry conditions	1.0000	1.3567	Traffic stress test
S_5	Controlled extreme stress	$1.20 \times P_{90}(\alpha_p)$	1.6281	Max. rain factor over the distribution window	1.1000	1.7909	Conservative upper bound

Table 4 and Fig 3 show that the pilot traffic factor is larger than the moderate-rain weather factor. The values $\alpha_{S_2} = 1.2927$ and $\alpha_{S_4} = 1.3567$ indicate that the TomTom observations capture a rise in travel time relative to free-flow conditions at the monitoring points. The value $\beta_h = 1.1000$, meanwhile, represents the moderate weather penalty used for the rainy scenario (Sakhare *et al.*, 2023; Xiong *et al.*, 2022). Combining the two in S_5 calls for cautious reading, because the traffic and weather effects may overlap.

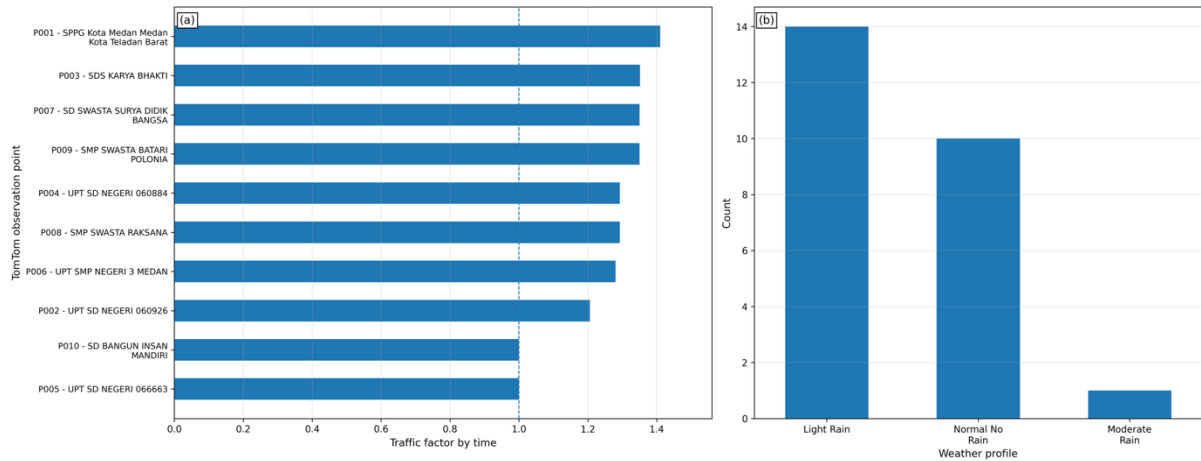


Fig 3. (a) Pilot TomTom Traffic Factor by Monitoring Point; (b) Weather Profiles over the Distribution Window

D. Main Results: Depot-to-School Travel Time

Because the service runs from a single depot to schools, the main results use the depot-to-school statistic. Table 5 reports travel time for the five scenarios. All values are depot-to-school statistics, and they scale directly with θ_h . The mean rises from 7.766 minutes under the normal scenario (S_1) to 13.908 minutes under the controlled-extreme scenario (S_5), and P90 from 11.044 to 19.779 minutes (Fig 4), showing that the traffic and weather factors enlarge distribution-time exposure, most of all for schools in the upper tail of the distance distribution.

S_5 is meant as a conservative ceiling, not a prediction for any single day. If the TomTom factor already carries part of the rain signal, multiplying it by the weather factor in the composite $\theta_{S5} = \alpha_{S5}\beta_{S5}$ would count that rain twice and overstate the combined impact. A non-composite comparator $\theta_{S5} = \max(\alpha_{S5}, \beta_{S5}) = \max(1.6281, 1.1000) = 1.6281$ gives a mean of $7.766 \times 1.6281 = 12.644$ minutes and a P90 of $11.044 \times 1.6281 = 17.981$ minutes. From S_1 through S_4 only one factor is ever active, so the composite and non-composite multipliers coincide and the comparator changes nothing there; it bears on S_5 alone. The band this opens for S_5 (12.644 minutes, non-composite) against 13.908 minutes (composite), is how we keep the double-counting risk in plain view.

Table 5. Depot-to-School Travel-Time Results by Scenario

Scenario	Scenario name	θ_h	Mean (min)	Median (min)	P90 (min)	Max. (min)	Mean increase from S_1
S_1	Normal baseline	1.0000	7.766	8.262	11.044	12.438	0.00%
S_2	Pilot TomTom traffic	1.2927	10.039	10.680	14.276	16.078	29.27%
S_3	Rainy weather	1.1000	8.543	9.088	12.148	13.682	10.00%
S_4	Severe congestion	1.3567	10.537	11.209	14.984	16.875	35.67%
S_5	Controlled extreme stress	1.7909	13.908	14.796	19.779	22.275	79.09%

For a planner, the use of these figures is in pointing to the schools that deserve a closer look, in when their vehicles should depart, whether an intermediate drop-off would help, or whether a backup vehicle should be held in reserve.

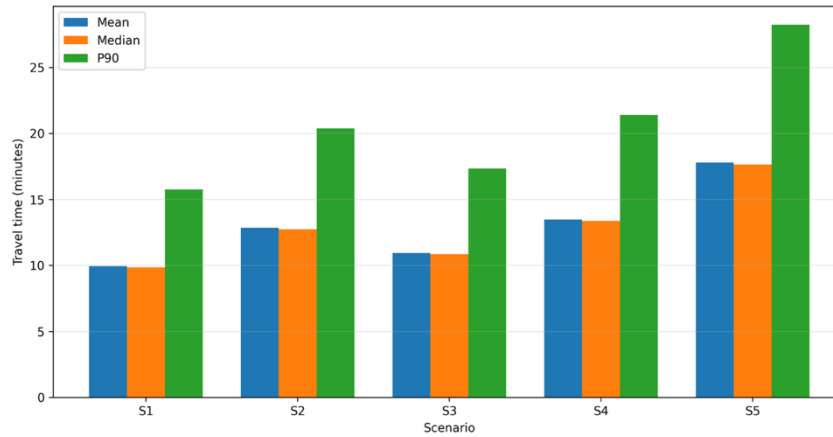


Fig 4. Comparison of Depot-to-School Travel Time by Scenario

Read at the network level (all-OD) instead, the mean travel time climbs from 9.935 minutes at S_1 to 17.793 minutes at S_5 , and the P90 from 15.763 to 28.230 minutes. These all-OD figures are higher than the depot-to-school figures because they include school-to-school trips that are not always relevant as primary service journeys. The all-OD statistic is therefore used to read network structure, not as a direct indicator of depot-to-school distribution performance. The effective speed under the extreme scenario for all OD pairs is:

$$v_{\text{eff},S_5} = \frac{\bar{L}_\varepsilon}{\bar{T}_\varepsilon^{S_5}/60} = \frac{4.966}{17.793/60} \approx 16.75 \text{ km/h} \quad (22)$$

This equals $v_0/\theta_{S_5} = 30/1.7909 \approx 16.75 \text{ km/h}$ and indicates that the controlled-extreme scenario implies a drop in the network's effective speed to about 16.75 km/h, averaged over all valid OD pairs.

E. Sensitivity to the Baseline Speed

We run the sensitivity analysis over $v_0 \in \{25,30,35,40\} \text{ km/h}$. On the all-valid-OD statistic, when $v_0 = 25 \text{ km/h}$ the mean S_1 travel time rises to 11.922 minutes and the mean S_5 travel time to 21.352 minutes. When $v_0 = 40 \text{ km/h}$, the mean S_1 travel time falls to 7.451 minutes and the mean S_5 to 13.345 minutes. Both movements track the scalar relationship set out in Eq. (18)-(20).

This sensitivity asks to be read with care. The unit relationship $\varepsilon_{v_0} = -1$ falls straight out of the form $\bar{T}^h = 60\bar{L}\theta_h/v_0$: change v_0 by some fraction and travel time shifts by the same fraction in the opposite direction. Real urban traffic is not so tidy, speed responds to vehicle volume, road capacity, intersections, queueing, and local disturbances, none of which the scalar form sees. Holding speed fixed at $v_0 = 30 \text{ km/h}$ can moreover under-state real travel time (Boeing & Zhou, 2026), which is just what the v_0 sweep is there to probe: how far the estimates move if the true field speed is lower. We use it to read the model's internal structure, then, not to stand in for a macroscopic traffic model or a dynamic simulation.

F. Implications for Education-Support Logistics

The framework can flag risk exposure in school meal distribution before full operational data exist, since delivery delay may conceptually disrupt the stability of meal schedules and since meal duration and scheduling are known to relate to student consumption (Cohen *et al.*, 2016); because the study does not directly test intake, concentration, attendance, or learning, we treat the link to educational outcomes as a conceptual motivation, the more so as the meta-analytic evidence is mixed and contingent on implementation quality (Wang *et al.*, 2021), so the educational contribution lies in offering an early instrument for reading logistical risk rather than in proving educational impact. Practically, the depot-to-school P90 values in Table 5 identify the upper-tail schools that planners can prioritize in service-zone planning, departure scheduling, the choice of intermediate distribution points, or the provision of

extra vehicles, and these can later be tied to data on vehicle capacity, school service times, actual arrival times, and food temperature. More broadly, WFP Route the Meals shows that school meal systems benefit from fuller distribution-network analysis once operational data, fleet capacity, and vehicle routes are available (World Food Programme Innovation Accelerator, 2026), and Gajpal *et al.* (2019) present a route-optimization formulation for mid-day meal delivery, whereas this study sits at an earlier stage, screening travel-time risk from scenarios and open data as a basis for moving toward route optimization in subsequent work.

G. Food Safety and the Limits of Interpretation

The travel-time results in this article are not used to infer food-safety compliance or violation. The analysis does not measure food temperature, vehicle-cabin temperature, food holding time, or food-handling procedures. The WHO Five Keys to Safer Food manual and the SPPG food-safety certification guidance provide context that food distribution requires risk control, but the travel-time values in this study only flag potential logistical exposure that warrants further examination (World Health Organization, 2006; Badan Gizi Nasional, 2025b).

The maximum of 22.275 minutes on a depot-to-school route under the controlled-extreme scenario, then, does not indicate that operations breach food safety. What it does say is narrower: under heavier traffic and rain, travel time can lengthen, and that lengthening belongs among the design considerations for thermal control, distribution scheduling, and service SOPs.

H. Limitations

Several limits bound how far these results should be pushed. On the data side, the school records stand in as simulative service nodes with no official MBG/SPPG status, the student counts are a service proxy rather than real portion figures, and the TomTom readings are a ten-point pilot snapshot of the segments around each monitoring point, not a multi-day profile, and not the end-to-end time of any full OD route. On the assumptions, the weather state at TomTom retrieval was not independently verified, which creates the double-counting risk that leads us to label S_5 a conservative upper bound and to report the non-composite comparator $\max(\alpha, \beta)$; the baseline speed $v_0 = 30$ km/h and the weather multipliers 1.05/1.10/1.20 are assumed rather than calibrated for Medan, and because a constant-speed model can under-predict real travel time (Boeing & Zhou, 2026), the absolute S_1 values should be read as indicative, while the weather factor is applied homogeneously across the network even though rain's effect varies with ponding, drainage, elevation, and traffic density, and the OSM network has not been validated against official local road data. In scope, the model does not yet include vehicle capacity, service times, school reception schedules, actual route ordering, or field validation. Future work should therefore pursue direct empirical validation, for example comparing the estimates against real vehicle-trip data or an independent routing service on selected routes, calibrate v_0 and β against pilot data, and add multi-day traffic data, vehicle departure and arrival times, food-temperature data, vehicle capacities, and a route-optimization model.

8. CONCLUSION

This article produces a data-driven travel-time scenario framework to support the preliminary analysis of school meal distribution as part of education-support logistics. The framework integrates depot and school node data, a road-network distance matrix from OSMnx, pilot TomTom traffic observations, and historical Open-Meteo weather profiles.

The mathematical contribution lies in transforming road-network distance L_{ij} into scenario travel-time estimates T_{ij}^h through the parameters v_0 , α_h , β_h , and θ_h . The model places the depot-to-school result as the primary metric because it matches the service context of distribution from a single depot to schools. The all-OD-pairs statistic is still reported as a secondary network reading.

The main results show the mean depot-to-school travel time rising from 7.766 minutes under the normal scenario to 13.908 minutes under the controlled-extreme scenario, while the non-composite comparator $\max(\alpha, \beta)$ places S_5 at 12.644 minutes; we interpret this range as a conservative upper

bound. The depot-to-school P90 increases from 11.044 to 19.779 minutes. On the all-OD statistic, mean travel time rises from 9.935 to 17.793 minutes, while P90 increases from 15.763 to 28.230 minutes. The sensitivity analysis shows that uncertainty in the baseline speed v_0 produces a proportional and opposite change in travel time ($\varepsilon_{v_0} = -1$) within the scalar model used here.

We position the framework as an early computational representation, not as final route optimization, an operational validation of SPPG, a food-safety model, or an operational digital twin, and it occupies an explicitly earlier stage than existing school-meal route-optimization work (Gajpal *et al.*, 2019; World Food Programme Innovation Accelerator, 2026). Subsequent research should add multi-day traffic data, actual vehicle arrival times, vehicle capacities, school service times, food-temperature conditions, parameter calibration, and a route-optimization model, so that the analysis can move from an early scenario toward firmer operational decision-making.

ACKNOWLEDGEMENTS

This research was funded by the PNPB (Non-Tax State Revenue) of Universitas Negeri Medan for the 2026 fiscal year under Rector's Decree No. 0236/UN33/KPT/2026. We are grateful to the OpenStreetMap contributors, OSMnx, TomTom, and Open-Meteo, whose freely shared resources made this work possible.

REFERENCES

- Acharya, S., Garikapati, V. M., Allen, M., Lyu, M., Hoehne, C., Sharda, S., & Fitzgerald, R. (2024). Enriching OpenStreetMap network data for transportation applications: Insights into the impact of urban congestion on accessibility. *Journal of Transport Geography*, 104096. <https://doi.org/10.1016/j.jtrangeo.2024.104096>
- Badan Gizi Nasional. (2025a). *Petunjuk teknis tata kelola penyelenggaraan Program Makan Bergizi Gratis Tahun Anggaran 2026* (Keputusan Kepala Badan Gizi Nasional Republik Indonesia Nomor 401.1 Tahun 2025, ditetapkan 29 Desember 2025). Badan Gizi Nasional Republik Indonesia. <https://www.bgn.go.id/juknis/IB1PKg-petunjuk-teknis-tata-kelola-penyelenggaraan-program-makan-bergizi-gratis>
- Badan Gizi Nasional. (2025b). *Pedoman sertifikasi keamanan pangan pada Satuan Pelayanan Pemenuhan Gizi* (Keputusan Kepala Badan Gizi Nasional Republik Indonesia Nomor 52.2 Tahun 2025). Badan Gizi Nasional Republik Indonesia. <https://www.bgn.go.id/juknis/xEm8Jv-pedoman-sertifikasi-keamanan-pangan-pada-satuan-pelayanan-pemenuhan-gizi>
- Badan Gizi Nasional. (2026, May 31). *Sejak 6 Januari 2025-29 Mei 2026, 8.182 SPPG pernah di-suspend, 2.213 SPPG kini masih dalam posisi suspend* [Press release]. Badan Gizi Nasional Republik Indonesia. <https://www.bgn.go.id/news/siaran-pers/sejak-6-januari-2025-29-mei-2026-8182-sppg-pernah-di-suspend-2213-sppg-kini-masih-dalam-posisi-suspend>
- Blauth, J., Held, S., Müller, D., Schlomberg, N., Traub, V., & Tröbst, T. (2022). *Vehicle routing with time-dependent travel times: Theory, practice, and benchmarks*. arXiv. <https://doi.org/10.48550/arXiv.2205.00889>
- Bliznashka, L., Blakstad, M. M., Berhane, Y., Ismail, A., & Fawzi, W. W. (2024). School feeding for improving child nutrition in conflict-affected settings: Feasibility and cost efficiency of alternative models in Yemen. *Annals of the New York Academy of Sciences*, 1540(1), 251–264. <https://doi.org/10.1111/nyas.15222>
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126-139. <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>
- Boeing, G. (2022). Street network models and indicators for every urban area in the world. *Geographical Analysis*, 54(3), 519-535. <https://doi.org/10.1111/gean.12281>
- Boeing, G. (2025). Modeling and analyzing urban networks and amenities with OSMnx. *Geographical Analysis*, 57(4), 567-577. <https://doi.org/10.1111/gean.70009>
- Boeing, G., & Zhou, Y. (2026). Travel time prediction from sparse open data. *International Journal of Geographical Information Science*. Advance online publication. <https://doi.org/10.1080/13658816.2026.2628193>

- Bundy, D., de Silva, N., Horton, S., Jamison, D. T., & Patton, G. C. (Eds.). (2017). *Disease Control Priorities: Child and Adolescent Health and Development* (3rd ed., Vol. 8). World Bank. <https://doi.org/10.1596/978-1-4648-0423-6>
- Cohen, J. F. W., Jahn, J. L., Richardson, S., Cluggish, S. A., Parker, E., & Rimm, E. B. (2016). Amount of time to eat lunch is associated with children's selection and consumption of school meal entrée, fruits, vegetables, and milk. *Journal of the Academy of Nutrition and Dietetics*, 116(1), 123-128. <https://doi.org/10.1016/j.jand.2015.07.019>
- Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, 2(1), 60-68. <https://doi.org/10.1175/2009WCAS1014.1>
- Drake, L., Woolnough, A., Burbano, C., & Bundy, D. (Eds.). (2016). *Global School Feeding Sourcebook: Lessons from 14 Countries*. Imperial College Press.
- Gajpal, Y., Roy, V., & Sahay, B. S. (2019). Vehicle routing for a mid-day meal delivery distribution system. *Heliyon*, 5(2), e01158. <https://doi.org/10.1016/j.heliyon.2019.e01158>
- Gendreau, M., Ghiani, G., & Guerriero, E. (2015). Time-dependent routing problems: A review. *Computers & Operations Research*, 64, 189-197. <https://doi.org/10.1016/j.cor.2015.06.001>
- Huber, S., & Rust, C. (2016). Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM). *The Stata Journal*, 16(2), 416-423. <https://doi.org/10.1177/1536867X1601600209>
- Ichoua, S., Gendreau, M., & Potvin, J.-Y. (2003). Vehicle dispatching with time-dependent travel times. *European Journal of Operational Research*, 144(2), 379-396. [https://doi.org/10.1016/S0377-2217\(02\)00147-9](https://doi.org/10.1016/S0377-2217(02)00147-9)
- Ludwig, C., Psotta, J., Buch, A., Kolaxidis, N., Fendrich, S., Zia, M., Fürle, J., Rousell, A., & Zipf, A. (2023). Traffic speed modelling to improve travel time estimation in openrouteservice. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W7-2023*, 109–116. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-109-2023>
- Malandraki, C., & Daskin, M. S. (1992). Time dependent vehicle routing problems: Formulations, properties and heuristic algorithms. *Transportation Science*, 26(3), 185-200. <https://doi.org/10.1287/trsc.26.3.185>
- Maze, T. H., Agarwal, M., & Burchett, G. (2006). Whether weather matters to traffic demand, traffic safety, and traffic operations and flow. *Transportation Research Record*, 1948(1), 170-176. <https://doi.org/10.3141/1948-19>
- Open-Meteo. (2026). *Historical Weather API documentation*. Open-Meteo. Retrieved 2026, from <https://open-meteo.com/en/docs/historical-weather-api>
- Sakhare, R. S., Zhang, Y., Li, H., & Bullock, D. M. (2023). Impact of rain intensity on interstate traffic speeds using connected vehicle data. *Vehicles*, 5(1), 133–155. <https://doi.org/10.3390/vehicles5010009>
- TomTom. (2025). *Flow Segment Data: Traffic API documentation*. TomTom Developer Portal. <https://developer.tomtom.com/traffic-api/documentation/tomtom-maps/v1/traffic-flow/flow-segment-data>
- Tsapakis, I., Cheng, T., & Bolbol, A. (2013). Impact of weather conditions on macroscopic urban travel times. *Journal of Transport Geography*, 28, 204-211. <https://doi.org/10.1016/j.jtrangeo.2012.11.003>
- Wang, D., Shinde, S., Young, T., & Fawzi, W. W. (2021). Impacts of school feeding on educational and health outcomes of school-age children and adolescents in low- and middle-income countries: A systematic review and meta-analysis. *Journal of Global Health*, 11, 04051. <https://doi.org/10.7189/jogh.11.04051>
- World Food Programme Innovation Accelerator. (2026). *Route the Meals: Optimizing delivery networks and routes for efficient humanitarian supply chains*. World Food Programme. Retrieved 2026, from <https://innovation.wfp.org/project/route-meals>
- World Health Organization. (2006). *Five keys to safer food manual*. World Health Organization. <https://www.who.int/publications/i/item/9789241594639>
- Xiong, M.-M., Han, Z.-Y., Lu, B., Guo, J., Xu, S., Yang, Y.-J., & Huang, H. (2022). Effects of rainfall on the weekday traffic flow in major cities of the Beijing–Tianjin–Hebei region, China, in 2021. *Advances in Climate Change Research*, 13(6), 858–867. <https://doi.org/10.1016/j.accre.2022.11.009>

Ethnomodelling Indigenous Knowledge for Decolonizing Mathematics Education: Insights from The Passaungang Tau Heritage Site

Muhammad Nurhusain¹, Rizky Ramadhana², Rezki Novianti³, Abdul Hadi⁴

^{1,2,4}Mathematics Education Study Program, STKIP YPUP Makassar, Makassar, Indonesia

³Management Study Program, Universitas Muhammadiyah Mamuju, Mamuju, Indonesia

¹zein.alhusain@gmail.com, ²rizkyramadhana53@gmail.com, ³novisyam85@gmail.com, ⁴abdulhadi030786@gmail.com

ABSTRACT

This study is grounded in the growing global discourse on epistemological inclusivity and culturally responsive mathematics education, which recognizes indigenous knowledge as a legitimate source of mathematical learning. The study aims to identify the mathematical meanings embedded in the Passaungang Tau cultural heritage site and examine their pedagogical implications for the decolonization of mathematics education through an ethnomodelling approach. An ethnographic qualitative design was employed, with data collected through observation, semi-structured interviews, documentation, and cultural interpretation involving community elders, cultural practitioners, and traditional stakeholders in the Gantarang Keke community, South Sulawesi, Indonesia. The findings indicate that the Passaungang Tau site contains implicit mathematical knowledge reflected in spatial geometry, traditional measurement systems, and numeracy practices rooted in local socio-cultural traditions. Through an ethnomodelling framework, these practices were systematically translated into formal mathematical concepts. Preliminary findings also suggest that culturally contextualized mathematics learning enhances students' engagement, conceptual understanding, and cultural awareness. Ethnomodelling is therefore positioned as a transformative pedagogical approach that bridges indigenous knowledge and formal school mathematics while supporting the development of contextualized and culturally sustaining mathematics instruction.

Keywords: Ethnomodelling, Ethnomathematics, Indigenous Knowledge, Decolonizing Mathematics Education, Culturally Responsive Pedagogy



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Muhammad Nurhusain,
Mathematics Education,
STKIP YPUP Makassar,
Jalan Andi Tonro No. 17 Makassar 90221, Indonesia.
zein.alhusain@gmail.com

1. INTRODUCTION

For a long time, mathematics has been framed as a universal, value-neutral body of knowledge, often separated from the sociocultural contexts in which it developed and is applied. However, contemporary research has increasingly challenged this view, showing that mathematical knowledge is socially constructed, culturally embedded, and historically situated (Bishop, 1991; Valero, 2020). This shift gave rise to the field of ethnomathematics, which examines how mathematical ideas emerge, are presented, and are applied across different cultural practices (D'Ambrosio, 1985; Rosa & Orey, 2021). At the same time, critical perspectives in mathematics education have argued that knowledge production is never neutral: it is shaped by researchers' positions and socio-historical contexts, which in turn shape how mathematics is defined and taught (Foote & Gau Bartell, 2011).

According to Meaney & Trinick (2020), Indigenous knowledge systems are valid sources of mathematical thinking, encompassing spatial reasoning, measurement systems, and counting practices rooted in daily cultural activities. Research conducted Prahmana & D'Ambrosio (2020) and Umbara et al. (2021) confirmed that the local mathematical structures embedded in cultural artifacts and traditional customs can be integrated into formal education. Marshall (2023) and Nolan & Xenofontos (2023) noted that this integration improves students' conceptual understanding, strengthens their cultural identity, and boosts their learning engagement. Combined with the sociocultural perspective put forward by Gresalfi & Hand (2019), this line of research demonstrates that mathematics learning is inherently deeply tied to identity construction.

Ethnomathematics research is a field dedicated to uncovering the mathematical content inherent in diverse cultures. It has accumulated considerable progress to date, yet one long-standing problem has remained unsolved: mathematical knowledge rooted in specific cultures has never been able to be

transformed into systematic classroom teaching content. Numerous prior studies have consecutively undertaken one common task: proving that mathematical-related concepts are hidden within a wide range of cultural activities and cultural products. For example, Supiyati et al. (2019) identified embedded geometric concepts in the architecture of the Sasak people; in the same year, Pathuddin & Raehana (2019) documented mathematical concepts in the traditional foods of the Bugis people, and discussed whether these contents could serve as materials for mathematics learning. In more recent research, Hidayah et al. (2025) sorted out the ethnomathematical ideas hidden in the Wiwitan Methik Pari ritual, while Rahmawati et al. (2026) discovered geometric shapes and patterns in the traditional snacks of the Sasak people.

The value of these studies is clear: all of them confirm that mathematical knowledge is indeed embedded in various cultural activities, and that the set of mathematics taught in schools is not the only valid form of mathematics. However, their core work only stay at identifying these hidden ethnomathematical elements within cultures, documenting them, and explaining what they are. No study has proposed a systematic process to convert these scattered elements into formal, general mathematical expressions, and then organize them into learning content that can be directly integrated into classrooms.

This widespread flaw aligns precisely with the perspective put forward early on by D'Ambrosio & Rosa (2017): they long argued that ethnomathematics research must not stop at merely recognizing mathematical practices within cultures, but must take a further step to develop actionable, teaching-related measures that can integrate these contents into mainstream mathematics education. Put simply, what this field currently lacks is not the ethnomathematical knowledge that has already been documented. Mathematical contents hidden in cultures around the world have been uncovered by many scholars. What is truly lacking is a systematic, mature framework that can smoothly convert these culturally rooted mathematical ideas into teaching content suitable for standard classrooms.

It so happens that a research method called ethnomodelling can fill this gap. It can build an orderly connection channel between culturally rooted indigenous mathematical knowledge and formal school mathematics. Rosa & Orey (2021) broke down ethnomodelling into three interrelated pathways: emic, etic, and dialogic. By advancing step-by-step through these three pathways, indigenous mathematical content can be transformed into usable teaching materials. The first is the emic pathway. It requires researchers to first understand local people's mathematical ideas from the perspective of insiders within the culture. They must not impose external standards to alter these ideas arbitrarily from the outset, but must preserve the meanings, practices, and classification methods originally recognized within the cultural context, and retain the original form of indigenous mathematics completely and accurately. Next is the etic perspective. Its task is to interpret these culturally rooted practices using external, formal mathematical classifications for academic analysis. It must not stop at merely understanding local usages, but must also align these indigenous practices with the general set of mathematics taught in schools, establish clear connections between indigenous cultural knowledge and academic mathematics, and enable the two sets of knowledge to communicate with each other. Finally, the dialogic pathway requires the emic and etic perspectives to collaborate and interact. It is not enough to simply piece the two sets of knowledge together; instead, indigenous cultural mathematical knowledge and formal academic mathematical knowledge must be refined and aligned with each other, and finally transformed into meaningful, interconnected mathematical expressions. At the same time, the respective origins of both must be preserved: the original intent of the indigenous culture is not distorted, and the norms of formal mathematics are also satisfied.

This method is not merely theoretical; it has been supported by empirical research that confirms its feasibility: (Umbara et al., 2021) verified that ethnomodelling can transform the mathematical ideas hidden in the calculation activities of Indonesia's Cigugur Indigenous people into academic mathematical concepts, including the specific contents taught in classrooms such as counting, integer operations, sets, relations, congruence, and modular arithmetic. Through this complete process of emic, etic, and dialogic pathways, ethnomodelling no longer stays at the level of "descriptively identifying mathematics within culture", but can systematically transform these contents into formal mathematical knowledge, and even organize them into materials for the classroom, smoothly integrating them into formal educational settings.

At the same time, global discussions surrounding the decolonization of mathematics education have further strengthened the core understanding that Indigenous knowledge holds epistemological legitimacy and pedagogical value. Decolonization efforts work to break the dominance of Eurocentric mathematical perspectives and integrate diverse ways of knowing into educational practice, and these ideas have received support from academic research (Garcia-Olp et al., 2022; Shultz et al., 2024). In multicultural and postcolonial contexts, this approach is especially well-suited, and it can advance inclusivity, equity, and cultural sustainability in education (Giroux & Nolan, 2025; Govender et al., 2026).

Within the local Indonesian context, practices that integrate Indigenous wisdom into education are supported by an independent curriculum that promotes contextualized, culturally responsive learning. Ethnomathematics provides a strategic pathway to achieve this policy goal, as it can connect mathematical concepts to students' real-life experiences (Nurhusain et al., 2025). Indonesia is home to diverse cultural resources, and past studies have unearthed mathematical meanings in architecture, food-related practices, and various cultural rituals (Hidayah et al., 2025; Pathuddin et al., 2021; Rahmawati et al., 2026; Supiyati et al., 2019). However, most of these studies have remained at the descriptive level, lacking a systematic framework to transform cultural knowledge into formal mathematics learning. Research that links cultural heritage sites to mathematics education through rigorous theoretical and methodological approaches remains very scarce.

The Passaungang Tau cultural heritage site offers a distinctive configuration of culturally embedded mathematical knowledge. Although traditional measurement systems based on the human body are also found in other cultural contexts, the specific practices identified at Passaungang Tau are embedded in the construction and use of a culturally significant stone-enclosed arena and are interconnected with local concepts of spatial balance, symmetry, and symbolic numeration. For comparison, the Nuabosi community in East Nusa Tenggara uses anthropometric units such as jengkal, hasta, depa, and footsteps in the construction of traditional houses, weaving, games, and customary activities (Bude et al., 2025). Similarly, traditional Minangkabau craftsmen have used dapu (fathom), eto (cubit), jangko (span), tampok (hand), and jari (finger) in constructing rumah gadang (Fitriza et al., 2019). These examples indicate that non-standard, body-based measurement is a recurring feature of indigenous mathematical practices, but they also demonstrate that such systems are culturally situated and function differently across communities. At Passaungang Tau, the use of depa and footsteps is embedded specifically in the spatial organization and construction of the stone-enclosed arena, where measurement practices are associated with dimensions, balance, and the ordering of the cultural space. The site therefore provides an opportunity to examine not merely the existence of non-standard measurement, but how measurement, spatial reasoning, and symbolic numeration operate as an interconnected system of situated mathematical knowledge.

Previous ethnomathematics research has already demonstrated the pedagogical potential of culturally embedded mathematical practices. For instance, Supiyati et al. (2019) identified mathematical concepts in Sasaknese architecture, while (Prahmana & D'Ambrosio, 2020) connected mathematical patterns in Yogyakarta batik with geometry learning and cultural values. More directly, Umbara et al. (2021) employed an ethnomodelling approach to transform the calculation practices of the Cigugur Indigenous community into formal mathematical concepts. These studies demonstrate important progress from cultural exploration toward educational application. However, research specifically examining the Passaungang Tau heritage site has not yet systematically connected its spatial geometry, traditional measurement practices, and culturally symbolic numeration within an integrated ethnomodelling framework. The present study therefore does not claim that the mathematical potential of cultural heritage has been previously unexplored; rather, it addresses a more specific gap by systematically interpreting the interconnected mathematical practices of Passaungang Tau and mapping them onto formal mathematical representations and potential classroom learning contexts.

Accordingly, this study centers on two core goals: exploring the mathematical meanings embedded in the site and systematically interpreting these meanings through an ethnomathematical modeling approach. This approach involves three core processes: (1) identifying the mathematical concepts inherent in the site's cultural practices, (2) interpreting these practices as mathematical models, and (3) exploring their potential integration into mathematics learning.

This study proposes an integrated ethnomathematical modeling framework, which transforms the system of local cultural practices in heritage sites into formal mathematical knowledge and instructional structures. This framework fills the long-standing gap between ethnomathematical explorations and their implementation in classrooms, and anchors the connection between learning and students' sociocultural identities. This study also advances research in related fields in two aspects: First, it overcomes the limitations of previous descriptive analyses, establishes a connection framework between cultural knowledge and formal mathematics, and expands the boundaries of ethnomathematical research. Second, it provides a foundation for integrating local knowledge into mathematics teaching, and supports culturally responsive inclusive practices that not only address local needs, but also contribute to global discussions on related topics.

2. RESEARCH METHOD

A. Research Design

This study adopts a qualitative research design rooted in ethnomathematics and ethnomodelling, with the aim of uncovering the mathematical connotations embedded in the pottery cultural heritage site of Passaungang Tau. Qualitative research was selected because this study focuses on interpreting the meanings of cultural practices, symbols, and spatial structures within their original context, rather than conducting quantitative statistics; supporting citations for this approach are drawn from Hirose & Creswell (2023) and Ishtiaq (2019). An ethnographic orientation was additionally adopted to conduct in-depth exploration of Indigenous knowledge systems and their associated mathematical structures, a choice supported by Almusaed et al. (2025). Ethnomodelling was used to connect emic, locally implicit knowledge with etic, formal mathematical representations, and the suitability of this approach is supported by findings of Umbara et al. (2021).

B. Participants

The study site for this research was the Passaungang Tau cultural heritage site, which was selected through purposive sampling. This site has a rich cultural foundation and holds research potential for uncovering embedded mathematical concepts. The study's participants were also selected via purposive sampling, to screen individuals familiar with the cultural practices of this heritage site. The participants included 1 traditional leader (KA1), 3 community members (M1, M2, M3), 2 teachers (G1, G2), and 1 student (S1). This sampling strategy allowed the inclusion of information-rich cases to deepen the research's analytical depth. The methodological support for this approach is drawn from study by Etikan et al. (2015).

C. Instrument

The core research instrument for this study was the researcher themselves, supported by three tools commonly used in qualitative research: an observation guide, a semi-structured interview outline, and a recording form. Participatory observation was used to document cultural practices, spatial layouts, and contextual interactions; interviews were conducted to elicit interviewees' interpretations of cultural symbols and related content; and analysis of documents and cultural relics relied on photos, sketches, and object inspections to identify embedded mathematical rules including geometric principles, symmetry, and measurement logic.

D. Procedures

To ensure the depth of this qualitative ethnographic study and the effectiveness of triangulation, this research adopted a phased multi-source data collection plan. First, preliminary observation was conducted to become familiar with the research site and its cultural context. Second, through participant observation, naturally occurring cultural practices and spatial structures were recorded. Third, semi-structured in-depth interviews were carried out with selected participants to obtain in-depth understanding of the cultural meanings and practices at play. Fourth, literature and artifacts were analyzed to verify the observational and interview data collected in the earlier steps. The entire process spanned a long period, to meet the fundamental requirements of this type of research for long-term engagement and contextual comprehension.

E. Ethnomodelling Framework and Data Analysis

This study conducted data analysis using an integrated ethnomodelling framework. This framework incorporates thematic analysis to transform systems of cultural practice into formal

mathematical representations, and the analysis in this study proceeded through three interconnected stages.

The first stage was the cultural exploration stage. We transcribed, organized, and coded three types of data collected from observations, interviews, and documentary records, and identified recurring themes related to cultural practices, spatial structures, symbolic meanings, and traditional measurement systems. This process followed the research protocol proposed by Braun & Clarke (2006). The second stage was the ethnomodelling analysis stage. We transformed four categories of ethnomathematical practices identified during fieldwork—spatial reasoning, geometric structures, measurement systems, and counting patterns—from an emic perspective into an etic mathematical framework. This translation method drew on the study by (Umbara et al. (2021), and through this process, we systematically generated formal mathematical models rooted in the local cultural context. The third and final stage was the instructional contextualization stage. We mapped the generated mathematical models onto potential classroom application scenarios, and developed contextualized mathematical tasks, conceptual learning trajectories, and instructional representations aligned with school mathematics curricula. This approach aligns with the view of culturally responsive contextualized mathematics learning, which emphasizes connecting mathematical concepts to learners' lived experiences. Supporting evidence for this view comes from study by Prahmana & D'Ambrosio (2020) and the study by Rosa & Orey (2021). This stage ensured that the outcomes of the ethnomodelling process had both theoretical significance and instructional applicability.

To ensure the credibility of this study, we adopted data source triangulation, long-term immersive participation, member checks for cultural validity verification, and reflexivity strategies throughout the entire research process. The theoretical basis for these methods draws on relevant studies by Lincoln et al. (2002), Nowell et al. (2017), and Tracy (2010).

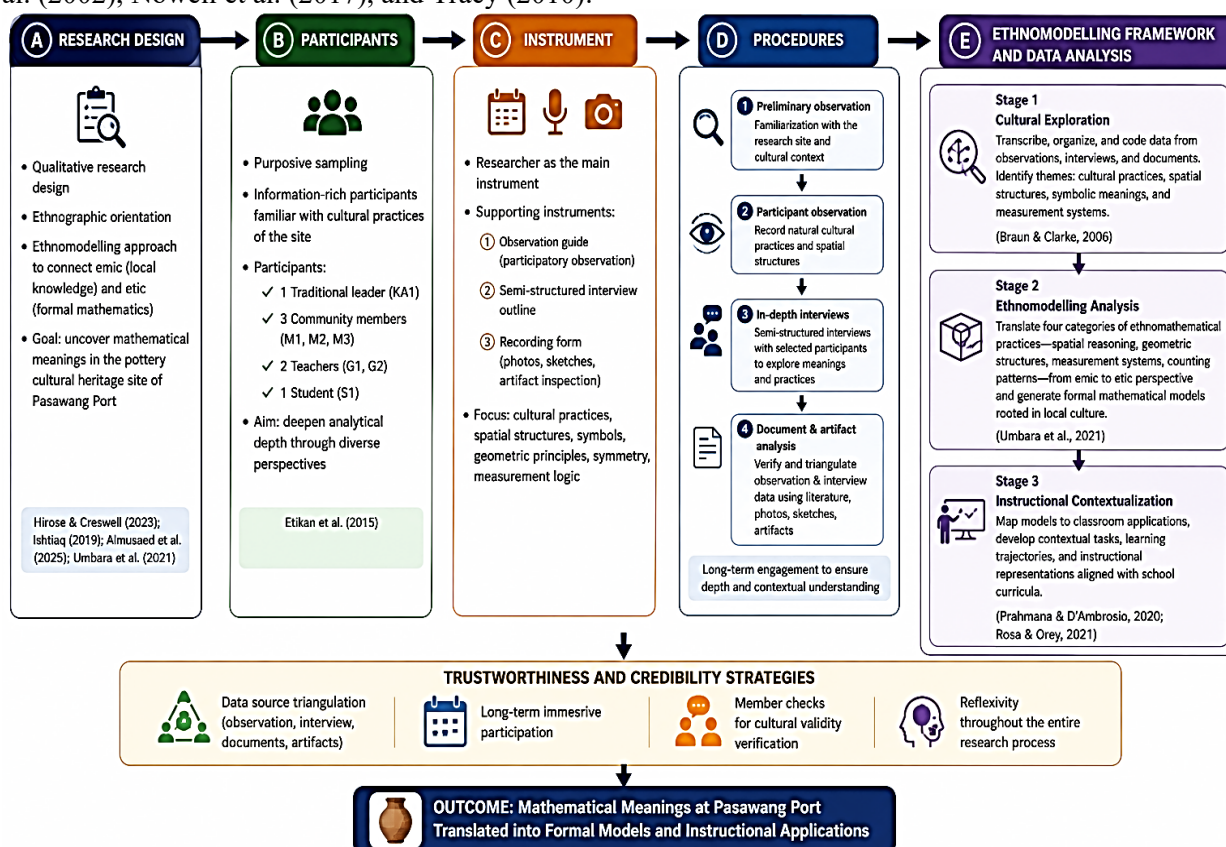


Fig 1. Research Method

3. RESULTS AND DISCUSSION

This study adopts thematic coding analysis to process three types of research data: interview data, observation data, and documentary data. Taking the Passaungang Tau site as its field research site for empirical investigation, the study identifies four core themes of ethnomathematics related to this site,

which are: geometric expressions in cultural architecture, the traditional measurement system as an indigenous mathematical practice, the counting system embedded in cultural symbols and social activities, and the approach to transforming cultural practices into formal mathematical concepts through ethnographic modeling. This study clarifies the connection between the cultural practices of the Gantarang Keke community and the construction of situated mathematical knowledge, and defines the potential of integrating such knowledge into mathematics teaching. Figure 1 presents the physical conditions of the site, which serves as the empirical background for this study.



Fig 2. Passaungang Tau Site in Gantarang Keke
 (Source: Research Documentation, 2025)

Figure 1 shows the physical conditions of the Passaungang Tau archaeological site. As a cultural space, this site served as the carrier for the development of mathematical practices among the local community.

A. Theme 1: Geometry as a Representation of Cultural Structure

The physical structure of the site contains implicit geometric concepts. Based on field observations and interview results, the local community perceives the basic form of this site as an enclosed space that resembles a cuboid with specific dimensions.

Interviewee M2 explained that the site is like a large box with its own length, width, and depth. This indicates that in the local perception, it is not merely a flat area, but a three-dimensional space: “Kalau dilihat dari bentuknya, ini seperti kotak besar... ada panjang, lebar, dan dalamnya juga. Jadi bukan hanya datar, tapi seperti ruang.” (Informant M2: “Just by its shape, this place is like a big crate... it has length, width, and depth. So it is not just a flat patch of land, but a real, tangible space.”).

This statement shows that the local community had an intuitive understanding of three-dimensional concepts, but this understanding was never expressed in formal mathematical terms. This finding is also supported by the account of a local traditional leader (interview code KA1), who explained that when the site was built, the requirement for balance was fully considered: the left and right sides must be completely equal to ensure the structure was solid and orderly. He stated, “Dulu dibuat dengan memperhatikan keseimbangan... kiri dan kanan harus sama, supaya kuat dan rapi.” (Informant KA1: “When we built things in the past, we always paid attention to balance... the left and right had to be the same, so that the structure would be firm and neat.”). Table 1 organizes all implicit geometric concepts identified at the site, including cuboid shapes, symmetry, and the spatial dimensions related to area and volume.

Table 1. Identification of Geometric Concepts in Passaungang Tau

Cultural Aspect	Description	Mathematical Concept
Site structure	Bounded arena	Three-dimensional shape (cuboid)
Structure	Flat and balanced	Symmetry
Dimensions	Length, width, and depth	Volume and area

These findings confirm that geometric principles are in fact embedded in the local community's cultural construction practices. The concept of balance corresponds to the mathematical concepts of symmetry and proportion, while the orderly spatial layout reflects the community's capacity for intuitive geometric reasoning. The geometric understanding of the Passaungang Tau community is entirely derived from culturally embedded spatial practices, and the ideas of balance and order are put into practice as part of their daily life traditions.

This paper argues that the geometric understanding of the Passaungang Tau people is a product of local cultural practices, rather than the outcome of formal mathematics education. A study by Prahmana et al. (2021) points out that mathematical knowledge does not possess inherent symbolic properties; instead, it is embedded in cultural experiences. Combined with this ethnic group's embodied spatial reasoning about balance and symmetry, Mania & Alam (2021) proposed that geometry is a culturally constructed cognitive system. The geometry discussed here is precisely a structurally and value-laden construct that carries the social order of harmony.

B. Theme 2: Traditional Measurement as a Local Mathematical System

During the construction of Passaungang Tau, the local community used non-standard measurement systems such as the arm span (*depa*) and stride length. These systems are not merely practical measurement tools; they also embody the community's understanding of space and distance. Informant M3 explained that in the past, people did not use meters to measure, but instead used strides or arm spans, and would adjust these measurements according to actual needs: "*Dulu tidak pakai meter, ukurannya pakai langkah atau depa... disesuaikan saja dengan kebutuhan.*" (Informant M3: "Back then, they didn't use meters; measurements were based on steps or arm spans, depending on what was needed.").

Informant G2, who works as a teacher, believes this system has educational potential. It can be used as an introductory scenario to let students compare traditional measurement units with standard units such as the meter, "*Ini bisa dijadikan awal pembelajaran. Siswa bisa membandingkan ukuran tradisional dengan meter.*" (Informant G2: "This can serve as a starting point for learning. Students can compare traditional units with meters.").

The links between traditional units and mathematical concepts have been systematically sorted out in Table 2, which presents the forms of various measurement units, their corresponding descriptions, and the derivable mathematical transformation pathways.

Table 2. Traditional Measurement System

Traditional Unit	Description	Mathematical Transformation
<i>Depa</i>	Arm span	Unit of length
Footsteps	Walking distance	Length estimation
Approximation	Non-standard measurement	Ratio and comparison

These research findings show that the traditional measurement system contains implicit mathematical reasoning such as thinking related to estimation, comparison, and proportion. From an ethnomodelling perspective, these practices can be systematically transformed into formal mathematical concepts, building a bridge between indigenous cultural knowledge and school mathematics. Traditional measurement practices also indicate that mathematical reasoning is rooted in experience and functional needs. Under the ethnomodelling framework, these practices belong to emic mathematical structures that can be transformed into formal measurement concepts. This transformation is not a simple technical adjustment, but an epistemological reconstruction. Umbara et al. (2021) previously proposed the core logic of this transformation pathway, and Roza et al. (2020) had earlier pointed out the teaching potential of traditional measurement. This study advances relevant discussions into a implementable plan for the structured transformation of teaching content.

C. Theme 3: Numeration in Socio-Cultural Values and Structures

The concept of numbers within Passaungang Tau appears not only in quantitative forms, but also carries strong symbolic meanings within the community's social structure. Numbers such as 7, 14, and 15 possess particular significance in customary traditions.

KA1, a traditional community leader, explained that the consequences of violating customary rules can affect up to "14 generations," a detail that demonstrates the severity of sanctions within the

traditional system. He stated in Indonesian: “*Kalau ada yang melanggar aturan, dampaknya bisa sampai 14 turunan... itu aturan yang sangat berat.*” Another community member, M1, noted that 7 and 15 frequently appear in various customary activities, and local people believe these numbers carry unique symbolic implications. He stated in Indonesian: “*Angka 7 dan 15 sering dipakai dalam adat... dianggap punya makna tertentu.*” The relationship between symbolic meanings and the mathematical functions of numbers in customary practices is systematically summarized in Table 3.

Table 3. Numerical Concepts in Culture

Number	Cultural Meaning	Mathematical Concept
7	Important customary number	Number
14	Generational sanction	Sequence / multiple
15	Cultural symbolism	Numeration

The findings of this study show that counting systems have both quantitative and sociocultural attributes. The expression of “14 generations” is a symbolic form of quantification used to reinforce social norms and authority. Meanwhile, the repeated appearance of the numbers 7 and 15 reflects a fixed counting pattern formed in cultural practices. From an ethnomathematics perspective, this is exactly cultural counting that carries dual meanings of mathematics and philosophy. The use of counting in sociocultural contexts highlights the semiotic and normative dimensions of mathematics. As carriers of cultural meaning, the three numbers 7, 14, and 15 not only regulate social behavior but also encode core philosophical values. This finding challenges the reductionist view that counting is only a purely quantitative tool, and instead positions it as a symbolic system embedded in cultural discourse. Relevant supporting evidence can be found in the studies by Boaler (2019) as well as Pathuddin & Raehana (2019). In this sense, counting is a construct that operates at both the cognitive and sociocultural levels, which expands its role in the field of mathematics education.

Through ethnomodelling, these patterns can be converted into formal mathematical concepts such as digital patterns, relational logic, and situational counting systems for use in classroom teaching. On the whole, the cultural practices of Passaungang Tau themselves contain structured mathematical ideas that can be systematically transformed into formal mathematical content.

D. Theme 4: Ethnomodeling and Pedagogical Contextualization

The study finds that Passaungang Tau has sufficient potential to serve as a situational resource for mathematics learning, and both teachers and students have given positive feedback on integrating culture into mathematics teaching.

A participating student coded S1 stated that learning based on this site makes mathematical concepts easier to understand, because the related forms can be observed directly: “*Kalau belajar dari tempat ini, lebih mudah dipahami... karena bisa langsung lihat bentuknya.*” (“If you learn from this place, it is easier to understand... because you can see the forms directly.”).

Similarly, Teacher G1 explained that when mathematics teaching breaks free from the limitations of numbers written on blackboards and is integrated with historical and cultural contexts, it becomes more meaningful and engaging: “*Matematika jadi lebih hidup... tidak hanya angka di papan, tapi ada hubungannya dengan sejarah.*” (“Mathematics becomes more vivid... it is no longer just numbers on a blackboard, but also connected to history.”).

These findings show that learning rooted in real-world contexts can improve students' conceptual understanding by supporting concrete visualization and promoting the abstraction of mathematical ideas. From an ethnomathematics perspective, this is exactly the process through which cultural practices are transformed into formal mathematical concepts via ethnomodelling, which confirms that mathematics is inherently embedded in sociocultural life.

Figure 2 presents the integrated ethnomodelling framework developed in this study. It illustrates the complete pathway through which cultural practices are transformed into formal mathematical concepts through exploration and modelling, and then integrated into teaching scenarios. The framework clarifies that ethnomodelling serves as a bridge connecting local indigenous knowledge to school mathematics.



Fig 3. Integrative Ethnomodeling Framework for Culturally Responsive Mathematics Education

This study organizes and presents the possibilities of integrating the above findings into mathematics learning, specifically using a table to show how cultural contexts map onto mathematical concepts and classroom activities.

Table 4. Potential Integration of Ethnomodelling in Mathematics Learning

Cultural Context	Mathematical Concepts	Learning Potential
Structure of the Passaungang Tau site	Spatial geometry, symmetry, and spatial reasoning	Identification of three-dimensional shapes and symmetry analysis
Measurement using <i>depa</i> and footsteps	Length units, estimation, ratios, and unit conversion	Comparison between non-standard and standard units
Symbolic numbers in customary traditions	Numeracy, number patterns, and numerical relations	Learning number patterns through cultural contexts

Overall study findings show that ethnomodelling not only supports learners' conceptual understanding, but also helps protect indigenous cultural knowledge in educational settings. From a broader perspective, the ethnomodelling used in this study acts as an epistemological bridge that connects indigenous knowledge systems and formal mathematics. It transforms cultural practice systems into structured mathematical representations, breaking the dominant narrative that mathematics is a universal, culturally neutral discipline. It highlights that mathematics is socially embedded knowledge shaped by historical and cultural contexts. This perspective advances discussions on the decolonization of mathematics education, recognizing indigenous knowledge as an equally valid cognitive resource (Garcia-Olp et al., 2022; Shultz et al., 2024).

It is important to highlight that integrating cultural contexts into mathematics learning can improve conceptual understanding when abstract ideas are grounded in learners' familiar sociocultural experiences. This aligns with culturally responsive and culturally sustaining pedagogical theories, which frame culture as a core component of knowledge construction rather than an add-on. Furthermore, ethnomodelling not only makes methodological contributions, but also can reshape the relationship between mathematics, culture, and learning.

E. Preliminary Pedagogical Exploration

To test the instructional applicability of the ethnomodelling framework, we carried out a small-scale learning activity for 8 middle school students at the Passaungang Tau Site. During the activity, students were asked to identify geometric shapes, use the traditional unit of measurement *depa*, and connect cultural practices with formal mathematical concepts. We also collected their feedback through reflective interviews.

The results of this study show that students have a high level of active participation, and are able to connect mathematical concepts with culturally meaningful representations. When concepts related to geometry and measurement are integrated with familiar cultural practices and the physical local environment, the difficulty of understanding these concepts decreases significantly. Students also hold positive attitudes toward the integration of local culture into mathematics learning, believing that this approach not only improves their conceptual understanding but also enhances their cultural awareness. As a preliminary exploration with limited scope, this study did not include quantitative validity measurements. The findings indicate that learning based on ethnomodelling can support culturally responsive teaching, and integrating mathematical concepts with students' sociocultural context can strengthen their participation. Table 5 summarizes students' feedback and their learning engagement.

Table 5. Students' Feedback and Learning Engagement in Preliminary Ethnomodelling Learning Activities

Activity	Student Response
Geometry identification	Students identified triangles, rectangles, and spatial patterns
Depa measurement	Students compared traditional and formal units
Reflection discussion	Students reported higher engagement and contextual understanding

The research findings show that integrating cultural heritage into mathematics learning can build a connection between abstract mathematical concepts and concrete life experiences; this approach not only supports learners' conceptual understanding, but also enhances their cultural perception and learning motivation. This study is exploratory, and its results provide preliminary empirical support for the teaching potential of mathematics learning rooted in local culture.

F. Pedagogical Implications and Global Relevance

The findings of this study have important reference value for mathematics teaching design and curriculum development. The mathematical elements identified in the Passaungang Tau cultural heritage site in this study can be transformed into learning materials, classroom activities, and situational assessment tasks. Developing teaching content based on cultural heritage allows students to access mathematics that aligns with their local context, which is consistent with the educational direction that emphasizes situational learning and the integration of local wisdom (Marshall, 2023; Nurhusain et al., 2025).

The small-scale teaching practice carried out in this study shows that cultural context is an effective entry point for introducing formal mathematical concepts. Cultural products and practices are by no means merely supplementary cases; instead, they can be systematically integrated into curriculum design to support students' exploration, discussion, and knowledge construction, while also connecting mathematical content with students' daily realities (Gresalfi & Hand, 2019; Kong et al., 2023; Marshall, 2023; Nolan & Xenofontos, 2023; Rosa & Orey, 2021). The positive outcomes of this practice have verified the suitability of the ethnomathematical modeling learning strategy, and all relevant conclusions are supported by existing research literature (Prahmana & D'Ambrosio, 2020; Supiyati et al., 2019).

Beyond its direct educational applications, this study also maps out directions for future research and innovation. By exploring the ethnomodelling framework built around Passaungang Tau, our work can be adapted for research on other cultural heritage scenarios, and can also support the development of curriculum resources, instructional media, and teacher professional development programs. Furthermore, this framework can serve as a research foundation to analyze how culturally rooted learning designs contribute to broader educational goals, including boosting student motivation and self-efficacy, and promoting formal education's recognition of diverse knowledge traditions. Relevant supporting evidence for this line of work can be found in a series of prior studies. This study also notes that collaboration between educational institutions and local cultural communities is a core condition for sustaining the applicability of culturally rooted mathematics education. The recording and interpretation of mathematical concepts embedded in cultural heritage requires the joint participation of community knowledge holders, educators, and researchers, to ensure the accurate presentation of local knowledge and the preservation of its cultural significance (Disnawati & Nahak, 2019; Garcia-Olp et al., 2022; Meaney & Trinick, 2020; Rosa & Orey, 2021; Shultz et al., 2024).

The findings of this study also indicate that greater attention must be paid to teacher preparation and professional development focused on culturally adaptive mathematics instruction. Most culture-rooted mathematical concepts fail to be implemented in classrooms, and the core reason for this is that teachers lack opportunities to identify, interpret, and integrate these concepts into formal learning settings. Strengthening teachers' ability to connect with local knowledge systems can support more diverse mathematical inquiries and classroom dialogues.

From a broader perspective, this study underscores the importance of expanding the knowledge base that shapes the content and perspectives drawn on in mathematics education. Treating cultural practices as a legitimate source of mathematical insights can foster more inclusive forms of cognition, support ongoing efforts to expand the participation of diverse stakeholders in mathematics education, and cultivate greater recognition of the knowledge contributions of local and Indigenous communities.

G. Limitations and Suggestions for Future Research

This study has several limitations. First, the research was conducted only within a single cultural context, which limits the generalizability of its conclusions. Second, the analysis relies on ethnographic interpretation; while every effort was made to ensure rigor, the potential for subjectivity remains. In addition, the teaching exploration was small in scale, exploratory in nature, and included only a small number of students, with no quantitative assessment of learning outcomes. This study also raises an

important epistemological concern: converting Indigenous cultural practices into formal mathematical representations through cultural modeling may simplify the complex cultural connotations of those practices, or reframe their meanings through an academic lens. It may also romanticize Indigenous knowledge as static and universally applicable, even though cultural practices are inherently dynamic and context-dependent.

Future research needs to integrate multicultural contexts and carry out larger-scale classroom-based practical implementations to strengthen the reliability of empirical evidence. Further exploration is also required on how to balance the three core dimensions of cultural authenticity, epistemological integrity, and pedagogical transformation in ethnomathematics-based teaching.

4. CONCLUSION

This study aimed to fill the research gap between the identification of ethnomathematics and its pedagogical transformation. It investigated the mathematical meanings embedded within the Passaungang Tau cultural heritage site, and systematically interpreted these meanings through an ethnomodelling framework, which aligns with the objectives proposed in the introduction. The study found that mathematical concepts are deeply embedded in cultural practices, manifesting in three forms: geometric structures, traditional measurement systems, and symbolic counting. These cultural elements have been converted into formal mathematical representations and learning-oriented models, confirming that the study's objectives, methodological design, and empirical results are highly consistent with one another.

This study confirms that the local context of the Passaungang Tau in Gantarang Keke community has nurtured a socioculturally embedded system of mathematical knowledge. Within this system, geometric logic is rooted in indigenous spatial structures and concepts of balance; measurement relies on an embodied, non-standard system; and counting carries symbolic cultural meanings. When this knowledge is transformed into school mathematics content through ethnographic modeling, it can improve students' level of engagement, comprehension, and cultural awareness. This finding verifies that cultural practices are a legitimate epistemological source for formal mathematics education.

This study verified that the conceptual expectations laid out in the introduction are fully consistent with the empirical findings presented in the discussion section. The ethnomodelling framework proposed in this study not only connects ethnomathematics exploration with classroom practice, but also provides a transferable model for culturally responsive mathematics education. Future research can promote this framework in multicultural settings, carry out large-scale teaching experiments and quantitative evaluations, and meanwhile optimize the balance between cultural interpretation and mathematical formalization, to give simultaneous consideration to cultural authenticity and teaching applicability.

ACKNOWLEDGEMENTS

The authors of this study would like to express our sincere gratitude to the Ujung Pandang Education Foundation for providing financial support that enabled the smooth completion of this project. We also extend our thanks to the leadership of STKIP YPUP Makassar, the university's Institute for Research and Community Service (Lembaga Penelitian dan Pengabdian kepada Masyarakat, LPPM), and the Mathematics Education program, for providing administrative support and assistance throughout the entire research process. We deeply thank all research participants who took time out of their schedules to take part in interviews and share their valuable insights. Special thanks go to the peer reviewers who offered constructive feedback on the initial draft of the manuscript, as well as the internal reviewers who helped refine the study's focus and the interpretation of its findings. The authors take full responsibility for any shortcomings that remain in this paper.

REFERENCES

- Almusaed, A., Almssad, A., & Yitmen, I. (2025). Ethnography and Case Study Methods. In A. Almusaed, A. Almssad, & I. Yitmen (Eds.), *Practice of Research Methodology in Civil Engineering and Architecture: A Comprehensive Guide* (pp. 317–353). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97393-2_10

- Bishop, A. J. (1991). *Mathematical Enculturation: A Cultural Perspective on Mathematics Education*. (Vol. 6). Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-009-2657-8>
- Boaler, J. (2019). Developing Mathematical Mindsets: The Need to Interact with Numbers Flexibly and Conceptually. *American Educator*, 42(4), 28–33. <https://eric.ed.gov/?id=EJ1200568>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bude, M. K., Ude, M. A., Tukan, G. A., Wea, Y. T., Mei, M. F., & Wondo, M. T. S. (2025). Eksplorasi Etnomatematika Aktivitas Pengukuran padaa Desa Nuabosi. *Journal of Mathematics Education*, 4(2), 542–553.
- D'Ambrosio, U. (1985). Ethnomathematics and its Place in the History and Pedagogy of Mathematics. *For the Learning of Mathematics*, 5(1), 44–48. <http://www.jstor.org/stable/40247876>
- D'Ambrosio, U., & Rosa, M. (2017). Ethnomathematics and Its Pedagogical Action in Mathematics Education. In M. Rosa, L. Shirley, M. E. Gavarrete, & W. V Alangui (Eds.), *Ethnomathematics and its Diverse Approaches for Mathematics Education* (pp. 285–305). Springer International Publishing. https://doi.org/10.1007/978-3-319-59220-6_12
- Disnawati, H., & Nahak, S. (2019). Pengembangan Lembar Kerja Siswa Berbasis Etnomatematika Tenun Timor pada Materi Pola Bilangan. *Jurnal Elemen*, 5(1), 64–79. <https://doi.org/10.29408/jel.v5i1.1022>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2015). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fitriza, R., Turmudi, T., Juandi, D., & Harisman, Y. (2019). Traditional measurement units: A study on the construction of rumah gadang of Minangkabau. *Journal of Physics: Conference Series*, 1157(4). <https://doi.org/10.1088/1742-6596/1157/4/042123>
- Foote, M. Q., & Gau Bartell, T. (2011). Pathways to equity in mathematics education: how life experiences impact researcher positionality. *Educational Studies in Mathematics*, 78(1), 45–68. <https://doi.org/10.1007/s10649-011-9309-2>
- Garcia-Olp, M., Nelson, C., & Saiz, L. (2022). Decolonizing Mathematics Curriculum and Pedagogy: Indigenous Knowledge Has Always Been Mathematics Education. *Educational Studies*, 58(1), 1–16. <https://doi.org/10.1080/00131946.2021.2010079>
- Giroux, M. A. G., & Nolan, K. T. (2025). Respect, Responsibility, Relevance, and Reciprocity: What the 4 Rs of Indigenous Research Offer Toward Decolonizing a Mathematics Classroom. *Education Sciences*, 15(12). <https://doi.org/10.3390/educsci15121579>
- Govender, N., Moheeput, K., Singh-Pillay, A., Cyparsade, M., James, A., Mabaso, B., Ngema, S., & Kolobe, L. (2026). Cultural decolonization and implications for pedagogy in integrating STEM-IKS education in the Global South. *Cultural Studies of Science Education*, 21(1), 103–127. <https://doi.org/10.1007/s11422-025-10270-6>
- Gresalfi, M., & Hand, V. M. (2019). Coordinating situated identities in mathematics classrooms with sociohistorical narratives: a consideration for design. *ZDM*, 51(3), 493–504. <https://doi.org/10.1007/s11858-019-01034-y>
- Hidayah, S. K., Lutfiyah, L., & Marsidi, M. (2025). Ethnomathematical Exploration of the Wiwitan Methik Pari Ritual Culture. *ETDC: Indonesian Journal of Research and Educational Review*, 4(4), 1157–1167. <https://doi.org/10.51574/ijrer.v4i4.3709>
- Hirose, M., & Creswell, J. W. (2023). Applying Core Quality Criteria of Mixed Methods Research to an Empirical Study. *Journal of Mixed Methods Research*, 17(1), 12–28. <https://doi.org/10.1177/15586898221086346>
- Ishtiaq, M. (2019). Book Review Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (4th ed.). Thousand Oaks, CA: Sage. *English Language Teaching*, 12(5), 40. <https://doi.org/10.5539/elt.v12n5p40>
- Kong, J. E., Arizmendi, G. D., & Doabler, C. T. (2023). Implementing the Science of Math in a Culturally Sustainable Framework for Students With and at Risk for Math Learning Disabilities. *TEACHING Exceptional Children*, 56(1), 44–51. <https://doi.org/10.1177/00400599221127385>
- Lincoln, Y. S., Guba, E. G., & Pilotta, J. J. (2002). *Naturalistic inquiry*: Beverly Hills, CA: Sage Publications, 1985, 416 pp. *International Journal of Intercultural Relations*, 9(4), 438–439. [https://doi.org/10.1016/0147-1767\(85\)90062-8](https://doi.org/10.1016/0147-1767(85)90062-8)

- Mania, S., & Alam, S. (2021). Teachers' Perception toward the Use of Ethnomathematics Approach in Teaching Math. *International Journal of Education in Mathematics, Science and Technology*, 9(2), 282–298. <https://doi.org/10.46328/ijemst.1551>
- Marshall, S. (2023). But What Does it Look Like in Maths? A Framework for Culturally Sustaining Pedagogy in Mathematics. *International Journal of Multicultural Education*, 25(1), 1–29. <https://doi.org/10.18251/ijme.v25i1.3251>
- Meaney, T., & Trinick, T. (2020). Indigenous Students in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 369–373). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_76
- Nolan, K., & Xenofontos, C. (2023). Mapping teachers' perspectives on culturally responsive pedagogy in mathematics: from academic achievement to insights and opportunities. *Intercultural Education*, 34(6), 550–567. <https://doi.org/10.1080/14675986.2023.2265855>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1). <https://doi.org/10.1177/1609406917733847>
- Nurhusain, M., Upu, H., & Djadir. (2025). Ethnomathematics-Based Learning & the Merdeka Curriculum in the Society 5.0 Era: Indonesian Educational Innovation. *International Journal of Integrative Sciences (IJIS)*. 4(4), 743–766. <https://doi.org/10.55927/ijis.v4i4.203>
- Pathuddin, H., Kamariah, & Ichsan Nawawi, M. (2021). Buginese Ethnomathematics: Barongko Cake Explorations as Mathematics Learning Resources. *Journal on Mathematics Education*, 12(2), 295–312. <https://eric.ed.gov/?id=EJ1313711>
- Pathuddin, H., & Raehana, S. (2019). Etnomatematika: Makanan Tradisional Bugis Sebagai Sumber Belajar Matematika. *MaPan*, 7(2), 307–327. <https://doi.org/10.24252/mapan.2019v7n2a10>
- Prahmana, R. C. I., & D'Ambrosio, U. (2020). Learning Geometry and Values From Patterns: Ethnomathematics on the Batik Patterns of Yogyakarta, Indonesia. *Journal on Mathematics Education*, 11(3), 439–456. <https://eric.ed.gov/?id=EJ1294663>
- Prahmana, R. C. I., Yunianto, W., Rosa, M., & Orey, D. C. (2021). Ethnomathematics: Pranatamangsa System and The Birth-Death Ceremonial In Yogyakarta. *Journal on Mathematics Education*, 12(1), 93–112. <https://eric.ed.gov/?id=EJ1294495>
- Rahmawati, F. I., Hastuti, I. D., Mariyati, Y., & Syaharuddin. (2026). Ethnomathematics: Exploring Geometric Shapes and Patterns in Sasak Snacks. *IJEMS Indonesian Journal of Education and Mathematical Science*, 7(1), 99–109. <https://doi.org/10.30596/ijems.v7i1.29298>
- Rosa, M., & Orey, D. C. (2021). An Ethnomathematical Perspective of Stem Education in a Glocalized World. *Bolema - Mathematics Education Bulletin*, 35(70), 840–876. <https://doi.org/10.1590/1980-4415v35n70a14>
- Roza, Y., Siregar, S. N., & Solfitri, T. (2020). Ethnomathematics: Design Mathematics Learning at Secondary Schools by Using The Traditional Game of Melayu Riau. *Journal of Physics: Conference Series*, 1470(1), 1–14. <https://doi.org/10.1088/1742-6596/1470/1/012051>
- Shultz, M., Close, E., Nissen, J., & Van Dusen, B. (2024). Enacting Culturally Relevant Pedagogy when “Mathematics Has No Color”: Epistemological Contradictions. *International Journal of Research in Undergraduate Mathematics Education*, 10(2), 486–515. <https://doi.org/10.1007/s40753-023-00219-x>
- Supiyati, S., Hanum, F., & Jailani. (2019). Ethnomathematics in Sasaknese Architecture. *Journal on Mathematics Education*, 10(1), 47–57. <https://eric.ed.gov/?id=EJ1204853>
- Tracy, Sarah J. (2010). Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, 16(10), 837–851. <https://doi.org/10.1177/1077800410383121>
- Umbara, U., Wahyudin, W., & Prabawanto, S. (2021). Exploring Ethnomathematics with Ethnomodeling Methodological Approach: How Does Cigugur Indigenous People Using Calculations to Determine Good Day to Build Houses. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(2), 1–19. <https://doi.org/10.29333/EJMSTE/9673>
- Valero, P. (2020). Political Perspectives in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 663–666). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_126

The Muffled Echoes of Reflection: Mapping the Fading Resonance of Mathematics Reflective Thinking Among Anxious Students

Muntazhimah¹, Indah Widiati², Joko Dwi Prasetyo³

¹Mathematics Education Study Program, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia

²Mathematics Education Study Program, Universitas Islam Riau, Riau, Indonesia

³Master of Mathematics Education, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta Selatan, Indonesia

¹muntazhimah@uhamka.ac.id, ²indahwidiati@uuir.ac.id, ³prasetyo.jodwi@gmail.com

ABSTRACT

Mathematical reflective thinking allows students to examine and reconstruct the steps they take when solving problems. Yet psychological factors, especially mathematics anxiety, can disrupt this mental process by causing cognitive interference. This study describes and maps how high school students in Jakarta engage in reflective thinking while solving contextual trigonometric problems, considering three levels of mathematics anxiety: high, medium, and low. The research used a descriptive qualitative approach and involved six students chosen through purposive sampling. Data came from a mathematics anxiety scale, a reflective thinking ability test, and in-depth interviews conducted with the think-aloud technique. Data analysis used a thematic approach grounded in Surbeck's three phases of reflective thinking: Reacting, Comparing, and Contemplating. Among students with high anxiety, the findings point to a pattern of "fading resonance": the voice of logic becomes muffled as reflective thinking progressively weakens, particularly when students must sustain deeper evaluation under increased cognitive demands. Low-anxiety students, by contrast, show complete and systematic reflective resonance across all three stages. The results indicate that students' psychological characteristics are a primary factor distinguishing the quality of their mathematical reflection, which calls for instructional strategies that can reduce emotional barriers.

Keyword : Mathematical Reflective Thinking, Mathematics Anxiety, Trigonometry, High School Students



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Muntazhimah,

Mathematics Education Study Program,

Universitas Muhammadiyah Prof. Dr. Hamka,

Jl. Tanah Merdeka No.20, RT.11/RW.2, Rambutan, Ciracas, Kota Jakarta Timur, Indonesia.

muntazhimah@uhamka.ac.id

1. INTRODUCTION

At its foundation, mathematics is a systematic discipline that develops logical, analytical, and critical thinking (Damastuti et al., 2023; Isran et al., 2025). Problem-solving therefore stands at the center of the learning process (Charolina, 2025; Munawwarah et al., 2024). In mathematics education, though, the emphasis has increasingly moved beyond procedural fluency and memorization toward higher-order thinking skills (*HOTS*) (Gradini et al., 2025; Muntazhimah et al., 2022). Mathematical reflective thinking is one part of *HOTS*. It involves a structured cognitive process in which students intentionally examine, assess, and reconstruct their reasoning, connecting what they already know with unfamiliar challenges (Sari et al., 2025; Setiyani et al., 2025). Through this kind of reflection, students can approach complex problems with greater focus and reduce misconceptions (Muntazhimah et al., 2022).

Despite its importance, reflective thinking is often neglected in education because so much attention is placed on final answers (Setiyani et al., 2024). As a result, students may lack the ability to assess the steps they take when solving problems (Mugniarningsih et al., 2022; Situmorang, 2025). The results of the Programme for International Student Assessment (PISA) 2022 in Indonesia, for example, show that a significant majority of students continue to struggle with even simple mathematical problems requiring them to apply concepts to real-world situations (Masito et al., 2024; Putriani et al., 2025). Many students find mathematics burdensome and confusing, seeing it as a difficult subject dominated by abstract formulas (Budhathoki et al., 2022; Cañeda et al., 2024). The difficulty becomes especially apparent in complex areas such as trigonometry, often perceived as a "labyrinth" of formulas that students struggle to apply in different formats (Latibjonovich, 2022; Reza et al., 2026).

Numerous earlier research has evidenced that students not only faced challenges memorizing and implementing the formulas in trigonometry, but also the conceptual interpretation of mathematical concepts and flexible reasoning. For example, studies concluded that students could be hindered by challenges relating to understanding the information provided in trigonometric questions, angle

comparison, and the appropriate trigonometry applied to them for completing a question (Nurmeidina & Rafidiyah, 2019). In a similar study of students' understanding of ratio and trigonometric ratio, the difficulty in visual conceptualisation of angles can restrict students' flexibility to work with trig relations, which suggests students were more likely to possess an instrumental understanding than relational understanding (Yiğit Koyunkaya, 2016). In other contemporary studies, it was noted that high school students faced difficulties in truly grasping concepts and ideas behind some of the contextual ideas, such as an angle of elevation, which thus led to students facing difficulties in identifying proper procedures to be used in solving the questions (Obeng et al., 2024). However, such problems are crucially meaningful in the light of Mathematical Reflective Thinking since, to solve non-routine trigonometric problems, students not only need to recognize the useful information, but also to make comparisons of potential methods, appraise the correctness of reasoning, and reexamine answers. Previous research has demonstrated a positive relationship between students' reflective thinking in problem solving and their mathematical achievement, as well as between reflective thinking and mathematical problem-solving ability (Syadid & Sutiarto, 2021; Toraman et al., 2020). Therefore, difficulties in interpreting and evaluating trigonometric information may constrain the reflective processes necessary for successful mathematical problem-solving and achievement. Much less research has been conducted on how these reflective processes operate at varying levels of mathematics anxiety, especially when students are working on contextual trigonometry problems. This gap highlights the importance of investigating the nature of the processes, in addition to the correctness of the solutions, that an affective variable like mathematics anxiety might be associated with.

Effectiveness of the reflective thinking process does not reflect a mere mental process but encompasses various psychological filters (Gómez et al., 2025; Ugli, 2025). Amongst the many emotional filters, mathematics anxiety is a psychological filter that frequently hinders mathematical progress (Joy et al., 2025). In accordance with the cognitive interference theory, anxiety symptoms produce negative thoughts such as worry and self-evaluative rumination and consume the limited working memory capacity available for the mathematical problem (Brockbank & Feldon, 2024; Cécillon et al., 2024). Resultant emotional dysregulation produces a 'mental block' where students' logical 'voice' or resonance is muffled, and inhibits their ability to initiate or sustain a reflective cycle of thinking (J. Liu et al., 2026). While consciousness of cognitive dissonance or 'confusion' can facilitate deeper reflection, high anxiety produces confusion as a form of mathematical incapacity and thwarted reflective thinking, resulting in avoidance or premature surrender (Joy et al., 2025; J. Liu et al., 2026).

The relationship between psychological factors such as the Adversity Quotient (AQ) or resilience and mathematics achievement has been long explored in the extant literature (Ginting, 2024; Mustafa & Andriani Yunus, 2025; Nurdin et al., 2026; Peng et al., 2025). Literature findings indicate that resilient high or climber students consistently yield an excellent cycle of reflective thinking, compared to resilient low or climber students who conceptually have a mediocre cycle, or non-resilient, quitter students who have an absent cycle of reflective thinking. However, most of these studies utilize quantitative path analyses or broad behavioral categorizations. There is a lack of qualitative research that specifically "maps" the internal dynamics and thematic "fading" of reflective resonance in high-pressure urban environments like Jakarta, particularly within the abstract domain of trigonometry.

This is a novel contribution of this study in that it is not only about the scores, but the mapping of the flow of reflective thinking in the context of anxiety for the whole spectrum. It is hoped that by using the reacting, comparing, and contemplating stages of reflective thinking, this study could pinpoint the exact moment when the logic resonance begins to fade to the anxious student. The innovation lies in reconstructing the "muffled echoes" of student responses to identify in what stages of mathematical reflection anxiety kills certain aspects of mathematical reflection. It is expected that this study will provide a general map of tenth-graders' reflective thinking process and to serve as a reference for Jakarta teachers on the steps to develop a diagnostic intervention, which not only leads to cognitive remediation but also provides psychological safety. Therefore, the aim of this study is to investigate and describe the dynamics of high school students' mathematical reflective thinking when solving contextual trigonometry problems, viewed through the lens of their mathematics anxiety levels.

2. RESEARCH METHOD

In this research, a qualitative descriptive method and a case study design are used. The purpose of this method is to explore and map the pattern of mathematical reflective thinking among students (Damastuti et al., 2023; Munawwarah et al., 2024). This design is selected as it allows for an in-depth, detailed investigation of students' mental processes in a natural setting without manipulating the variables (Fatimah et al., 2025).

Participants

The research was conducted in a Senior High School located in Jakarta, Indonesia. The research participants were selected by applying Purposive sampling methodology. Purposive sampling is a sampling technique where the researchers intentionally select the participants based on certain criteria to suit the purpose of the research (Situmorang, 2025). The initial group of students was given the Mathematics Anxiety Scale (MAS) to identify the levels of mathematics anxiety of the students. Then, the researcher selected six students from the initial group of students as the research subjects: two students of high mathematics anxiety, two students of medium mathematics anxiety, and two students of low mathematics anxiety (Charolina, 2025). Such small number of subject selection is in line with case study research, which emphasizes the variety of cases rather than the generalizability of the findings (Herlina et al., 2025).

Research Instruments

In this qualitative study, researcher is the main instrument, assisted by some supplemental instruments (Nasution et al., 2025):

1. Mathematics Anxiety Scale (MAS): An adapted questionnaire to group students among the three spectrums of anxiety (Acharya et al., 2026).
2. Trigonometric Reflective Thinking Test: A written test that comprises contextual non-routine problems, aimed to bring out the three phases of reflective thinking: *reacting*, *comparing*, and *contemplating* (Zakiyah & Muntazhimah, 2025).
3. Interview Guidelines: a semi-structured interview used to explore more in depth the students' thinking process, employing a think-aloud protocol (Muntazhimah et al., 2023).

Based on Charolina (2025) and Munawwarah et al. (2024). Two experts in Mathematics education validated every instrument in both of them before administering them to any data. They evaluated the items and instructions to ensure they are understandable by users, the fit of the indicators between the instruments designed for each of the constructs, the appropriateness of the trigonometric problems related to the three sub-stages of reflection: *Reacting*, *Comparing*, and *Contemplating*, and the compatibility of the tools for the participants to comprehend. Using the responses from both experts to modify their text and presentation and to put indicators aligned from each instrument.

Research Procedure

The chronological course of this research followed four distinct stages:

- Preparation Stage: Developing the theoretical framework, designing instruments, and conducting expert validation.
- Anxiety Categorization: Administering the MAS questionnaire to the student to identify and select the six representative subjects.
- Data Collection Stage: Subjects completed the mathematical reflective thinking trigonometric test while engaging in think-aloud protocols. The subjects were then engaged in in-depth semi-structured interviews to clarify the ambiguities presented in their written responses.

Analysis and Synthesis Stage: Interpreting and analyzing the data gathered to formulate the pattern of reflective thinking at each level of anxiety.

Data Analysis and Validity

The data analysis process utilizes the interactive model as proposed by Miles and Huberman with three intertwined processes, namely data reduction, data display, and drawing and verifying conclusions (Haqiqiyah & Putri, 2025; Muntazhimah et al., 2023). The data reduction process involved selecting the data and coding them according to the three stages of reflection outlined by Surbeck. Coding

included searching for evidence on students' responses, thinking aloud scripts, and responses during the interview sessions based on the three phases of reflective thinking: *Reacting* (interpreting the problem for the first time), *Comparing* (relating current and previously learned information, and comparing strategies and solutions), and *Contemplating* (providing deeper evidence and thinking back about a solution and solution strategy). The coded data were then displayed in verbal and matrix narratives that are used to trace and compare students' coherence of reflection from one math anxiety level to another. To ensure the trustworthiness of the findings, methodological triangulation was done through inter-corroborating the evidence from all the research methods- mathematics anxiety questionnaire, written test, thinking aloud procedures, and interviews (Hanson-DeFusco, 2023). The study also included member checking, in which the researcher's findings were given back to students to see whether their original meaning was being accurately represented.

3. RESULTS AND DISCUSSION

This research modeled the mathematical reflective thinking processes of six senior high school students in Jakarta on the solution of contextual trigonometry problems. The six research subjects were differentiated into low, medium, and high MAS anxiety spectrum according to the data from the Mathematics Anxiety Scale (MAS). The results are depicted by a resonance metaphor to show the mental processes in the act of reflective thinking.

A. Mapping the Reflective Spectrums: Profiles of Reflective Resonance

The responses from the students show clear evidence that there is a great variation in the extent to which the students have been reflecting mathematically within the phases of *Reacting*, *Comparing*, and *Contemplating*.

A.1 The Clear Resonance: Low Anxiety Student Profile (Students 1 & 2)

The following contextual problem was given to the students: On a certain day, two members of the school flag-raising team were told by their coach to get a new rope in order to operate the pulley of the newly installed flagpole. Since the flagpole had already been erected and could not be laid horizontally, direct measurement of its height was impossible. To determine its height, one student, whose height was known to be 1.5 meters, stood 8 meters away from the pole and measured the angle of elevation to the top of the flagpole as 45° . The coach informed them that the rope length required was twice the height of the flagpole plus an additional 2 meters for knots and fastening. The rope costs IDR 12,000 per meter, and the coach provided IDR 300,000.

- Determine the height of the flagpole.
- Determine whether the money provided by the coach is sufficient, excessive, or insufficient. Explain your reasoning.
- If the coach uses the remaining money to purchase additional rope, how many meters of rope can be purchased?

The answer sheets of Students 1 and 2 are presented in Figure 1.

Students in this group showed the most complete and explicit thinking resonance. In the *Reacting* phase, they were able to develop the mathematical model in terms of the original trigonometric context. A remarkable case emerged from Student 2, where resonance was sharpened at the *Comparing* phase. Student 2 used the geometric property of the isosceles right triangle (45° - 45° - 90°) based on prior knowledge to determine the pole height directly without having to memorize tangent formulas repeatedly. In the *Contemplating* phase, they were definitive with their response and ended their thinking with a reflective statement, which is "jadi pas (it fits perfectly)", reflective of strong metacognition to check for consistency.

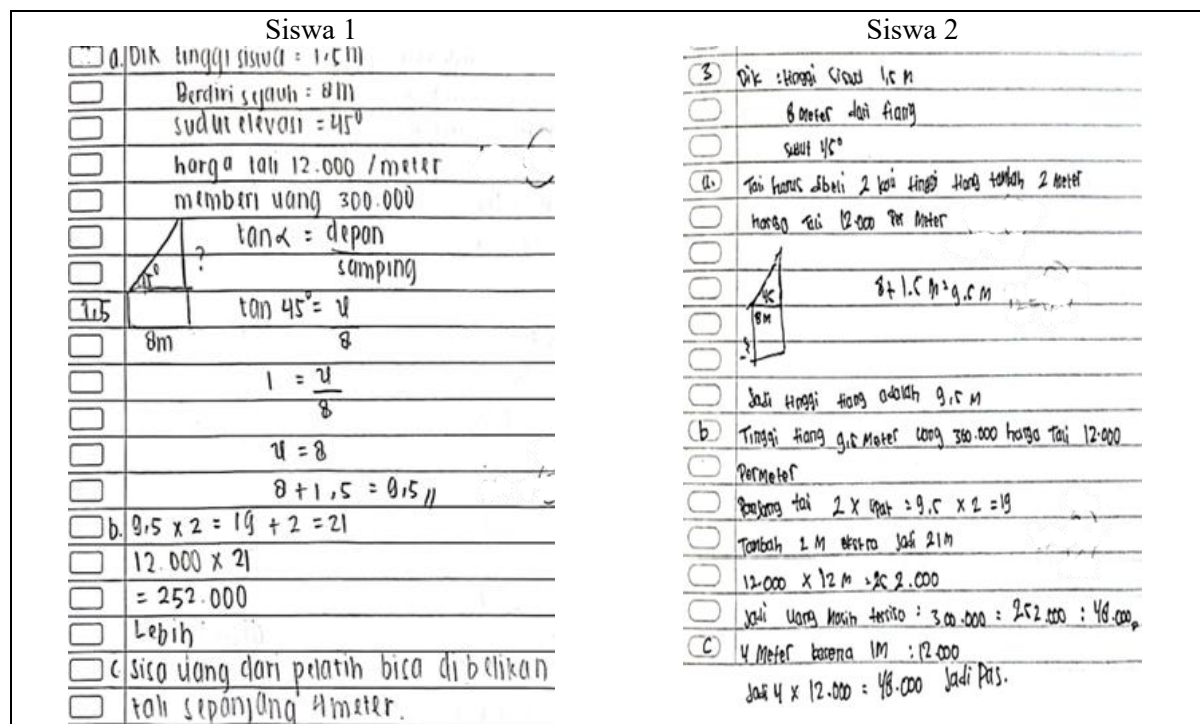


Figure 1. The First and Second Students' Answers

A.2 The Distorted Signal: Medium-Anxiety Student Profile (Students 3 & 4)

Within this group, students' ability to engage in the reflective thinking process was assessed by examining their responses to the question: Toni needs to save his pet cat, which is stranded on the roof of a house, with the aid of a ladder. The cat is positioned 7m above the ground, and the ladder must be placed such that an angle of 30° with the ground is achieved for safety purposes.

- If Toni's ladder is 8 meters long, determine whether it is sufficient to rescue the cat. Explain your reasoning.
 - What is the minimum ladder length required to reach the cat?
 - If Toni's ladder is either too short or longer than necessary, determine the deficiency or excess length.
- The answer sheets of Students 3 and 4 illustrate the reflective thinking processes they employed, as presented in Figure 2.

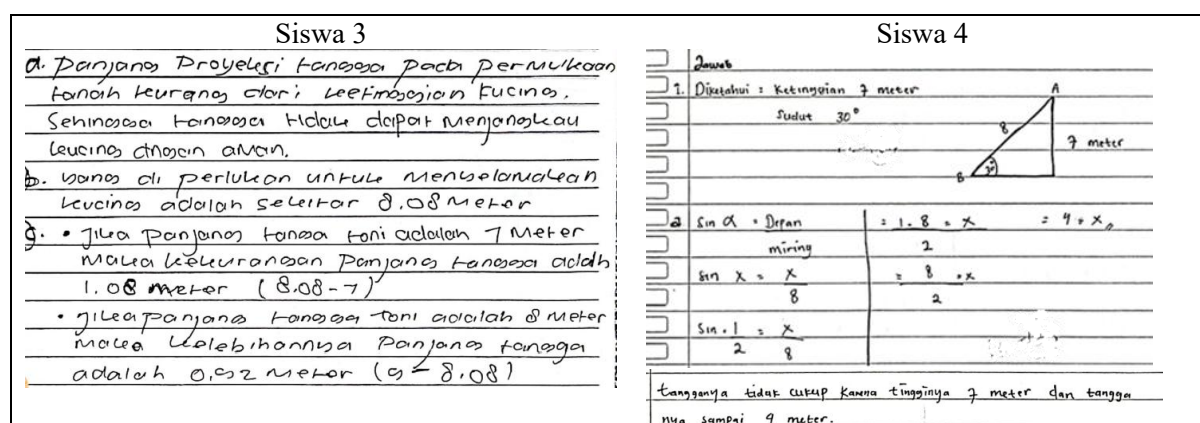


Figure 2. The Third and Fourth Students' Answers

Following figure 2, students in this spectrum resonated up to about the midpoint and then started to become distorted. While the *Reacting* phase was traversed during the phases of understanding the problem and identifying the model, the "signal" was lost when the strategy was entered. For example, Student 3 sought to utilize all available information but did not reconsider the selected strategy. During the *Contemplating* phase, the reflective "signal" was cut; students often became comfortable with the

final numeric answer without performing an in-depth "check" for errors in procedure and reached an ineffective.

A.3 The Fading Echo: High-Anxiety Student Profile (Students 5 & 6)

Students were provided with the following problem: Dani took part in a traditional pole climbing game in Sawit Village. After a long climb, he managed to get himself to the top of a 6m pole and wanted to throw down some prizes to his friends Budi and Dedi. On the look forward, Dani saw Budi at an angle of depression of 45 and on the looking sideways, he saw Dedi at an angle of depression of 30.

- a. Assuming that a stronger throw results in a greater distance, determine whether Dani must throw harder toward Budi or Dedi.

Figure 3 illustrates the reflective thinking processes exhibited by Students 5 and 6.

Siswa 5	Siswa 6
$\tan \alpha = \frac{de}{sana}$ $\tan 30^\circ = \frac{x}{6}$ $\frac{1}{2}\sqrt{3} = \frac{x}{6}$ $\frac{1}{2}\sqrt{3} \cdot 6 = x$ $3\sqrt{3} = x$ $x = 6$	$\tan \alpha = \frac{6}{x}$ $1 = \frac{6}{x}$ $x = \frac{6}{1} = 6$ $\tan (30^\circ) = \frac{6}{x}$ $\frac{1}{2}\sqrt{3} = \frac{6}{x}$ $x = \frac{6}{\frac{1}{2}\sqrt{3}} = 10.39 \text{ meter}$ jarak antara Dani dengan Budi adalah 6 meter, jarak antara Dani dengan Dedi adalah sekitar 10.39 meter.

Fig 3. The Fifth and Sixth Students' Answers

In this group, the reflective thinking resonance was linearly diminished until a state of "cognitive silence". During the Reacting stage, Student 5 could only understand superficial content. She could calculate the number using the tangent formula, but failed to analyze this result, which could help to address contextual questions. Student 6 failed in calculation from the beginning, but continued solving the problem without one clear direction of attention. The resonance was depressed almost to the extent of disappearance in the Comparing stage because students experienced "cognitive blocking" and thus lacked the drive to push to more demanding strategies. During the Contemplating stage, there was basically no reflection.

B. Discussion: Why Does the Resonance Fade?

The findings reveal that reflective thinking does not just exist as one particular cognitive framework but that it is a process where students psychological circumstances severely filter this cognitive process, which explains the disappearance of reflection (fading resonance) by indicating that the reflection process appears blocked at the critical junction that shifts the cognitive processes of simple understanding into a more advanced level of evaluation, and that Mathematics Anxiety seems to suppress a logical process.

B.1 Cognitive Interference and Working Memory Paralysis

The findings reveal that not only was reflective thinking not a standalone cognitive structure, but it was very much filtered by students' psychological status. Indeed, the fading resonance we observed demonstrates a degradation of transitions from an initial understanding and evaluation process, implying that math anxiety may hamper the continuity of logical thinking during the complex phases of thinking.

These findings are coherent with prior works that demonstrate that cognitive interference is one of the potential pathways that may lead to the impact of math anxiety on students' mathematical performance. For example, Suárez-Pellicioni et al. (2016) found that math anxiety is related to working memory and attentional control and showed that this might be mediated by the quality of working memory and attentional control. More recently, Finell et al. (2022) provided, on the basis of meta-analytic evidence, that working memory is significant in the relation between math anxiety and

mathematics performance. These findings provide a possible cognitive explanation for the fading resonance observed in the present study. When students experience anxiety while solving complex trigonometric problems, cognitive resources may be diverted toward managing anxiety and uncertainty, making it more difficult to sustain the evaluative processes required to move from initial understanding toward deeper reflection.

In the present study, this interpretation is highly relevant since our work is on reflective thinking.

Therefore, the results demonstrate that lower- vs. higher-anxiety students didn't differ in the initiation of the reasoning process, but in the ability to sustain the thinking across the other stages of thinking, namely Reacting, Comparing, and Contemplating. That is, lower-anxiety students were more capable of maintaining their reasoning and checking for strategy consistency, while the higher-anxiety students were struggling during the evaluation and reflection phase when facing difficulties and non-routine challenges. Consequently, fading resonance doesn't imply that the students do not reason anymore, but implies that the continuum of the thought may fade over time; the length and the depth of reasoning decline.

Moreover, as noted in previous studies, this cognitive pathway should not be interpreted as fatalistic. Indeed, as pointed out in Caviola et al. (2017) work about the effects of math anxiety, stress, and time pressure on strategies in mathematics, while the psychological conditions like anxiety and stress can have a detrimental impact on strategizing and working cognitive mechanisms, research findings are not fully concordant with respect to the explanation that underlies this. It is not the case here since we haven't collected any measures on working memory, so we should refer to this phenomenon as a likely pathway. All things considered, the present study suggests math anxiety acts as a filter on students who engage in a continuous thinking process when math requires deeper reasoning processes.

We also show how this potential interference shows up in the actual course of the reflective reasoning: resonance remains for a while, starts to decrease in the comparison and evaluation phase, and fades out of sight by the time students begin to contemplate. "Fading resonance" in this context refers to whether students get the right solution, or to what extent their reflective thought process could sustain as math becomes more challenging.

B.2 Urban Pressures and the Silent Reflection

We may attempt to interpret students' moderate and high anxiety responses in the context of earlier research exploring student beliefs about the role of social pressure and pressures associated with performance on tasks. In the present study, Students 3 and 4 both reported Moderate anxiety, and although they began working on mathematical reasoning, when evaluating their approach and response, their ability to maintain reflection faltered. Students 5 and 6, who both reported High anxiety, were even less able to sustain reflective thinking, especially when they expressed doubt and the fear that they would make an error. There is evidence to support the claims made in relation to the influence of external pressure on students who suffer mathematics anxiety; Hwang & Lee (2019), in their research of 439 middle school students, demonstrated a positive relation between parental pressure for academics and their levels of math anxiety. The study further concluded that parental support would moderate the relationship between academic pressure and math anxiety, yet the moderation pattern is differentiated by gender (parents' support moderating the relationship positively for the boys and negatively for the girls), thus suggesting external demands are more mentally taxing under conditions of lacking psychological support and of external demands which are perceived as pressures.

Peer pressures could also have an effect on how cognitive responses impact student behaviour when responding to mathematics problems. Competitive conditions resulted in an increase in state anxiety; one of an experimental study conducted by Liu et al. (2024) found reduced performance and attention in competitive conditions. These findings can be integrated to inform the interpretation of the effect in the present work since both result from the fading of reflective activity, and the high anxiety may similarly see the cognitive resources shifted to deal with the anxiety, which will hinder the evaluation process. There is evidence to suggest that in conditions of pressure there is an interference in strategic activity and the appearance of concern words at the expense of mental resources needed to solve math problems (Caviola et al. 2017). Nevertheless, Caviola and others stressed that the evidence is far from consistent: pressure can impact differently, according to external and individual influences.

Notably, external pressure does not always operate in an anxiety-provoking way. The study of (Hollenstein et al., 2024) utilized longitudinal data from 509 students and 28 teachers, indicating that teacher expectations predicted increases in math achievement and math self-concept and decreases in math anxiety. This research offers a crucial caveat to this interpretation: it is not necessarily external expectations themselves that are anxiety-provoking, but how individual students appraise their expectations. When students perceive teacher expectations and parental pressure as supportive and challenging, these demands may be positively motivating; conversely, when these demands are perceived as threatening and evaluative, anxiety may increase, and interference will likely emerge.

Taken in tandem, these lines of research show that the path to mathematics anxiety under duress is not a directly causal one, but a more intricate link. Here, the fact that the three anxiety groups performed differently from each other shows us that an individual student's psychological reaction to struggling mathematically may either give it prominence in her/his mind (leading to reflection) or give it salience as threatening and anxiety-provoking (leading to withdrawal). Although neither parental pressure nor teacher expectations nor competition from peers were directly tested in the present study, these findings parallel previous research indicating that the anticipation of performance and social demands contributes to mathematical anxiety and engagement. Therefore, "silent reflection," as described in highly anxious students, should be framed as more than math ignorance, but also as a potential by-product of an interaction of cognitive demands and students' psychological appraisal of those challenges.

B.3 Mathematics Anxiety as a Gatekeeper of Logic Resonance

This provides support for viewing Mathematics Anxiety as a barrier which prevents reflective thinking from realizing its potential at its deepest level. It is suggested that student psychological traits at the time of engagement with the problems were a significant variable. In particular, low-anxiety students appeared more persistent, persevering, and self-regulated when engaging with difficult mathematics, in particular displaying a more regulated, calm approach and utilizing self-coaching during the assessment.

In contrast, the psychological traits that the high-Mathematics Anxiety cohort identified at the beginning of the learning journey during trigonometry- confusion, rather than being an impetus to reflect- led to a failure of mathematical reasoning persistence. Their limited persistence when facing non-routine challenges caused their reflective resonance to fade systematically. Evidence exists in the literature to interpret these differential findings by using the concept of mathematical resilience. For example, Gokhool et al. (2022) found that mathematics anxiety may lead to maladaptive behavior to cope with challenges such as avoidance and postponement of assistance, reducing perseverance in mathematics. More recently, evidence from Hodds (2026) provided that an effective Mathematics Anxiety and mathematical resilience intervention reduced levels of Mathematics Anxiety and enhanced student Mathematics achievement. These findings are reflected in the current analysis, where high levels of mathematics anxiety are associated with reduced persistence in evaluation of strategies leading to failure to attain the Contemplating level of reflective thinking.

Although it is perhaps intuitive to suggest that mathematical resilience and maths anxiety might be inversely related overall, this should not be read as general or automatic. For example, Casinillo et al. (2022), found no significant relationship between mathematical resilience and maths anxiety in senior high school students ($p = .4725$), indicating that resilience alone might not reduce anxiety and that other affective and contextual factors might also shape students' experience of mathematics. This finding offers a useful caveat to the present results. Instead of signifying that maths anxiety directly predicts ability to be reflective, anxiety may instead modulate a student's ability to persist, control affect and maintain engagement in times of mathematical struggle.

This is arguably what can be most readily observed in the transition from Comparing to Contemplating. It was the low anxiety students who were more able to 'hang on in there' whilst comparing strategies and checking for consistency of answers. High anxiety students on the other hand appeared more inclined to opt out of closer scrutiny as soon as it becomes somewhat unclear or difficult. So, while mathematical resilience offers an appealing explanatory framework in this respect it is important to note that it was not directly measured in the present study. Thus, the "gate keeping"

function attributed to maths anxiety should perhaps be framed as a qualitative interpretation of how this anxiety might restrict the extensiveness or continuity of reflective practice.

C. Visualization of Reflective Resonance Flow

The flowchart of the dynamics of mathematical reflective thinking resonance based on mathematics anxiety level below is used to clarify this mapping. It shows where the flow of reflective resonance would be stopped at various levels of mathematics anxiety. The flow of reflective thinking goes through three stages of Reacting Comparing Contemplating.

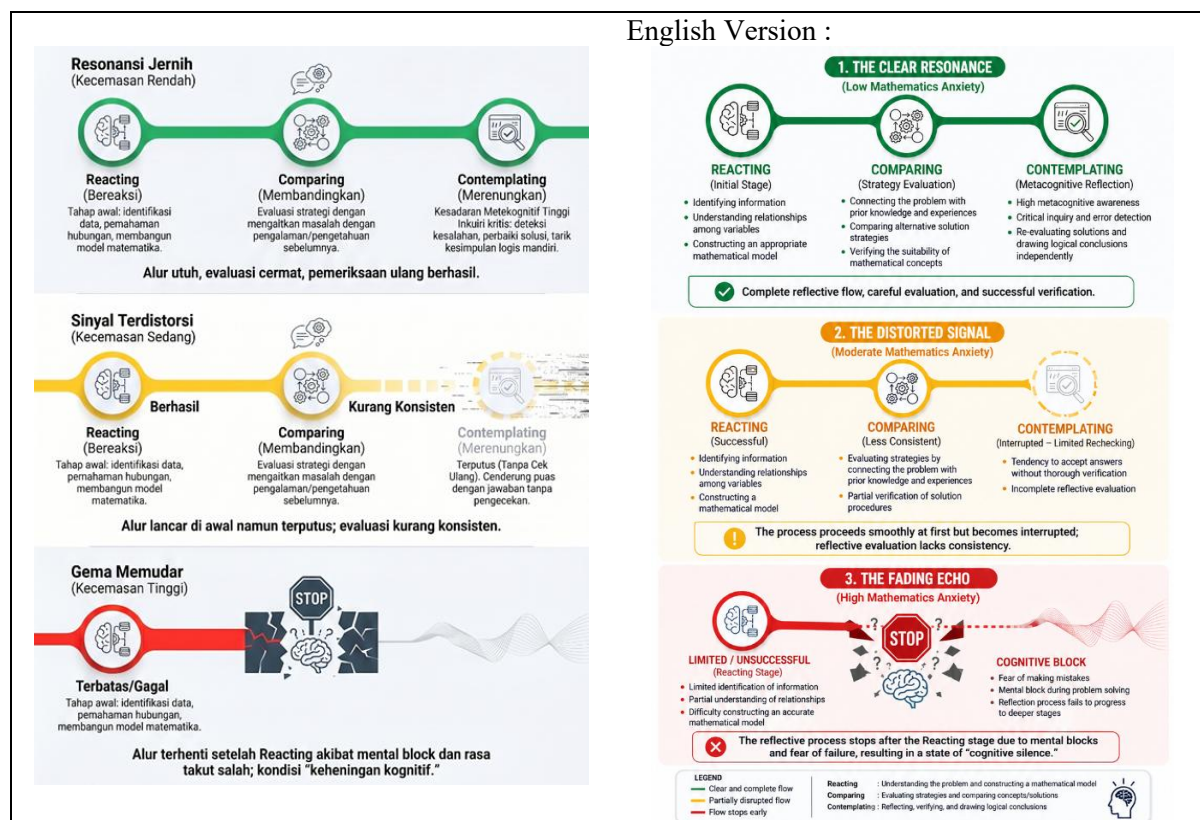


Figure 4. Dynamics of Mathematical Reflective Thinking Resonance Based on Mathematics Anxiety Levels

Figure 4 shows the trends of mathematical reflective thinking resonance in relation to mathematics anxiety levels. Compared to the students with low anxiety, the students with low reflects relatively weak but stable reflective process, which indicated a progressive development of the resonance of thinking from *Reacting*, *Comparing*, and to *Contemplating*, respectively. The students with moderate reflects a similar process, but the resonance of thinking is less stable as the students faced more challenges gradually when transitioning to the deeper contemplative stage. The students with high mathematics anxiety show relatively strong variance in the resonance of thinking and a sharp decrease in the process of continuous development of resonance of thinking when progressing to the *Contemplating* stage. The students tend to be cognitively confused and less persistent toward non-routine trigonometry problems. To conclude, the figure shows that the increase in mathematics anxiety will be simultaneously associated with a stepwise break in the flow and depth in reflective thinking to the *Contemplating* stage.

4. CONCLUSION

This study shows strong support for the original premise expressed in the introduction: that mathematical reflective thinking is not just a matter of any mathematical architecture but is heavily

filtered by the students' psychological processes. Results seem to substantiate that the "resonance" of mathematical logic is muffled by mathematics anxiety, which appears to be the biggest gatekeeper in reflective success. The findings indicate that students with low mathematics anxiety demonstrate relatively stable reflective resonance across Reacting, Comparing, and Contemplating, whereas students with high anxiety exhibit a systematic "fading resonance," in which the voice of logic becomes increasingly muffled as reflective engagement weakens. Intrusive thoughts and emotional interference may constrain the continuity of reflection, particularly during the transition from initial problem framing to deeper evaluative reflection. This pattern may be related to increased demands on working-memory resources, although working memory was not directly measured in the present study.

This study makes a novel contribution with a "thematic map" of the nature of cognitive decay in high-pressure urban settings such as Jakarta, in which the state of confusion is ultimately productive only if the student has enough psychological capacity (that is, emotional regulation and psychological safety) to use it. A policy implication would be that mathematics instruction in Jakarta should move away from the unhelpful generic remedial approach to mathematics toward a more diagnostic one that focuses on both psychological safety and emotional regulation as well as content mastery.

For future research, a promising avenue would be to create and test several "reflective scaffolds" for the anxious student to reinitiate the "thinking pulse" when it begins to fade. Moreover, conducting this kind of investigation with larger samples and other abstract mathematical topics besides trigonometry would also strengthen the applicability of the "fading resonance" model for emotion and higher-order cognition.

REFERENCES

- Acharya, A. K., Das, S., Bariha, K., & Behera, S. P. (2026). Mathematics Anxiety Among Secondary School Students in Puri District of Odisha: An Empirical Study. *International Journal of Innovative Science and Research Technology*, 3217. <https://doi.org/10.38124/ijisrt/26jan1630>
- Brockbank, R. B., & Feldon, D. F. (2024). Cognitive Reappraisal: The Bridge between Cognitive Load and Emotion. *Education Sciences*, 14(8). <https://doi.org/10.3390/educsci14080870>
- Budhathoki, B. B., Acharya, B. R., Belbase, S., Kshetree, M. P., Khanal, B., & Panthi, R. K. (2022). High School Students' Mathematics Anxiety: Discouragement, Abuse, Fear, and Dilemma Induced through Adults' Verbal Behaviour. *International Journal of Learning, Teaching and Educational Research*, 21(6), 247–269. <https://doi.org/10.26803/ijlter.21.6.15>
- Cañeda, M. E., Mata, L. P., & Bedrijo, R. W. (2024). Investigating Students' Lived Experiences in Mathematics Classroom Activities. *International Journal of Research and Innovation in Social Science (IJRISS)*, 8(11). <https://doi.org/https://dx.doi.org/10.47772/IJRISS.2024.8110122>
- Casinillo, L. F., Casinillo, E. L., Lagumbay, C. T., Abad, H. R. F., & Dagongdong, M. L. (2022). REVISITING MATHEMATICAL RESILIENCE AND ANXIETY AMONG SENIOR HIGH STUDENTS. *IJJET (International Journal of Indonesian Education and Teaching)*, 6(2), 193–203. <https://doi.org/10.24071/ijiet.v6i2.4661>
- Caviola, S., Carey, E., Mammarella, I. C., & Szucs, D. (2017). Stress, time pressure, strategy selection and math anxiety in mathematics: A review of the literature. In *Frontiers in Psychology* (Vol. 8, Number SEP). Frontiers Media S.A. <https://doi.org/10.3389/fpsyg.2017.01488>
- Cécillon, F. X., Mermillod, M., Leys, C., Bastin, H., Lachaux, J. P., & Shankland, R. (2024). The Reflective Mind of the Anxious in Action: Metacognitive Beliefs and Maladaptive Emotional Regulation Strategies Constrain Working Memory Efficiency. *European Journal of Investigation in Health, Psychology and Education*, 14(3), 505–530. <https://doi.org/10.3390/ejihpe14030034>
- Charolina, S. (2025). Mathematical Problem-Solving Ability of Students in Terms of Adversity Quotient and Self-Efficacy. *Jurnal Eduscience (JES)*, 12(3), 826.
- Damastuti, A. S., Triyanto, T., & Nurhasanah, F. (2023). Reflective Thinking Students with Different Adversity Quotients in Solving Mathematics Problems. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, 7(4), 1128. <https://doi.org/10.31764/jtam.v7i4.16552>
- Fatimah, A. E., Siagian, M. D., & Rini Sulastri. (2025). Correlation of mathematical resilience, attitudes and learning habits, gender and their impact on student academic performance. *Jurnal Elemen*, 11(3), 614–629. <https://doi.org/10.29408/jel.v11i3.30058>
- Finell, J., Sammallahti, E., Korhonen, J., Eklöf, H., & Jonsson, B. (2022). Working Memory and Its Mediating Role on the Relationship of Math Anxiety and Math Performance: A Meta-Analysis. In *Frontiers in Psychology* (Vol. 12). Frontiers Media S.A. <https://doi.org/10.3389/fpsyg.2021.798090>
- Ginting, S. S. B. (2024). The role of adversity quotient (AQ) for successful mathematics learning. *JPP (Jurnal Pendidikan Dan Pembelajaran)*, 31(2), 91–104. <https://doi.org/10.17977/um047v31i22024p91-104>

- Gokhool, F., Lawson, D., & Hodds, M. (2022). Investigating the relationship between mathematics anxiety, mathematical resilience and mathematics support engagement: an analysis of demographic and cohort factors. *MSOR Connections*, 20(2), 82–93. <https://doi.org/https://doi.org/10.21100/msor.v20i2.1321>
- Gómez, J. D. L., Álvarez Maestre, A. J., Parada Trujillo, A. E., Pérez Fuentes, C. A., Bedoya Ortiz, D. H., & Sanabria Alarcón, R. K. (2025). Determining Factors for the Development of Critical Thinking in Higher Education. In *Journal of Intelligence* (Vol. 13, Number 6). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/jintelligence13060059>
- Gradini, E., Firmansyah B, F. B., Noviani, J., & Ulya, K. (2025). Fostering Higher-Order Thinking Skills in Mathematics Education: Strategies, Challenges, and Classroom Practices. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, 13(2), 135. <https://doi.org/10.33394/j-ps.v13i2.15099>
- Hanson-DeFusco, J. (2023). What data counts in policymaking and programming evaluation – Relevant data sources for triangulation according to main epistemologies and philosophies within social science. *Evaluation and Program Planning*, 97. <https://doi.org/10.1016/j.evalprogplan.2023.102238>
- Haqiqiyah, A., & Putri, R. K. (2025). Analysis of Mathematics Problem Solving Ability Viewed from Self-Regulated Learning. *IJEMS Indonesian Journal of Education and Mathematical Science*, 6(1), 13–17. <https://doi.org/10.30596/ijems.v6i1.21723>
- Herlina, A., Syaifuddin, M., & Wuriyanto, A. B. (2025). Verbal Representation Abilities and Processes in HOTS Problem Solving Based on Adversity Quotient-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). In *Jurnal Eduscience (JES)* (Vol. 12, Number 6).
- Hodds, M. (2026). Support ting nursing students with their mathematical anxiety and mathematical resilience. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, 45, 317–330. <https://doi.org/https://doi.org/10.1093/teamat/hraf015>
- Hollenstein, L., Rubie-Davies, C. M., & Brühwiler, C. (2024). Teacher expectations and their relations with primary school students' achievement, self-concept, and anxiety in mathematics. *Social Psychology of Education*, 27(2), 567–586. <https://doi.org/10.1007/s11218-023-09856-1>
- Hwang, H., & Lee, K. (2019). The Moderating Effects of Parental Support in the Relations between Parental Academic Pressure and Mathematics Anxiety: Gender-Specific Relations. *Journal of Research in Curriculum Instruction*, 23(4), 290–304. <https://doi.org/10.24231/rici.2019.23.4.290>
- Isran, D., Susanta, A., Rahimah, D., & Syafri, F. S. (2025). Exploring Mathematical Proofs and Solutions in Higher Education: A Case Study on Real Analysis. *Mathematics Education Journal*, 19(1), 163–180. <https://doi.org/10.22342/jpm.v19i1.pp163-180>
- Joy, I., Jayson, N., & Orongan, R. C. (2025). Problem Solving Abilities and Anxiety on Students' Mathematical Resiliency. *International Journal of Research and Innovation in Social Science (IJRISS)*, 9(3). <https://doi.org/https://dx.doi.org/10.47772/IJRISS.2025.90300236>
- Latibjonovich, S. E. (2022). Features Of Teaching Trigonometry In Specialized Schools. *JournalNX- A Multidisciplinary Peer Reviewed Journal*, 9(10).
- Liu, C., Fang, M., Wang, M., Wu, Y., Chen, W., & Cheng, Y. (2024). The Effects of Peer Competition-Induced Anxiety on Massive Open Online Course Learning: The Mediating Role of the Behavioral Inhibition System. *Behavioral Sciences*, 14(4). <https://doi.org/10.3390/bs14040324>
- Liu, J., Sun, Y., & Li, Y. (2026). The impact of the mathematical adversity quotient on preparatory students' mathematics learning engagement: moderated mediation effect analysis. *Frontiers in Psychology*, 17. <https://doi.org/10.3389/fpsyg.2026.1725090>
- Masito, D., Tuhu, S., Morin, S., & Saputri, V. (2024). The Relationship Between Indicators of Mathematical Problem-Solving Ability and Performance on The PISA Test: Systematic Literature Review. In *Didaktika: Jurnal Kependidikan* (Vol. 13, Number 4). <https://jurnaldidaktika.org5149>
- Mugniarningsih, A. N., Awiria, & Rony, Z. T. (2022). Students' Learning Difficulties in Problem Solving Ability in Face-to-Face Learning Limited Elementary School. *Jurnal Pendidikan & Pengajaran Guru Sekolah Dasar*, 5(2), 66–72. <http://journal.unpak.ac.id/index.php/jppguseda>,
- Munawwarah, S., Lidyasari, A. T., & Munafiroh, V. (2024). Mathematical problem-solving ability of 5th grade students based on adversity quotient. *Jurnal Penelitian Ilmu Pendidikan*, 2(17), 87–98. <https://doi.org/10.21831/jpip.v17i1.90771>
- Muntazhimah, Khusna, H., & Anjarwati, R. (2023). Reflective thinking ability in terms of mathematical resilience of mathematics student teachers. *AIP Conference Proceedings*, 2698. <https://doi.org/10.1063/5.0122354>
- Muntazhimah, M., Turmudi, T., Prabawanto, S., Anwar, A., & Wahyuni, R. (2022). Bibliometric Analysis of Mathematics Reflective Thinking Based on Scopus Database. *European Online Journal of Natural and Social Sciences*, 11(4), 1132–1143. <http://www.european-science.com>
- Mustafa, S., & Andriani Yunus, A. (2025). Exploring the Impact of Adversity Quotient on Students' Mathematics Achievement in Junior High School. *Jurnal Pendidikan MIPA*, 26(2), 1068–1085. <https://doi.org/10.23960/jpmipa/v26i2.pp1068-1085>
- Nasution, E. Y. P., Yulia, P., & Handican, R. (2025). Analysis of students' statistical reasoning ability based on personality types. *Al-Jabar : Jurnal Pendidikan Matematika*, 16(1), 151–167. <https://doi.org/10.24042/ajpm.v16i1.26539>

- Nurdin, Hasma, E. A., Mutmainna, & Ma'rufi. (2026). The role of self-efficacy and adversity quotient in predicting basic mathematics competence within merdeka curriculum framework. *Al-Jabar Jurnal Pendidikan Matematika*, 17(1), 339–360. <https://doi.org/10.24042/ajpm.v17i1.30152>
- Nurmeidina, R., & Rafidiyah, D. (2019, November 6). Analysis of Students' Difficulties in Solving Trigonometry Problems. *Proceedings of the 4th Progressive and Fun Education International Conference, Profunedu 2019*. <https://doi.org/10.4108/eai.7-8-2019.2288422>
- Obeng, B. A., Banson, G. M., Owusu, E., & Owusu, R. (2024). Analysis of senior high school students' errors in solving trigonometry. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2385119>
- Peng, Q., Zhang, Y., Liang, S., & Xu, M. (2025). Influence of adversity quotient, emotional support, mental toughness and blunt sensitivity on college students' well-being: A framework based on structural equation modeling and multi-criteria decision-making. *Acta Psychologica*, 259. <https://doi.org/10.1016/j.actpsy.2025.105437>
- Putriani, J. D., Rosyada, A., & Hartono, Y. (2025). Students' Mathematical Modeling Ability in Answering 2022 PISA-released items in Scientific Context. *Edumatsains Jurnal Pendidikan Matematika Dan Sains*, 10(2), 15–29. <https://doi.org/10.33541/edumats>
- Reza, H. S. F., Amelia Defrianti Putri, & Dhanachat Anuniwat. (2026). Realistic Mathematics Education Improves Trigonometric Ratios: A Classroom Action Research Study. *SJME (Supremum Journal of Mathematics Education)*, 10(1), 55–67. <https://doi.org/10.35706/sjme.v10i1.13148>
- Sari, Y., Swastika, A., & Tanyono, D. (2025). Mathematical Reflective Thinking Ability in Solving Higher Order Thinking Skills (HOTS) Type Problems as Seen From Cognitive Style. *Jurnal Dimensi Pendidikan Dan Pembelajaran*, 1(13), 65–78. <https://journal.umpo.ac.id/index.php/dimensi/index>
- Setiyani, Gunawan, Waluya, S. B., Sukestiyarno, Y. L., Cahyono, A. N., Mutia, & Yunianto, W. (2025). The role of scaffolding in shaping reflective mathematical thinking of dependent field students in numeracy problems. *Infinity Journal*, 14(2), 531–550. <https://doi.org/10.22460/infinity.v14i2.p531-550>
- Setiyani, Waluya, S. B., Sukestiyarno, Y. L., & Cahyono, A. N. (2024). Construction of reflective thinking: A field independent student in numerical problems. *Journal on Mathematics Education*, 1(15), 151–172. <https://doi.org/10.22342/jme.v15i1.pp151-172>
- Situmorang, P. L. (2025). Analysis of students' problem solving abilities with symbolic representation through the implementation of the case method. *Education and Social Sciences Review*, 6(2), 17–29. <https://doi.org/10.29210/07essr630400>
- Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. In *Cognitive, Affective and Behavioral Neuroscience* (Vol. 16, Number 1, pp. 3–22). Springer New York LLC. <https://doi.org/10.3758/s13415-015-0370-7>
- Syadid, R. A. A. C., & Sutiarto, S. (2021). Hubungan Kemampuan Berpikir Reflektif Matematis dengan Kemampuan Pemecahan Masalah Matematis Peserta Didik. *JKPM (Jurnal Kajian Pendidikan Matematika)*, 327–336. <http://journal.lppmunindra.ac.id/index.php/jkpm/>
- Toraman, Ç., Orakçı, Ş., & Aktan, O. (2020). Analysis of the Relationships between Mathematics Achievement, Reflective Thinking of Problem Solving and Metacognitive Awareness. *International Journal of Progressive Education*, 16(2), 72–90. <https://doi.org/10.29329/ijpe.2020.241.6>
- Ugli, N. A. N. (2025). Psychological Characteristics of the Development of Analytical Thinking in Adolescents. *Jurnal Psikologi*, 2(4), 10. <https://doi.org/10.47134/pjp.v2i4.4460>
- Yiğit Koyunkaya, M. (2016). Mathematics education graduate students' understanding of trigonometric ratios. *International Journal of Mathematical Education in Science and Technology*, 47(7), 1028–1047. <https://doi.org/10.1080/0020739X.2016.1155774>
- Zakiah, S., & Muntazhimah, M. (2025). Brain-Based Learning Model to Improve Mathematical Reflective Thinking Ability in Trigonometry : An Experimental Study. *Jurnal Paedagogy*, 12(3), 626. <https://doi.org/10.33394/jp.v12i3.16011>