*Journal of Education and e-Learning Research*
*Vol. 12, No. 3, 507-519, 2025*
*ISSN(E) 2410-9991 / ISSN(P) 2518-0169*
*DOI: 10.20448/jeelr.v12i3.7398*


# Predictive model of mathematical literacy ability, specifically university students' critical thinking skills: A case study on the inappropriateness of multilevel regression

**Pitriana Tandililing[1]**
**Bettisari Napitupulu[2]**
**Okky Riswandha Imawan[3]**
**Raoda Ismail[4]**

(Corresponding Author)

[1,2,3,4]*Mathematics Education, Universitas Cenderawasih, Papua, Indonesia.*
[1]*Email: fitrianawill@gmail.com*
[2]*Email: napitupulubettisari@gmail.com*
[3]*Email: okkyriswandha.2021@student.uny.ac.id*
[4]*Email: raodaismail26@gmail.com*

## Abstract

This study aims to analyze the influence of Project-Based Learning (PjBL) and Guided Discovery Learning (GDL) on students' critical thinking skills in geometry focusing on mathematical literacy within geometry content. Specifically, it examines the effects of self-confidence, mathematical literacy scores, and gender on critical thinking skills and compares the predictive accuracy of multiple linear regression and multilevel regression models. A quantitative approach with regression analysis was used. Multiple linear regression and multilevel regression models were applied to assess the relationships among variables. Model accuracy was evaluated using intraclass correlation coefficients (ICC) to determine the impact of class-level grouping. Self-confidence and mathematical literacy in geometry significantly influenced students' critical thinking skills. Gender showed no significant effect. The multiple linear regression model outperformed the multilevel model as reflected by lower ICC values, indicating that class grouping did not substantially affect critical thinking outcomes. Internal factors, such as confidence and subject mastery are more critical to students' critical thinking development than external factors like gender or class grouping. Teachers should prioritize enhancing students' confidence and content mastery through well-designed instructional strategies to foster critical thinking in mathematics learning.

**Keywords:** Critical thinking skills, Guided discovery Learning, Higher-order thinking, Multilevel regression, Project-based Learning, Mathematical literacy, Self-confidence.

**Citation** | Tandililing, P., Napitupulu, B., Imawan, O. R., & Ismail, R. (2025). Predictive model of mathematical literacy ability, specifically university students' critical thinking skills: A case study on the inappropriateness of multilevel regression. *Journal of Education and E-Learning Research, 12*(3), 507–519. 10.20448/jeelr.v12i3.7398

**History:**
Received: 24 February 2025
Revised: 29 July 2025
Accepted: 25 August 2025
Published: 11 September 2025

**Publisher:** Asian Online Journal Publishing Group

**Funding:** This work is supported by the Ministry of Education, Culture, Research, and Technology, Center for Higher Education Funding, Indonesia (Grant number: 0668/J5.2.3/BPI.06/10/2021).
**Institutional Review Board Statement:** The Ethical Committee of the Yogyakarta State University, Indonesia has granted approval for this study on 20 October 2024 (Ref. No. T/60.1/UN34.9/PT.01.04/2024).
**Transparency:** The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
**Competing Interests:** The authors declare that they have no competing interests.
**Authors' Contributions:** All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

## Contents

**Contribution of this paper to the literature**

This study uniquely compares the predictive power of multiple linear and multilevel regression models in analyzing critical thinking skills while integrating self-confidence and mathematical literacy in geometry as key predictors—an approach that has been scarcely explored in mathematics education research.

## 1. Introduction

Critical thinking is a fundamental competency that students must develop to effectively address challenges in academic and professional contexts. It involves the capacity to analyze, interpret, and draw reasoned conclusions based on relevant evidence. The development of critical thinking skills can be significantly enhanced through the application of well-designed instructional strategies and appropriate learning resources. Research has shown that the use of problem-based learning modules is particularly effective in promoting the advancement of students' critical thinking abilities (Mujiyati, Warto, & Sutimin, 2019).

Assessing and predicting students' critical thinking abilities has become a key concern in educational research (Coulacoglou & Saklofske, 2017; Neukrug & Fawcett, 2015). Multilevel regression is often used to analyze hierarchical data where student information is nested within larger groups like classrooms or faculties with advancements in data analysis.

Although multilevel regression is a valuable analytical tool, it may not always be the most appropriate choice for certain datasets. When essential assumptions such as adequate variability between groups or minimal measurement error are violated, the model may yield inaccurate or biased estimates. In such circumstances, relying on multilevel regression could lead to misleading conclusions that fail to accurately represent the underlying patterns in the data.

This study investigates a case in which multilevel regression was employed to predict students' critical thinking skills but did not achieve the anticipated level of predictive accuracy. The findings reveal specific challenges faced during the modeling process and propose alternative analytical approaches that may be more effective under similar conditions.

This article seeks to support researchers and teachers in making more informed decisions when selecting statistical methods for analyzing hierarchical data by addressing these limitations. Furthermore, it explores potential improvements to enhance the precision and reliability of models used to assess students' critical thinking skills.

## 2. Literature Review

Critical thinking skills are a vital component in the development of students' intellectual capacity. They encompass the ability to objectively evaluate, interpret, and analyze information to make well-reasoned and informed decisions. In the context of higher education, cultivating critical thinking is a core objective as it prepares students to effectively navigate complex challenges in academic settings, professional environments, and everyday life.

The PISA 2022 Mathematical Literacy Framework maintains key content areas, such as quantity, uncertainty and data, change and relationships, and space and shape while emphasizing the role of critical thinking skills as a crucial 21st-century competency (OECD, 2022). Students must apply mathematical reasoning to solve real-world problems across the following four contextual areas: personal, occupational, societal, and scientific. In this context, critical thinking involves analyzing, evaluating, and interpreting information logically. This study assesses critical thinking through multiple-choice tests, measuring students' ability to identify valid arguments, recognize biases, and make reasoned decisions within mathematical literacy.

In the digital era, critical thinking is essential for navigating vast information sources, filtering reliable data, and avoiding misinformation. It also supports problem-solving in complex, unstructured situations by fostering analytical reasoning and creative solutions. This study highlights their importance in preparing individuals to tackle modern challenges effectively by examining students' critical thinking skills in mathematics.

Critical thinking is one of the essential higher-order thinking skills required in the 4.0 era, particularly during the pandemic. Critical thinking is not solely reliant on innate talent but can be cultivated through practice and training (Handoko, Mardiati, Ismail, & Imawan, 2023; Imawan, Ismail, Tandililing, & Aisyah, 2023; Ismail, Retnawati, Sugiman, & Imawan, 2025; Rott & Rott, 2021; Tiruneh, De Cock, & Elen, 2018; Tiruneh, De Cock, Weldeslassie, Elen, & Janssen, 2017). Various educational models have been shown to enhance critical thinking skills (Dyer & Sherin, 2016; Ismail, Imawan, & Nadhifah, 2023; Kuntze, Aizikovitsh-Udi, & Clarke, 2017). Students majoring in mathematics are expected to develop strong critical thinking skills to become professional mathematics teachers in the 4.0 era (Csapodi & Hoffmann, 2021; Imawan, Retnawati, Haryanto, & Ismail, 2024).

Pedagogical models such as Project-based Learning (PjBL) and Guided Discovery Learning (GDL) have been implemented in geometry courses, particularly in mathematical literacy within the geometry content to improve students' critical thinking abilities. However, the application of these models depends heavily on the teachers, which results in different teaching approaches across different classes (Tiruneh et al., 2017). These variations in teaching methods are likely to affect the diverse characteristics of student variables (Imawan et al., 2024; Purnomo, Arifin, Rahmawati, & Rahmawati, 2024). Therefore, when analyzing a specific variable from students, such as critical thinking test scores, it is essential to consider class differences.

Geometry has long been included in the school mathematics curriculum where students learn to reason through mathematical axiomatic structures (National Council of Teachers of Mathematics (NCTM), 2000). The competencies for the solid geometry course are outlined in the 2012 curriculum of the bachelor's degree program in mathematics and consist of three main components: (1) solving problems related to 3D shapes, the relationships between spatial elements, perpendicular lines and planes and distances between geometric objects. (2) Solving problems related to polyhedrons. (3) Solving problems related to curved-surface solids.

As adult learners, university students exhibit the following characteristics: (1) They participate in educational programs or training with a high motivation to learn. (2) They want to understand what they are learning will benefit them. (3) Time is a critical factor for students. (4) They respect teachers who are knowledgeable and can deliver material effectively. (5) They view the classroom as a space to gain broad experiences for both personal and professional growth. (6) They are capable of self-direction and independent action. (7) They prefer to be involved in decision-making processes. (8) They enjoy collaborating with groups and socializing with their peers (Cirillo & Hummer, 2021; Ismail, Retnawati, Sugiman, & Imawan, 2024; Kemp, Morrison, & Ross, 1994).

Different scholars have offered various perspectives on critical thinking. Orlich, Harder, and Calahan (2007) highlight that two key levels in Bloom's Taxonomy, i.e., analysis and synthesis play a fundamental role in developing higher-order thinking. Analysis involves examining information by breaking it into smaller components to understand its structure while synthesis refers to the ability to combine different ideas to create something new. These cognitive processes are essential for problem-solving and evaluating information effectively, making them central to critical thinking.

Similarly, Cottrell (2005) describes critical thinking as a mental process that involves actively engaging the mind. Developing critical and analytical thinking skills requires utilizing cognitive functions, such as focusing attention, organizing information into categories, identifying key details and making reasoned judgments. This definition emphasizes that critical thinking is not just about acquiring knowledge but also about processing and evaluating information effectively.

In addition, Ennis (1996) offers another definition stating that "critical thinking is reasonable, reflective thinking focused on deciding what to believe or do. The emphasis is on reasonableness, reflection, and the process of making decisions." This suggests that critical thinking involves thoughtful, rational reflection aimed at determining what one should believe or what course of action to take. The focus here is on the reasonableness of the thought process, reflection, and decision-making.

From the perspectives discussed above, it can be concluded that critical thinking is a higher-order cognitive skill. It involves the ability to analyze and synthesize information. Analysis refers to the capacity to identify the components that make up a concept while synthesis is the ability to combine various elements into a unified whole.

In line with the previous expert opinions, Arends and Kilcher (2010) explain that critical thinking involves a deliberate and reflective approach to evaluating arguments, identifying potential biases and logical errors, and making conclusions based on credible evidence and careful reasoning. This definition underscores the reflective nature of critical thinking, highlighting its purpose of analyzing arguments, identifying errors and biases, and drawing conclusions grounded in evidence and sound reasoning. Critical thinking skills have indicators that align with higher-order thinking skills (Ismail et al., 2024). Studies suggest that accurately assessing higher-order thinking skills requires the development of valid and reliable instruments (Hamdi, Suganda, & Hayati, 2018). Reliable assessments help ensure that the evaluation of students' abilities reflects their actual cognitive skills rather than external factors.

Self-confidence which is thought to influence critical thinking skills consists of two key components: competence and self-assurance. Competence refers to having the necessary knowledge or skills to complete a task successfully while self-assurance reflects a person's belief in their own ability to succeed regardless of their actual skill level. These two elements shape how individuals approach challenges and persist in achieving their goals (Akbari & Sahibzada, 2020; Imawan, Retnawati, Haryanto, & Ismail, 2025; Kiverstein, Rietveld, Slagter, & Denys, 2019; Liu, Xu, Montes, Dong, & Herrera, 2019). McElmeel (2002) defines confidence as a belief in oneself and one's abilities emphasizing the role of self-perception in personal achievement.

In education, predictive models are often used to examine the factors that impact student learning, including the development of critical thinking skills. Understanding these influences can help teachers design more effective learning strategies that foster both confidence and cognitive growth in students.

Different statistical techniques are commonly used to develop predictive models, including linear regression, logistic regression, and multilevel regression. These methods help researchers understand the relationships between individual and group-level variables in shaping educational outcomes, such as critical thinking skills. Regression analysis plays an important role in education, particularly in improving assessment tools. For instance, previous studies have demonstrated that logistic regression can effectively detect item bias, leading to more valid and reliable evaluation instruments (Liestari & Muhardis, 2021; Mujiyati et al., 2019; Sumin, Sukmawati, & Nurdin, 2022; Ulwatunnisa, Retnawati, Muhardis, & Yusron, 2023).

Multilevel regression also referred to as hierarchical or random-effects modeling are useful for analyzing data with a nested structure, such as students grouped within different classes or schools. This approach allows researchers to examine how variations at different levels contribute to individual learning outcomes. In educational research, multilevel regression is often applied to account for clustering effects when comparing academic performance across schools or classrooms.

However, the effectiveness of multilevel regression depends on meeting specific statistical assumptions. Studies suggest that when data do not conform to normality or when between-group variability is low, the results may be unreliable (Chua, 2014; Ramos, Dolipas, & Villamor, 2018; Subali, Paidi, & Mariyam, 2017). Measurement errors at the student or class level can also introduce bias leading to misleading interpretations. This is especially relevant in studies with small sample sizes or minimal differences between groups where applying multilevel regression might not yield meaningful insights and could instead result in inaccurate conclusions. Therefore, it is important to check whether the data structure truly requires a multilevel model and whether the variability at the group level is sufficient to support such an analysis. When these assumptions are not met, other techniques, such as simple linear regression or alternative statistical models, may provide more accurate results.

Following a comprehensive review of the relevant literature, the objectives of this study are formulated as follows: (1) to investigate the significance of self-confidence (affective), mathematical literacy in geometry test scores (cognitive), and gender in relation to university students' critical thinking skills in geometry. (2) To construct a regression model that most accurately predicts students' critical thinking skills.

## 3. Methodology

### *3.1. Research Design*

This study employed a descriptive quantitative research design. To achieve the first research objective, multiple linear regression analysis was utilized. For the second objective, a comparison was conducted between multiple linear regression and multilevel regression models. The analysis explored the necessity of applying multilevel regression in this particular case as there were indications of differences between classes, especially regarding the use of different teaching models despite the students belonging to the same university.

Multilevel regression is a statistical modeling technique used to estimate relationships between variables observed at different hierarchical levels within nested or structured data (Rencher, 2001; Stevens, 2009; Tabachnik & Fidell, 2014). The two-level regression model is the simplest form of multilevel regression where the first level consists of individual data, and the second level consists of group data (Johnson & Wichern, 2007; West, Welch, & Galecki, 2006).

### 3.2. Sample and Data Collection

The population of this study consists of 116 students from the Mathematics Education Study Program at the bachelor's degree level. Data were collected using a critical thinking skills test on spatial geometry material, a mathematical literacy in geometry test, and a self-confidence questionnaire. All instruments in this study have provided good results when their reliability was estimated and their validity was tested. The instrument for measuring critical thinking skills must be valid and reliable to provide accurate results (Rosnawati, Kartowagiran, & Jailani, 2015).

### 3.3. Analyzing of Data

The data in this study was analyzed using Jamovi software following a structured process to ensure accurate and reliable results. The steps undertaken in the analysis are outlined below:

1. Testing assumptions and running multiple linear regression.

The first step involved checking key assumptions to confirm the suitability of multiple linear regression. These included tests for residual normality, multicollinearity, homoscedasticity, and data linearity (Gudono, 2017; Johnson & Wichern, 2007; Rencher, 2001; Tinsley & Brown, 2000). Once these assumptions were met, a multiple linear regression analysis was conducted to examine how the independent variables influenced students' critical thinking skills.

2. Running a Null Multilevel Regression Model.

In this phase, a multilevel regression model without independent variables was applied. This allowed for an evaluation of the variance between different classrooms serving as a baseline for further analysis (Denis, 2020).

3. Conducting Multilevel Regression with a Random Intercept.

The next step involved applying a multilevel regression model with a random intercept. This model recognized that different classes might have varying baseline levels capturing the differences between groups more effectively (Denis, 2019).

4. Applying Multilevel Regression with a Random Slope.

The final stage incorporated a multilevel regression model that allowed both the intercept and slopes to vary across classrooms. This approach provided deeper insights into how the relationships between variables differed between groups (Hair, Babin, & Anderson, 2019).

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were compared to determine the most suitable model for predicting students' critical thinking skills (Gagné & Dayton, 2002; Portet, 2020; Sanquetta et al., 2018; Snipes & Taylor, 2014). The model with the lowest AIC and BIC was selected as the best fit for the study. Smaller AIC and BIC values indicate a better-fitting model as they suggest a model with fewer residuals and a more accurate fit to the data.

## 4. Findings

The results of this study are described classically and modernly. The classical theory was used because the subjects of this study were not too many, namely only 102 students. Meanwhile, modern theory is used in this research to support the results of classical theory and to develop it because modern theory can produce more detailed output using the DIF analysis model.

### 4.1. Model 1: Multiple Linear Regression

Assumption tests were performed to ensure the model's validity. Below are the results of these tests before conducting data analysis using multiple linear regression.

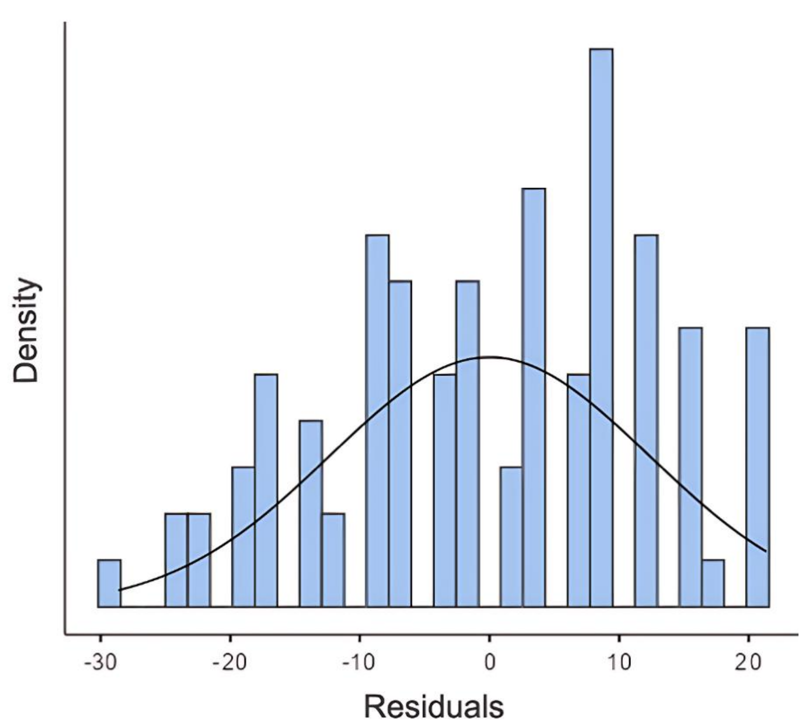



**Figure 1.** Histogram of residuals

### 4.1.1. Normality Assumption

The histogram in Figure 1 provides a visual representation of the residuals' distribution. A normal distribution of residuals should follow a bell-shaped curve indicating that errors are normally distributed.

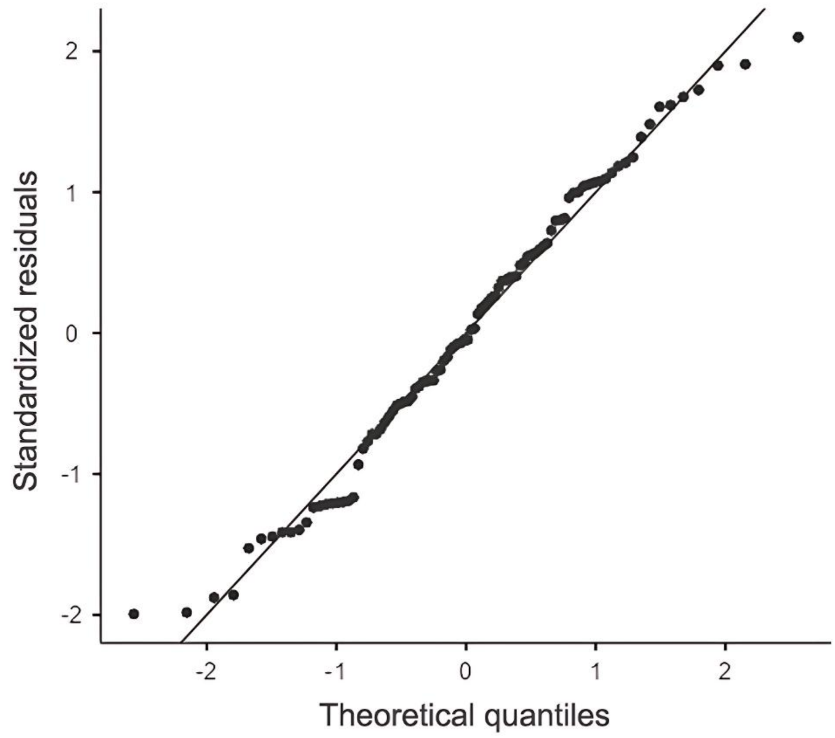



**Figure 2.** Residual plot

The residual plot in Figure 2 shows the relationship between the predicted values and residuals. For a well-fitted model, the residuals should be randomly scattered, indicating no discernible pattern that would suggest homoscedasticity and that the linear model is appropriate.

Figures 1 and 2 provide visual insights into the residuals' distribution and the pattern of residuals in relation to predicted values. These plots help verify whether the assumptions of normality and homoscedasticity hold for the regression model.

**Table 1.** Normality test of the data

| Normality test | Statistic | P |
|---|---|---|
| Shapiro-Wilk | 0.983 | 0.232 |
| Kolmogorov-Smirnov | 0.076 | 0.632 |
| Anderson-Darling | 0.360 | 0.442 |

Table 1 shows the results of three different normality tests (Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling) conducted on the data. Since all p-values are greater than 0.05, we can conclude that the data is normally distributed.

**Table 2.** Residual normality test

| Residual normality | Statistic | P |
|---|---|---|
| Shapiro-Wilk | 0.978 | 0.119 |
| Kolmogorov-Smirnov | 0.079 | 0.578 |

Table 2 presents the results of normality tests on the residuals. The p-values for both the Shapiro-Wilk and Kolmogorov-Smirnov tests are greater than 0.05 with the data normality tests. This indicates that the residuals follow a normal distribution satisfying the normality assumption.

Tables 1 and 2 demonstrate that the data and residuals meet the normality assumption as indicated by p-values greater than 0.05 across multiple normality tests. This means the residuals are normally distributed, which is crucial for the validity of the multiple linear regression model.

**Table 3.** Multicollinearity test

| Variables | VIF | Tolerance |
|---|---|---|
| Self- confidence (X1) | 1.05 | 0.953 |
| Spatial geometry test results(X2) | 1.05 | 0.953 |

### 4.1.2. Multicollinearity Assumption

Table 3 presents the results of the multicollinearity test using the Variance Inflation Factor (VIF) and tolerance. Both independent variables (X1: Self- confidence and X2: spatial geometry test results) have VIF values below 10, and the tolerance values are close to 1. This indicates that there is no significant multicollinearity between the independent variables, fulfilling the assumption of multicollinearity. Table 3 confirms that multicollinearity is not an issue as all VIF values are below 10, and tolerance values are close to 1. This means the independent variables are not highly correlated and the model can reliably estimate the effects of each predictor.

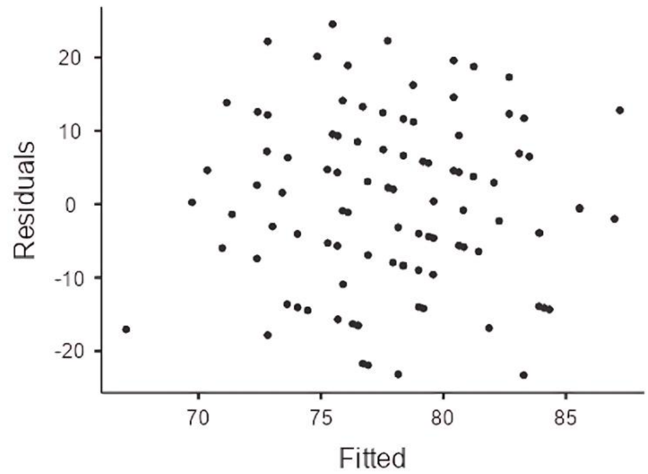



**Figure 3.** Fitted plot

### *4.1.3. Homoscedasticity Assumption*

Figure 3 shows the residuals plotted against the fitted values. This scatter plot helps to visually inspect for any systematic patterns in the residuals. In a model where homoscedasticity holds, the residuals should be randomly distributed with no clear pattern or trend.

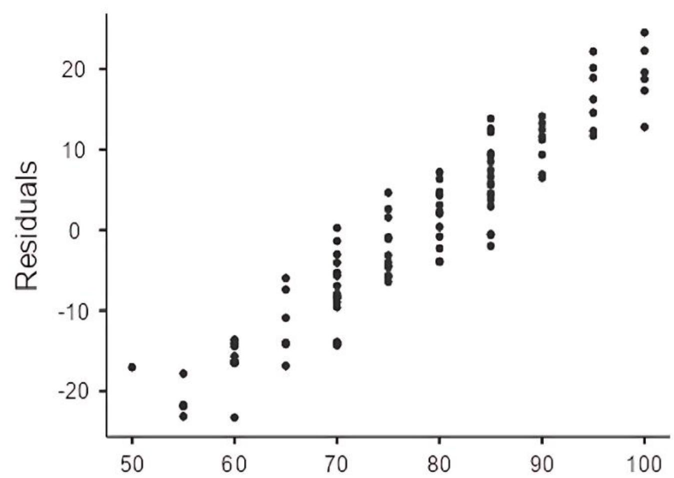
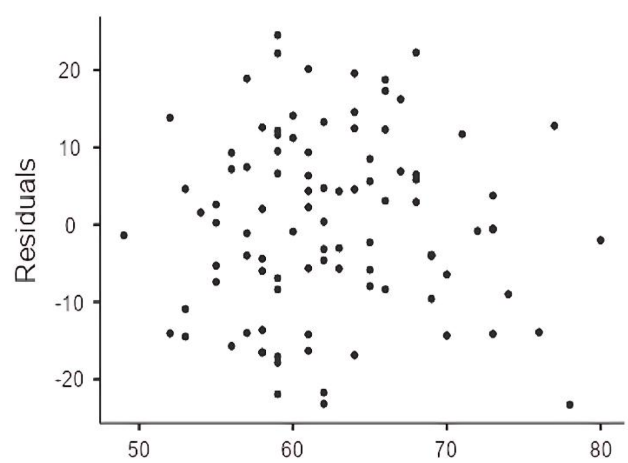
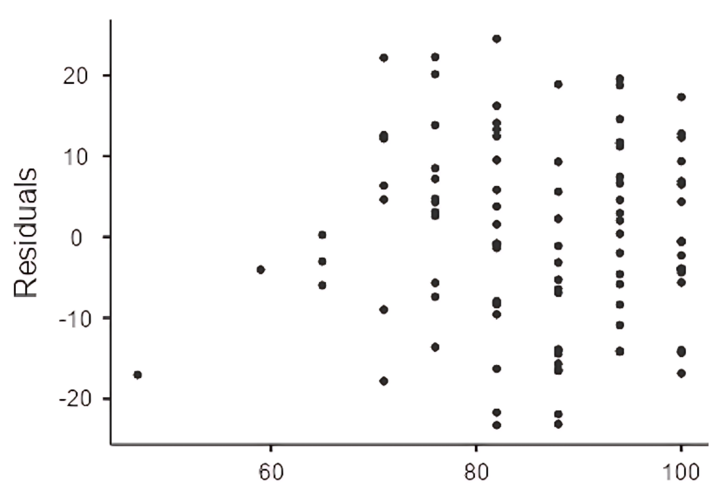



**Figure 4.** Plot of variable Y.

Figure 4 displays the relationship between the dependent variable (Y) and residuals. Similar to the fitted plot, this helps ensure that residuals are randomly distributed across the values of Y.




**Figures 5.** Plots of variables X1.




**Figures 6.** Plots of variables X2.

Figures 5 and 6 show the residuals plotted against the independent variables (X1: self- confidence and X2: spatial geometry test results). The expectation is that the residuals should not exhibit any clear patterns across the predictor values indicating that the variance of the errors is constant for all values of X1 and X2.

**Table 4.** Homoscedasticity test results

| Homoscedasticity test | Statistic | P |
|---|---|---|
| Breusch-Pagan | 0.941 | 0.625 |
| Goldfeld-Quandt | 0.600 | 0.955 |
| Harrison-McCabe | 0.616 | 0.938 |

Table 4 presents the results from three different statistical tests to assess homoscedasticity: Breusch-Pagan, Goldfeld-Quandt, and Harrison-McCabe. For each test, the p-values are all greater than 0.05 indicating that the assumption of homoscedasticity is satisfied. This means that the variance of residuals is consistent across the range of predicted values and independent variables, fulfilling a key assumption for valid multiple linear regression analysis.

A multiple linear regression analysis was carried out to examine the connection between the independent variables (X1: self- confidence and X2: spatial geometry test results) and the dependent variable (Y: critical thinking skills in geometry) after confirming that all assumptions were met (normality of residuals, no multicollinearity, homoscedasticity, and linearity).

**Table 5.** Model 1 summary

| Model | R | $R^2$ | Adjusted $R^2$ | AIC | BIC |
|---|---|---|---|---|---|
| 1 | 0.330 | 0.109 | 0.0899 | 751 | 762 |

Table 5 presents the model summary. The $R^2$ value of 0.109 indicates that the two independent variables (X1 and X2) collectively explain 10.9% of the variance in the dependent variable (Y). This indicates a modest explanatory power while it still suggests a statistically significant relationship. Additionally, the AIC (751) and BIC (762) values provide information on the model's goodness-of-fit and can be used to compare this model with alternative models.

**Table 6.** Coefficients of model 1

| Predictor | Estimate | SE | T | p |
|---|---|---|---|---|
| Intercept | 31.510 | 13.965 | 2.260 | 0.026 |
| (X1) | 0.410 | 0.196 | 2.090 | 0.039 |
| (X2) | 0.241 | 0.113 | 2.130 | 0.036 |

Table 6 provides the estimated coefficients, standard errors (SE), and p-values for each predictor in the regression model. The intercept has a coefficient estimate of 31.510 ($p = 0.026$) representing the baseline level of critical thinking skills when both X1 and X2 are zero. For X1 ( self- confidence), the coefficient estimate is 0.410 with a p-value of 0.039 indicating a statistically significant positive relationship with the dependent variable. This means that as students' self-confidence increases, their critical thinking skills in geometry improve with each unit increase in self-confidence contributing to a 0.410 increase in the critical thinking score. For X2 ( spatial geometry test results), the coefficient estimate is 0.241 with a p-value of 0.036 also indicating a statistically significant positive relationship. This suggests that higher performance on the spatial geometry test is associated with better critical thinking skills with each unit increase in test results contributing to a 0.241 increase in the critical thinking score.

The multiple linear regression models show that both self-confidence and performance on the spatial geometry test have significant positive effects on students' critical thinking skills in geometry. The model explains 10.9% of the variation in critical thinking skills, and the assumptions for linear regression (normality, no multicollinearity, homoscedasticity, and linearity) were adequately met. Therefore, both predictors contribute meaningfully to understanding the factors that influence students' critical thinking abilities in the context of geometry.

### *4.2. Model 2: Null Multilevel Regression ( Without Including Independent Variables)*

This model investigates the variation in critical thinking skills in geometry (Y) across two levels: individual students (level 1) and classrooms (level 2). The analysis was conducted without including any independent variables (X1: self-confidence or X2: spatial geometry test results) allowing for an exploration of whether classroom-level differences (level 2) contribute to variations in critical thinking skills. The Intra-Class Correlation Coefficient (ICC) is used to measure the proportion of variance explained by the classroom differences at level 2.

**Table 7.** Model 2 overview

| Info | |
|---|---|
| **Estimate** | **Linear mixed model fit by REML** |
| Call | (Y) Critical thinking skills ~ 1 + (1 \| classroom no) |
| AIC | 760.342 |
| BIC | 765.546 |
| LogLikel. | -375.927 |
| R-squared Marginal | 0.000 |
| R-squared conditional | 0.016 |
| Converged | Yes |
| Optimizer | Bobyqa |

Table 7 summarizes the null multilevel model fit. The model evaluates the overall mean of critical thinking skills (without considering independent variables) while accounting for classroom-level variance. The AIC of 760.342 and BIC of 765.546 provide fit statistics for comparing models. The R-squared marginal is 0.000, indicating that the model explains no variance at the individual level. However, the R-squared conditional is 0.016, meaning that approximately 1.55% of the variance in critical thinking skills can be attributed to classroom-level differences.

**Table 8.** Random components

| Groups | Name | SD | Variance | ICC |
|---|---|---|---|---|
| No class | Intercept | 1.550 | 2.390 | 0.0155 |
| Residual | Residual | 12.320 | 151.770 | - |

Table 8 presents the random effects for classroom-level variance. The Intra-Class Correlation Coefficient (ICC) is 0.0155 indicating that 1.55% of the total variance in critical thinking skills is due to classroom-level differences. The influence of classroom grouping is present; it is relatively small suggesting that the variation between individual students within classrooms accounts for the majority of the variance in critical thinking skills. The standard deviation (SD) for classroom intercepts is 1.55 and the corresponding variance is 2.39. The residual variance (within-classroom variance) is 151.77 showing that most of the variation lies at the student level.

**Table 9.** Fixed effects parameter estimates

| Names | Confidence interval (95%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Estimate | SE | Lower | Upper | df | T | p |
| Intercept | 77.9 | 1.67 | 74.7 | 81.2 | 1.000 | 46.7 | 0.014 |

Table 9 shows the fixed effect for the intercept which represents the overall mean score for critical thinking skills across all students regardless of classroom membership. The estimate for the intercept is 77.9, with a standard error (SE) of 1.67, and the confidence interval ranges from 74.7 to 81.2. The intercept is statistically significant with a p-value of 0.014 indicating that the overall average critical thinking skill score differs significantly from zero.

In the null multilevel regression model, approximately 1.55% of the variance in students' critical thinking skills in geometry can be attributed to differences between classrooms. Although this is a small proportion, it indicates that classroom-level factors (such as differences in teaching style or classroom environment) play a minor role in students' critical thinking skills. The remaining 98.45% of the variance occurs within classrooms, highlighting that individual-level factors are the primary drivers of critical thinking skill differences in this context. The small yet significant ICC suggests that while classrooms influence critical thinking outcomes, most of the variation is within the classroom, emphasizing the importance of individual student characteristics.

### *4.3. Model 3: Multilevel Regression with Random Intercepts*

Model 3 incorporates random intercepts to account for classroom-level variability (level 2) while analyzing the influence of individual-level factors (level 1: X1, X2). In this model, the intercept is treated as a random component, allowing for variation between classrooms, while the coefficients (or slopes) for the independent variables (X1: self-confidence and X2: spatial geometry test results) are treated as fixed. This setup allows us to explore the effect of classroom differences on students' critical thinking skills in geometry.

**Table 10.** Model 3 overview

| Info | |
|---|---|
| **Estimate** | **Linear mixed model fit by REML** |
| Call | (Y) Critical thinking skills ~ 1 + (X1) Self-confidence + (X2) Geometry test results + (1 \| Classroom no) |
| AIC | 753.091 |
| BIC | 767.087 |
| LogLikel. | -372.132 |
| R-squared marginal | 0.111 |
| R-squared conditional | 0.145 |
| Converged | Yes |
| Optimizer | Bobyqa |

Table 10 shows the fit statistics for model 3 with an AIC of 753.091 and a BIC of 767.087. This model includes individual-level variables (X1: self- confidence and X2: geometry test results) and takes into account the random effect of classrooms (level 2). The R-squared marginal is 0.111 indicating that 11.1% of the variance in critical thinking skills is explained by the individual-level predictors (X1, X2). The R-squared conditional is 0.145, showing that 14.5% of the total variance is explained when accounting for both individual and classroom-level effects.

**Table 11.** Random components

| Groups | Name | SD | Variance | ICC |
|---|---|---|---|---|
| No class | Intercept | 2.340 | 5.470 | 0.038 |
| Residual | Residual | 11.680 | 136.44 | - |

Table 11 provides the random components of model 3. The Intra-Class Correlation Coefficient (ICC) is 0.038, meaning that 3.85% of the total variance in critical thinking skills is attributed to differences between classrooms (level 2). This is a slight increase compared to the null model; it still suggests that most of the variance lies at the individual level rather than the classroom level. The standard deviation (SD) for the random intercept (classroom-level effect) is 2.340 with a corresponding variance of 5.470, and the residual variance is 136.44.

**Table 12.** Fixed effects parameter estimates

| Names | Confidence interval (95% ) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Estimate** | **SE** | **Lower** | **Upper** | **df** | **t** | **p** |
| Intercept | 77.917 | 2.038 | 73.922 | 81.912 | 0.992 | 38.230 | 0.017 |
| (X1) | 0.410 | 0.195 | 0.0291 | 0.792 | 92.000 | 2.110 | 0.038 |
| (X2) | 0.254 | 0.112 | 0.0343 | 0.475 | 92.610 | 2.270 | 0.026 |

Table 12 shows the fixed effects of model 3. Both independent variables have statistically significant effects on critical thinking skills, X1 ( self- confidence). The estimate is 0.410 with a p-value of 0.038 indicating that self-confidence has a significant positive effect on students' critical thinking skills in geometry, X2 ( geometry test results). The estimate is 0.254 with a p-value of 0.026 meaning that spatial geometry test results also have a significant positive impact on critical thinking skills.

In model 3, incorporating random intercepts allows us to examine classroom-level differences while controlling for individual-level factors (self-confidence and geometry test results). The results indicate that while classroom differences account for a small proportion of the variance in critical thinking skills (ICC = 0.038), individual-level factors, such as self-confidence and geometry test results significantly predict critical thinking skills.

This model improves over the null model ( model 2) with a lower AIC and BIC suggesting a better fit by including the fixed effects of the individual-level predictors. The random intercept for classrooms indicates that students' critical thinking skills vary slightly between classrooms though most of the variation is explained by individual characteristics.

### *4.4. Model 4: Multilevel Regression with Random Slopes*

Model 4 is a multilevel regression model that treats the independent variables as random slopes, unlike model 3, which considers the independent variables as fixed effects. This model also includes a level 1 variable (students), but the coefficients (slopes) of the independent variables are allowed to vary across classrooms (level 2). The estimation was done using the Restricted Maximum Likelihood (REML) method.

**Table 13.** Model 4 overview

| **Info** | |
|---|---|
| **Estimate** | **Linear mixed model fit by REML** |
| Call | (Y) Critical thinking skills ~ 1 + (X1) Self-confidence + (X2) Geometry test results + (1 + (X1) Self-confidence + (X2) Geometry test results \| Classroom No) |
| AIC | 760.999 |
| BIC | 787.077 |
| LogLikel. | -370.717 |
| R-squared marginal | 0.096 |
| R-squared conditional | 0.194 |
| Converged | Yes |
| Optimizer | Bobyqa |

Table 13 shows that model 4's AIC is 760.999 and its BIC is 787.077 indicating that this model is less optimal than model 3 in explaining the variance. The R-squared marginal is 0.096 meaning that 9.56% of the variance in critical thinking skills is explained by the level 1 predictors (X1 and X2). The R-squared conditional is 0.194 suggesting that 19.35% of the variance is explained when accounting for both individual and classroom-level effects.

**Table 14.** Random components

| **Groups** | **Name** | **SD** | **Variance** | **ICC** |
|---|---|---|---|---|
| No class | Intercept | 2.587 | 6.695 | 0.0486 |
| | (X1) | 0.160 | 0.026 | - |
| | (X2) | 0.243 | 0.059 | - |
| Residual Level | Residual | 11.452 | 131.158 | - |

Table 14 provides the random effects for model 4. The Intraclass Correlation Coefficient (ICC) is 0.0486, meaning that 4.86% of the total variance in critical thinking skills is attributed to differences between classrooms. This indicates a slightly larger influence of classroom-level effects compared to previous models, although the effect is still relatively small. The standard deviations for the random slopes of X1 ( self- confidence) and X2 ( geometry test results) are 0.160 and 0.243, respectively. This indicates that the slopes for these variables vary slightly across classrooms.

**Table 15.** Fixed effects parameter estimates

| Names | 95% confidence interval | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Estimate** | **SE** | **Lower** | **Upper** | **Df** | **t** | **p** |
| Intercept | 78.124 | 2.173 | 73.865 | 82.382 | 1.070 | 35.950 | 0.014 |
| (X1) | 0.426 | 0.222 | -0.0091 | 0.861 | 2.150 | 1.920 | 0.186 |
| (X2) | 0.215 | 0.204 | -0.1855 | 0.615 | 1.040 | 1.050 | 0.479 |

Table 15 shows the fixed effects of model 4. The results reveal X1 (self- confidence). The estimate is 0.426, but with a p-value of 0.186 which means the effect is not statistically significant. Thus, self-confidence does not have a significant impact on critical thinking skills in this model. X2 ( geometry test results): The estimate is 0.215 but the p-value is 0.479, indicating that geometry test results also do not have a significant impact on critical thinking skills in this model. These findings differ significantly from previous models, as both X1 and X2 show no significant effects on the dependent variable in model 4.

### 4.5. Model Selection: Choosing the Best Model

When selecting the best model, both AIC and BIC are used as criteria. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) help evaluate how well the model explains the variance while penalizing complexity. The model with the lowest AIC and BIC values is considered the best fit.

**Table 16.** Comparison of AIC and BIC values

| Model number | Model type | AIC | BIC |
|---|---|---|---|
| 1 | Multiple linear regression | 751 | 762 |
| 2 | Multilevel null model (No independent variables) | 760.342 | 765.546 |
| 3 | Multilevel regression with random intercept | 753.091 | 767.087 |
| 4 | Multilevel regression with random slope | 760.999 | 787.077 |

Table 16 shows that model 1 (Multiple Linear Regression) has the lowest AIC and BIC values, making it the best model for this analysis. The formula for model 1 is as follows:

$$\hat{Y} = 31.510 + 0.41X_1 + 0.241X_2$$

Where

$\hat{Y}$ = predicted critical thinking skills.

$X_1$= self- confidence.

$X_2$= geometry test results.

These results suggest that multilevel regression models do not necessarily outperform multiple linear regression, especially when the grouping variable (classrooms) is limited as is the case here with only two classrooms. The ICC values are small, indicating that the contribution of classroom-level effects to the total variance is minimal. Therefore, in this case, a simpler multiple linear regression model provides a better fit.

## 5. Discussion

### 5.1. Model 1: Multiple Linear Regression

The results of the multiple linear regression analysis indicate that X1 ( student self- confidence) has a p-value < 0.05. The results indicate that self-confidence has a significant impact on students' critical thinking skills in geometry. A lower p-value for this variable suggests that students with higher self-confidence tend to demonstrate stronger critical thinking abilities. Similarly, student performance on geometry tests also shows statistical significance implying that better test results are associated with improved critical thinking skills.

These findings align with previous studies research by Harjo, Kartowagiran, and Mahmudi (2019), Scriven and Paul (2007) and Shaheen (2016) which highlight the link between self-confidence, academic performance, and higher-order thinking skills. Research has consistently shown that students who believe in their abilities often perform better academically and develop stronger analytical and reasoning skills.

The significance of both self-confidence and academic achievement emphasizes the crucial role these factors play in fostering critical thinking. This suggests that building students' confidence in their abilities by reinforcing their academic foundation can contribute to the development of higher-order thinking skills, particularly in subjects like geometry.

### 5.2. Model 2: Multilevel Null Model ( without Independent Variables)

To assess whether classroom differences influence critical thinking skills in geometry, the multilevel null model ( model 2) was applied without incorporating individual-level variables. This model helps determine how much of the variation in critical thinking skills can be attributed to classroom factors by calculating the Intraclass Correlation Coefficient (ICC).

The ICC result was 0.0155, meaning that only 1.55% of the total variance in critical thinking skills is linked to classroom differences. While this indicates some level of classroom influence, the effect is quite small. A higher ICC value would typically suggest a stronger impact of classroom grouping, but in this study, the low ICC could be due to the limited number of classrooms (only two), reducing the likelihood of detecting a substantial effect.

Additionally, the AIC and BIC values for model 2 were higher than those for model 1, suggesting that the multiple linear regression model (model 1) provides a better overall fit. This indicates that considering individual-level factors, such as self-confidence and test performance, leads to a more accurate prediction of critical thinking skills than relying solely on classroom-level influences.

The comparison between models 1 and 2 reinforces the idea that individual student factors play a more significant role in shaping critical thinking skills. The inclusion of self-confidence and test results in model 1 substantially improves its predictive power, while model 2, which only considers classroom differences, accounts for a much smaller portion of the variation. Given the small ICC value, the influence of classroom factors appears to be minimal, which may be due to the small number of classrooms analyzed. This suggests that studies examining classroom-level effects should ensure a sufficiently large sample at this level to draw more reliable conclusions.

### 5.3. Model 3: Multilevel Regression with Random Intercept

A multilevel regression model with a random intercept was applied to evaluate the influence of classroom differences on students' critical thinking skills in geometry. In this model, the independent variables were based on individual student data while the intercept varied across classrooms to account for potential group-level differences. However, the slopes remained fixed, assuming a consistent relationship between the independent and dependent variables across all classrooms. The estimation process was carried out using the Restricted Maximum Likelihood (REML) method to ensure accurate parameter estimation.

In model 3, X1 ( student self- confidence) has a p-value < 0.05 indicating a statistically significant effect on Y ( critical thinking skills in geometry). This suggests that higher self-confidence among students leads to better critical thinking skills. Similarly, X2 ( student geometry test results) also shows a p-value < 0.05, demonstrating a significant positive effect on critical thinking skills. These results align with previous research by Hong et al. (2021), Jasiulewicz–Kaczmarek et al. (2021) and Wulandari, Rochmad, and Sugianto (2020) which found that students with higher self-confidence tend to perform better in their studies and develop stronger higher-order thinking skills, including critical thinking.

However, despite the significance of both X1 and X2 in this model, the AIC and BIC values for model 3 are higher than those of model 1 ( multiple linear regression). This indicates that model 1 still provides a better fit for predicting critical thinking skills. In other words, although model 3 accounts for classroom differences, its predictive power is inferior to that of the simpler multiple regression model. This suggests that, in this case, classroom-level variations may not contribute substantially to the prediction of critical thinking skills, reinforcing the idea that individual factors (such as self-confidence and test performance) are more crucial in shaping these skills.

### *5.4. Model 4: Multilevel Regression with Random Coefficients (Slopes)*

Model 4 is a multilevel regression model in which the independent variables, X1 ( student self- confidence) and X2 ( geometry test results) are treated as random coefficients (slopes). This is different from model 3 where the slopes were fixed. In model 4, the relationship between the independent variables and the dependent variable (critical thinking skills) is allowed to vary across classrooms, providing more flexibility in capturing potential classroom-level differences. The independent variables from level 1 (students) are still included but their effects are now allowed to change depending on the classroom.

However, the results in Table 15 show that X1 ( student self- confidence ) has a p-value > 0.05, indicating that it does not have a significant effect on Y (critical thinking skills ) in this model. Similarly, X2 (geometry test results ) also has a p-value > 0.05 suggesting that it does not significantly influence critical thinking skills. These findings are in stark contrast to the results from models 1, 2, and 3 where both X1 and X2 were significant predictors of critical thinking skills. In model 4, the inclusion of random slopes leads to the conclusion that neither X1 nor X2 has a statistically significant impact on Y.

Moreover, the AIC and BIC values for model 4 are higher than those of model 1, further indicating that model 1 remains the best predictive model for critical thinking skills in this context. The lack of significance for X1 and X2 in model 4 suggests that allowing the coefficients to vary across classrooms may not provide additional explanatory power. It could also imply that the differences in the relationships between the independent variables and the dependent variable across classrooms are not substantial enough to justify the complexity of a random slope model.

The comparison between models 3 and 4 highlights the potential limitations of multilevel modeling when applied to small group-level effects. In Model 3, where the intercept is treated as random, both X1 and X2 remain significant predictors of critical thinking skills. However, model 4, which allows for random slopes shows that the effects of X1 and X2 become non-significant, and the model's overall fit (as indicated by AIC and BIC) worsens.

Ultimately, model 1 ( multiple linear regression) remains the best model for predicting critical thinking skills in this study. Its simplicity combined with the significant effects of self-confidence and geometry test performance, makes it a more effective predictive tool compared to the more complex multilevel models. Additionally, the small ICC values and the limited number of classrooms in level 2 suggest that classroom-level differences do not play a major role in this case, further supporting the use of a simpler regression model over more elaborate multilevel approaches.

### *5.5. The Selection of the Best Regression Model*

After comparing the analyses from models 1, 2, 3, and 4, the best model for this case is model 1, which is the multiple linear regression model. This outcome suggests that multilevel regression does not always outperform multiple linear regression, even when accounting for level 2 factors such as class grouping (Gagné & Dayton, 2002; Sanquetta et al., 2018). A key consideration in multilevel modeling is the number of groups at level 2 and the Intraclass Correlation Coefficient (ICC) (Arnold, 2010). In this case, the level 2 grouping (classrooms) consists of only two classes which may be too few to adequately capture any significant class-level differences in the outcome variable.

The ICC values obtained in this study are relatively low, suggesting that classroom differences contribute only minimally to the overall variance. For instance, in the null model, the ICC is 0.0155, indicating that classroom-level factors account for just 1.55% of the total variance. This finding implies that the effect of classroom grouping on critical thinking skills is relatively minor. Similarly, low ICC values in subsequent models reinforce the conclusion that classroom differences have a limited influence on variations in students' critical thinking abilities.

In contrast, model 1 provides a more straightforward and effective prediction of critical thinking skills without the complexity of accounting for classroom groupings. Based on Table 6, the fitted regression equation for model 1 is as follows:

$$\hat{Y} = 31.510 + 0.41X_1 + 0.241X_2$$

Where

$\hat{Y}$ is the predicted critical thinking skills.

$X_1$ represents student self-confidence. $X_2$ represents student geometry test results.

This equation shows that both self-confidence (X1) and test results (X2) positively contribute to predicting students' critical thinking skills. The coefficients indicate that for every unit increase in self-confidence (X1), critical thinking skills increase by 0.41 units, and for every unit increase in test results (X2), critical thinking skills increase by 0.241 units.

In summary, although multilevel models offer a more complex approach by incorporating classroom-level differences, the limited number of level 2 groups and the low ICC values in this study suggest that these differences have a negligible effect on critical thinking skills. Therefore, model 1 ( multiple linear regression) proves to be the most effective and parsimonious model for predicting students' critical thinking skills without the need for additional multilevel structure.

## 6. Conclusion

This study demonstrates that self-confidence ( affective factor) and mathematical literacy in geometry test scores (cognitive factor) have a significant influence on critical thinking skills. However, gender does not show a significant effect. Additionally, the findings reveal that the AIC and BIC values of the multilevel regression model are higher than those of the multiple linear regression model. Since lower AIC and BIC values indicate better model performance due to reduced residual variance, it can be concluded that the multiple linear regression model is more effective than the multilevel regression model for this particular case.

The superiority of the multiple linear regression model in this study is primarily due to the limited number of level 2 groupings, as only two classes were analyzed. Additionally, the Intraclass Correlation Coefficient (ICC) indicates that class groupings contributed minimally to the total variance. Therefore, multilevel regression models are more appropriate when there are a substantial number of level 2 groupings and when ICC values show a significant contribution to variance. This ensures that the multilevel approach is justified and effectively captures group-level effects.

These findings provide valuable insights for future research. First, they highlight the importance of having a sufficient number of level 2 groupings in multilevel models to accurately assess group-level influences. Researchers planning to use multilevel regression should ensure a larger second-level sample to improve model accuracy and reliability. Additionally, this study emphasizes the role of self-confidence and cognitive ability in developing critical thinking skills. Future research could explore these relationships further by incorporating more diverse populations or different educational settings to enhance generalizability. Examining other affective and cognitive factors may also provide deeper insights into what shapes critical thinking leading to improved educational interventions and assessment strategies.

## 7. Recommendations

Based on these findings, several recommendations can be made. First, future research using multilevel regression should include a greater number of level 2 groupings to obtain more accurate and reliable results. A larger and more diverse sample is also recommended for studies investigating the relationship between self-confidence, cognitive abilities, and critical thinking, as this would allow for a more comprehensive analysis across different educational settings. Additionally, further research should explore other affective and cognitive factors, such as motivation and learning experiences to gain a broader understanding of the elements that influence critical thinking development. Lastly, educational approaches should be designed to enhance both students' self-confidence and cognitive skills, particularly in geometry and related subjects, as these factors have been shown to contribute significantly to critical thinking.

## 8. Limitations

Despite its contributions, this study has several limitations. The sample size was small as only two classes were analyzed in the multilevel regression model which limits the generalizability of the findings. Future research should include a larger and more diverse sample to improve external validity. Additionally, the study focused solely on geometry test performance as a measure of cognitive ability which may not fully capture critical thinking skills across different subjects. Future studies should consider a broader range of disciplines to gain a more comprehensive understanding of critical thinking development. Another limitation is the use of multilevel regression models with only two groups, resulting in low ICC values, which indicate minimal class-grouping effects. Future studies should involve more groups and a more in-depth multilevel analysis to better assess the relevance of this model. Lastly, this study focused only on self-confidence and cognitive abilities while other important factors, such as socioeconomic background, learning experiences, and social skills, were not explored. Future research should consider these additional variables for a more holistic understanding of what influences critical thinking skills.

## References


Akbari, O., & Sahibzada, J. (2020). Students' self-confidence and its impacts on their learning process. *American International Journal of Social Science Research, 5*(1), 1-15. https://doi.org/10.46281/aijssr.v5i1.462

Arends, R. I., & Kilcher, A. (2010). *Teaching for student learning*. New York: Routledge, Taylor & Francis Group.

Arnold, T. W. (2010). Uninformative parameters and model selection using Akaike's information criterion. *The Journal of Wildlife Management, 74*(6), 1175-1178.

Chua, Y. P. (2014). *Regression, factor analysis and structural equation modelling*. New York: McGraw-Hill Education.

Cirillo, M., & Hummer, J. (2021). Competencies and behaviors observed when students solve geometry proof problems: An interview study with smartpen technology. *ZDM–Mathematics Education, 53*, 861-875. https://doi.org/10.1007/S11858-021-01221-W

Cottrell, S. (2005). *Critical thinking skills*. Basingstoke: Palgrave Macmillan.

Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and psychological assessment: Principles and applications*. San Diego: Academic Press.

Csapodi, C., & Hoffmann, M. (2021). Changes in mathematics core curriculum and matriculation exam in the light of the COVID-19-shock. *Education Sciences, 11*(10), 610. https://doi.org/10.3390/EDUCSCI11100610

Denis, D. J. (2019). *Analysis for univariate, bivariate, and multivariate statistics*. Hoboken: John Wiley & Sons, Inc.

Denis, D. J. (2020). *Univariate, bivariate, and multivariate statistics using R: Quantitative tools for data analysis and data science*. Hoboken: John Wiley & Sons, Inc.

Dyer, E. B., & Sherin, M. G. (2016). Instructional reasoning about interpretations of student thinking that supports responsive teaching in secondary mathematics. *ZDM, 48*(1), 69-82. https://doi.org/10.1007/S11858-015-0740-1

Ennis, R. H. (1996). Critical thinking dispositions: Their nature and assessability. *Informal Logic, 18*(2), 165-182.

Gagné, P., & Dayton, C. M. (2002). Best regression model using information criteria. *Journal of Modern Applied Statistical Methods, 1*(2), 479–488. https://doi.org/10.22237/jmasm/1036110180

Gudono. (2017). *Analisis data multivariat*. Yogyakarta: BPFE-Yogyakarta.

Hair, J. F., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Boston: Cengage Learning.

Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *REiD (Research and Evaluation in Education), 4*(2), 126-135. https://doi.org/10.21831/reid.v4i2.22089

Handoko, S. T., Mardiati, Y., Ismail, R., & Imawan, O. R. (2023). Employing higher order thinking skills-based instruction in history course: A history high school teacher's perspective. *AIP Conference Proceedings, 2679*(1), 070004. https://doi.org/10.1063/5.0127631

Harjo, B., Kartowagiran, B., & Mahmudi, A. (2019). Development of critical thinking skill instruments on mathematical learning high school. *International Journal of Instruction, 12*(4), 149-166. https://doi.org/10.29333/iji.2019.12410a

Hong, J.-C., Hsiao, H.-S., Chen, P.-H., Lu, C.-C., Tai, K.-H., & Tsai, C.-R. (2021). Critical attitude and ability associated with students' self-confidence and attitude toward "predict-observe-explain" online science inquiry learning. *Computers & Education, 166*, 104172. https://doi.org/10.1016/J.COMPEDU.2021.104172

Imawan, O. R., Ismail, R., Tandililing, P., & Aisyah, F. N. (2023). Application of cooperative learning model type of teams games tournament using question card learning media on sine and cosine rules. *AIP Conference Proceedings, 2811*(1), 020002.

Imawan, O. R., Retnawati, H., Haryanto, & Ismail, R. (2024). Confirmatory factor analysis and differential item functioning analysis on mathematical literacy instruments for prospective Indonesian elementary school teachers. *AIP Conference Proceeding*, *3150*(1), 080009. https://doi.org/10.1063/5.0228174

Imawan, O. R., Retnawati, H., Haryanto, H., & Ismail, R. (2025). Innovations in assessment methods: Computerized adaptive testing (CAT) for sustainable energy efficiency. In Lecture Notes in Civil Engineering. In (2nd ed., Vol. 557, pp. 161–168): Springer Nature Link. https://doi.org/10.1007/978-981-97-8348-9_20

Ismail, R., Imawan, O. R., & Nadhifah, A. (2023). The influence of interest in learning mathematics on mathematics learning achievement of junior high school students in Papua. *AIP Conference Proceedings*, *2811*(1), 020013. https://doi.org/10.1063/5.0142257

Ismail, R., Retnawati, H., Sugiman, & Imawan, O. R. (2024). Construct validity of mathematics high order thinking skills instrument with cultural context: Confirmatory factor analysis. *AIP Conference Proceeding*, *3150*(1), *080008*. https://doi.org/10.1063/5.0228143

Ismail, R., Retnawati, H., Sugiman, & Imawan, O. R. (2025). A sustainable development approach to math higher-order thinking skills: Culture and green technology. In Lecture Notes in Civil Engineering. In (2nd ed., Vol. 557, pp. 137–143): Springer Nature Link. https://doi.org/10.1007/978-981-97-8348-9_17

Jasiulewicz–Kaczmarek, M., Antosz, K., Żywica, P., Mazurkiewicz, D., Sun, B., & Ren, Y. (2021). Framework of machine criticality assessment with criteria interactions. *Eksploatacja i Niezawodność*, *23*(2), 207-220. https://doi.org/10.17531/ein.2021.2.1

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education Inc.

Kemp, J. E., Morrison, G. R., & Ross, S. M. (1994). *Designing effective instruction*. New York: Macmillan College Publishing Company.

Kiverstein, J., Rietveld, E., Slagter, H. A., & Denys, D. (2019). Obsessive compulsive disorder: A pathology of self-confidence? *Trends in Cognitive Sciences*, *23*(5), 369-372. https://doi.org/10.1016/J.TICS.2019.02.005

Kuntze, S., Aizikovitsh-Udi, E., & Clarke, D. (2017). Hybrid task design: Connecting learning opportunities related to critical thinking and statistical thinking. *ZDM*, *49*, 923-935. https://doi.org/10.1007/S11858-017-0874-4

Liestari, S. P., & Muhardis, M. (2021). Hierarchical linear modeling for determining the effect of ICT literacy on mathematics achievement. *REID (Research and Evaluation in Education)*, *7*(1), 78-87. https://doi.org/10.21831/reid.v7i1.39181

Liu, X., Xu, Y., Montes, R., Dong, Y., & Herrera, F. (2019). Analysis of self-confidence indices-based additive consistency for fuzzy preference relations with self-confidence and its application in group decision making. *International Journal of Intelligent Systems*, *34*(5), 920-946. https://doi.org/10.1002/INT.22081

McElmeel, S. L. (2002). *Character education*. Westport, CT: Greenwood Publishing Group, Inc.

Mujiyati, N., Warto, W., & Sutimin, L. A. (2019). Developing a problem-based local history module to improve the critical thinking ability of senior high school students. *REID (Research and Evaluation in Education)*, *5*(1), 30-40. https://doi.org/10.21831/reid.v5i1.13334

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Neukrug, E. S., & Fawcett, R. C. (2015). *Essentials of testing and assessment* (3rd ed.). Boston: Cengage Learning.

OECD. (2022). *PISA 2022 mathematics framework*. Paris: OECD Publishing.

Orlich, D. C., Harder, J. R., & Calahan, R. C. (2007). *Teaching strategies: A guide to effective instruction*. Boston: Houghton Mifflin Company.

Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, *5*, 111-128. https://doi.org/10.1016/J.IDM.2019.12.010

Purnomo, Y. W., Arifin, P. P., Rahmawati, R. D., & Rahmawati, A. P. (2024). Children's mathematics-related beliefs: An exploratory qualitative study in an Indonesian elementary school. *Issues in Educational Research*, *34*(3), 1148-1169.

Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2018). Higher order thinking skillss and academic performance in physics of college students: A regression analysis. *International Journal of Innovative Interdisciplinary Research*, *20*(1), 48-60.

Rencher, A. C. (2001). *Methods of multivariate analysis* (2nd ed.). New York: John Wiley & Sons, Inc.

Rosnawati, R., Kartowagiran, B., & Jailani, J. (2015). A formative assessment model of critical thinking in mathematics learning in junior high school. *REiD (Research and Evaluation in Education)*, *1*(2), 186-198. https://doi.org/10.21831/reid.v1i2.6472

Rott, B., & Rott, B. (2021). Inductive and deductive justification of knowledge. *Epistemological Beliefs and Critical Thinking in Mathematics: Qualitative and Quantitative Studies with Pre-Service Teachers*, *106*(2), 121-145.

Sanquetta, C. R., Dalla Corte, A. P., Behling, A., de Oliveira Piva, L. R., Péllico Netto, S., Rodrigues, A. L., & Sanquetta, M. N. I. (2018). Selection criteria for linear regression models to estimate individual tree biomasses in the Atlantic Rain Forest, Brazil. *Carbon balance and management*, *13*, 1-15. https://doi.org/10.1186/S13021-018-0112-6

Scriven, M., & Paul, R. (2007). *Defining critical thinking. The Critical Thinking Community: Foundation for Critical Thinking. The Critical Thinking Community: Foundation for Critical Thinking*. Retrieved from https://www.criticalthinking.org/pages/defining-critical-thinking/766

Shaheen, N. (2016). International students' critical thinking–related problem areas: UK university teachers' perspectives. *Journal of Research in International Education*, *15*(1), 18-31. https://doi.org/10.1177/1475240916635895

Snipes, M., & Taylor, D. C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices. *Wine Economics and Policy*, *3*(1), 3-9. https://doi.org/10.1016/J.WEP.2014.03.001

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences, by James P. Stevens. In International Statistical Review* (5th ed.). New York: Routledge.

Subali, B., Paidi, P., & Mariyam, S. (2017). Mapping elementary school students' creativity in science process skills of life aspects viewed from their divergent thinking patterns. *REiD (Research and Evaluation in Education)*, *3*(1), 1-11. https://doi.org/10.21831/reid.v3i1.13294

Sumin, S., Sukmawati, F., & Nurdin, N. (2022). Gender differential item functioning on the Kentucky Inventory of Mindfulness Skills instrument using logistic regression. *REID (Research and Evaluation in Education)*, *8*(1), 55-66. https://doi.org/10.21831/reid.v8i1.50809

Tabachnik, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Harlow, England: Pearson Education Limited.

Tinsley, H. E. A., & Brown, S. D. (2000). *Handbook of applied multivariate statistics and mathematical modeling. In Handbook of Applied Multivariate Statistics and Mathematical Modeling*: Elsevier Science & Technology Books. https://doi.org/10.1016/b978-0-12-691360-6.x5000-9

Tiruneh, D. T., De Cock, M., & Elen, J. (2018). Designing learning environments for critical thinking: Examining effective instructional approaches. *International Journal of Science and Mathematics Education*, *16*, 1065-1089. https://doi.org/10.1007/S10763-017-9829-Z

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, *15*, 663-682. https://doi.org/10.1007/S10763-016-9723-0

Ulwatunnisa, M., Retnawati, H., Muhardis, M., & Yusron, E. (2023). Revealing the characteristics of Indonesian language test used in the national-standardized school examinations. *REID (Research and Evaluation in Education)*, *9*(2), 210-222. https://doi.org/10.21831/reid.v9i2.31999

West, B. T., Welch, K. B., & Galecki, A. (2006). *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: Chapman & Hall/CRC.

Wulandari, I. P., Rochmad, R., & Sugianto, S. (2020). Integrated between DAPIC problem solving model and RME approach to enhance critical thinking ability and self confidence. *Anatolian Journal of Education*, *5*(2), 73–84. https://doi.org/10.29333/aje.2020.526a