JPM
Jurnal Pendidikan Matematika
UNIVERSITAS ISLAM MALANG

ARTICLE INFORMATION

Students' mathematical critical thinking during project-based probability learning: A repeated-measures study

Febi Widarwati, Dian Devita Yohanie, Darsono
Corresponding author: febiwidarwati23365@gmail.com

How to cite (APA Style):
Widarwati, F., Yohanie, D. D., & Darsono. (2026). Students' mathematical critical thinking during project-based probability learning: A repeated-measures study. Jurnal Pendidikan Matematika (JPM), 12(2), 248–260. https://doi.org/10.33474/jpm.v12i2.25861

DOI	https://doi.org/10.33474/jpm.v12i2.25861
Contributions	**Febi Widarwati:** CCP, MTD, DAC, ANL, WOD, WRE. **Dian Devita Yohanie**: SPV, INV, VAL, WRE. **Darsono**: SPV, VAL, WRE.
Funding	This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Conflict of Interest	The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.
Copyright ©	2026 Febi Widarwati, Dian Devita Yohanie, Darsono (the authors).
Declaration of AI Use	The authors declare that generative artificial intelligence (AI) tools were used solely to assist with language editing, grammar checking, clarity, and/or organization of the manuscript. AI tools were not used to generate research data, conduct statistical analyses, formulate research findings, or make scientific conclusions. The authors take full responsibility for the accuracy, originality, integrity, and final content of the manuscript and have reviewed and verified all AI-assisted content.

*** CRediT roles:** CCP: Conceptualization; VAL: Validation; MTD: Methodology; INV: Investigation; DAC: Data Curation; ANL: Formal Analysis; WOD: Writing - Original Draft; WRE: Writing - Review & Editing; SPV: Supervision.

JPM
JURNAL PENDIDIKAN MATEMATIKA (JPM)
Vol. 12, No. 2, 2026, pp 248–260 e-ISSN: 2656–4564
https://doi.org/10.33474/jpm.v12i2.25861
online at: https://riset.unisma.ac.id/index.php/jpm/article/view/25861

Students' mathematical critical thinking during project-based probability learning: A repeated-measures study

Febi Widarwati*, Dian Devita Yohanie, Darsono
Universitas Nusantara PGRI Kediri, Indonesia
Email: *febiwidarwati23365@gmail.com*

ABSTRACT

In the 21st century, mathematical critical thinking is a core competency that students must master. However, high school students still struggle to analyze word problems in mathematics. This study examines how students' mathematical critical thinking skills develop over three Project-Based Learning (PjBL) sessions on probability. The study employs a quantitative approach with a pre-experimental, single-group repeated measures (time-series) design, involving 35 10th-grade students at SMAN 3 Kediri (Public Senior High School 3 Kediri). Data were collected through essay-based word-problem tests at the end of each session and assessed using a rubric based on the six FRISCO indicators (Focus, Reasoning, Inference, Situation, Clarity, and Overview), adapted from Ennis. The data were analyzed using the nonparametric Friedman and Wilcoxon Signed-Rank tests because the distributions of Test 2 and Test 3 scores deviated from normality. Average scores increased gradually from 59.86 (Test 1) to 61.61 (Test 2) and 69.04 (Test 3), with a statistically significant increase only observed from Test 2 to Test 3. At the indicator level, "Focus" and "Situation" increased sharply in the third session; "Reason" remained the strongest indicator, while "Inference" and "Overview" remained the weakest. These findings suggest that students' mathematical critical thinking skills developed gradually throughout the PjBL sequence, although the single-group design without a control group limits causal claims regarding PjBL's contribution. Teachers are advised to supplement PjBL with explicit scaffolding during the inference and answer evaluation stages.

Keywords: Project-Based Learning, Mathematical Critical Thinking, Probability, the FRISCO Framework, Repeated Measurement Design

*** Corresponding Author Email:** febiwidarwati23365@gmail.com

Submission	Revised	Accepted	Published
July 4, 2026	August 20, 2026	August 26, 2026	August 31, 2026

How to cite (in APA style):
Widarwati, F., Yohanie, D. D., & Darsono. (2026). Students' mathematical critical thinking during project-based probability learning: A repeated-measures study. *Jurnal Pendidikan Matematika (JPM)*, *12*(2), 248–260. https://doi.org/10.33474/jpm.v12i2.25861

INTRODUCTION

Twenty-first-century education aims to foster individuals' full potential in intellectual, moral, and skill-based domains (Ansya & Salsabilla, 2025). Mathematics plays a central role in this context, serving as a universal discipline that supports technological advancement and the development of cognitive capacities (Aizah & Dewi, 2024). As a result, mathematics instruction has shifted from repetitive calculation exercises to emphasizing the "4Cs" competencies, with critical thinking as a primary focus (Difinubun et al., 2022). In mathematics, critical thinking enables students to analyze, evaluate, and interpret information to make rational decisions (Maesaroh, 2021). Qohar and Sulandra (2021), further, critical thinking is defined as the capacity to solve real-life problems through structured reasoning, interpretation, analysis, and evaluation of information.

This study is based on Ennis's theory of critical thinking, which is generally summarized by the acronym FRISCO: Focus, Reason, Inference, Situation, Clarity, and Overview (Ennis, as cited in Setiana et al., 2020). The FRISCO framework was chosen because it captures, in an operational form, the elements that must conceptually be present for critical thinking to occur. According to Triani Putri et al. (2022)the six indicators in question can be elaborated as follows. Focus involves pinpointing the core issue embedded in a problem. Reasoning concerns the formulation of logically sound justifications. Inference denotes the capacity to derive conclusions that are appropriately warranted. Situation captures an awareness of the contextual circumstances underlying a given problem. Clarity reflects the exactness with which language and mathematical notation are employed. Finally, Overview entails a retrospective evaluation of the coherence exhibited throughout the entire problem-solving procedure.

The FRISCO framework is considered appropriate for probability material because numerically correct answers do not in themselves indicate the presence of critical thinking: students may obtain the correct numerical result but still err in identifying relevant events, selecting the appropriate rule (addition or multiplication), understanding the relationships between events, or interpreting the meaning of probability results in context. All six FRISCO indicators address these error patterns, from problem identification (Focus, Situation) and justification of rule selection (Reason) to drawing conclusions and contextual interpretation (Inference, Overview), making this framework well suited to probabilistic reasoning rather than mere computational accuracy.

Strengthening these indicators is increasingly urgent given Indonesia's performance in international assessments. The 2022 Program for International Student Assessment (PISA) ranked Indonesia 69th out of 81 participating countries, with the country's average mathematics score continuing to lag behind the OECD average (OECD, 2023). This gap is widely attributed to weak critical thinking skills among high school students, which are most evident when they are asked to solve word problems: tasks that require reading comprehension, logical reasoning, and the translation of real-world narratives into formal mathematical models before solutions can be attempted (Tunu et al., 2022). Difficulties at this translation stage have been well documented; students often misinterpret the information provided, fail to identify what is actually being asked, or apply memorized procedures without considering the context of the problem (Anisa et al., 2023).

One pedagogical response to this challenge is Project-Based Learning (PjBL), a student-centered model that engages students in solving authentic problems through structured project work (Azizah, 2022). PjBL gives students space to engage in constructive inquiry, independent exploration, and collaboration while producing tangible, authentic outcomes (Muhammad Rafik et al., 2022). Because PjBL requires students to answer real questions rather than merely imitate demonstrated procedures, this approach is believed to develop a broader range of competencies, including data analysis, problem-solving, communication, and teamwork, while connecting classroom learning to real-world practices (Amir & Solida, 2022). In this study, PjBL was implemented as a three-session project sequence: real-world context to project question, then data collection, then probability distribution to analysis of addition rules, then analysis of multiplication rules, and finally interpretation of results, report, presentation, and reflection. In the first session, students were given the challenge "Analysis of Probability in Everyday Life"; each group then independently determined its own context and formulated its project question and data needs. In the second session, the groups processed the collected data into probability distribution tables and identified events suitable for the addition rule. They identified independent events suitable for the multiplication rule and began drafting their reports. In the third session, the groups finalized their analysis into a final report and presented it; the assessment rubric used evaluated not only the accuracy of calculations but also the depth of analysis and the extent to which students connected probability concepts to the real-world contexts they had chosen. This structure distinguishes the intervention from a mere series of contextual probability problems, as it requires students to formulate their own investigative questions, collect and analyze authentic data, and produce and defend a substantive report. Several studies report that implementing PjBL is associated with measurable improvements in critical thinking and broader mathematics learning outcomes (Aini et al., 2022; Kristiyanto, 2020). On the other hand, PjBL also presents practical challenges: teachers generally cite the time required to design and manage project-based learning as the main obstacle to consistent implementation (Rosa et al., 2024).

Initial observations at SMAN 3 Kediri (Public Senior High School 3 Kediri) revealed a pattern consistent with the national picture. Students tend to be passive learners who claim to understand the material when it is explained but cannot apply that understanding independently to contextual word problems. Many struggle to distinguish relevant information from irrelevant information and cannot identify systematic problem-solving strategies. Probability is closely related to real-world, everyday situations. However, it remains challenging for students because it requires reasoning and decision-making under uncertainty rather than merely applying a single fixed formula.

While PjBL has been widely examined in relation to critical thinking, studies that monitor the progression of students' critical thinking skills through repeated measurements across consecutive sessions, rather than relying solely on pre-test/post-test designs, are relatively scarce, especially in the context of Probability (Fisher et al., 2020). This research addresses this gap by analyzing students' mathematical critical thinking across three PjBL sessions, using the FRISCO framework. By assessing performance after each session, the study seeks to elucidate not only whether PjBL enhances critical thinking, but also the trajectory of this enhancement and the relative development of each FRISCO indicator. Accordingly, the research question is: How do students' mathematical critical thinking skills, as measured by the FRISCO indicators, evolve over three consecutive Project-Based Learning sessions on Probability? Consistent with the repeated-measures design, the hypothesis posits that students' mathematical critical thinking scores will change significantly across the three sessions.

METHOD

This study employs a quantitative approach, characterized by the collection and statistical analysis of numerical data obtained through measurement, observation, and hypothesis testing, as described by Muhajirin et al., (2024). This study employs a one-group repeated-measures design, adapted from the time-series design described by Sugiyono (2019), in which the same group of students received the PjBL treatment over three sessions and was assessed at the end of each session, rather than just once before and once after the intervention. The study used no control group and no baseline measurement before the intervention; therefore, this design is classified as pre-experimental, not quasi-experimental, because it does not include a comparison group. These repeated measurements track the trajectory of change within the same group over time.

The study population consisted of all 10th-grade students at SMAN 3 Kediri (Public Senior High School 3 Kediri). The sample was selected through purposive sampling, in which class placement was based on the professional judgment of the mathematics teacher responsible for that grade level Sugiyono, (2019). Class 10-8, consisting of 35 students, was selected as the research subjects and received instruction using the PjBL model throughout the study.

The PjBL implementation took place over three sessions, each lasting 90 minutes (2 class periods). Class X-8 was divided into 7 working groups. In the first session, the teacher presented the project challenge, "Analysis of Probability in Everyday Life," as the driving question; each group then independently chose a real-world context to investigate and formulated its own investigative questions and data requirements. In the second session, each group processed the collected data into probability distribution tables, identified events that followed the addition rule, identified independent events for the multiplication rule, and began drafting their reports. In the third session, each group finalized and presented their report to the class; this session concluded with a reflection in the form of a class discussion, during which students and the teacher discussed the challenges they faced and the connection between probability concepts and the real-world contexts they had chosen. The final products of the project, a written report and a group presentation, are assessed using a rubric that takes into account the depth of analysis and the connection between probability concepts and real-world contexts, thereby distinguishing this intervention from a mere series of contextual probability problems.

This study was conducted at SMAN 3 Kediri (Public Senior High School 3 Kediri), East Java, during the even semester of the 2025–2026 academic year, from May 11 to 18, 2026. This period was chosen because it fell shortly after the semester break, when students had returned to their regular classroom routines and learning conditions could be considered stable. All stages of the study, from coordination with the school and the implementation of instruction to data collection, were completed within a one-month timeframe, with critical thinking tests administered at the end of each session.

Two categories of instruments were used: instructional materials (PjBL teaching modules) and data collection instruments (essay-based critical thinking tests).

The PjBL teaching modules were developed based on the Learning Objective Sequence derived from the applicable Learning Outcomes and designed to support the competencies of the Pancasila Student Profile (Difinubun et al., 2022). Before use, two expert validators validated these modules. As shown in Table 1, both validators rated the modules as "very good," indicating they are suitable for classroom implementation.

Table 1. Result of Teaching Module Validation

Validator	Score	Category
Validator 1 (Mathematics education lecturer)	83.00%	Very Good
Validator 2 (Subject teacher)	92.19%	Very Good

The mathematical critical thinking test consisted of essay questions based on probability word problems and was administered at the end of each of the three learning sessions. The questions were developed from the previous instrument, refined through consultation with the supervising professor, and their content validity was verified by a professor and a subject teacher at the research site. Students' answers were assessed using a rubric adapted from Ennis's FRISCO framework and modified by the researcher to suit the assessment needs of this study. Table 2 presents the complete rubric, with a scale of 0 to 4 for each indicator.

Table 2. Indicators of Students' Mathematical Critical Thinking Skills

Indicator	Description	Score
Focus	Writes the "Known" and "Asked" sections completely	4
	Writes the "Known" and "Asked" sections correctly, but one piece of data or condition is missing	3
	Writes the "Known" and "Asked" sections, but contains an error in identifying the data	2
	Only writes some data from the "Known" section without writing "Asked," or vice versa	1
	Does not write either the "Known" or "Asked" sections	0
Reason	Performs mathematical calculation steps coherently and correctly	4
	Calculation steps are correct and coherent, but there is a minor error in the final simplification, so the final result is slightly inaccurate	3
	The initial substitution is correct, but a significant calculation error occurs mid-process, resulting in a wrong final answer	2
	Only writes the values from the problem without performing further calculations	1
	Does not write any calculations at all	0
Inference	Writes a conclusion that answers the question precisely and contextually	4
	Writes a correct conclusion, but it is not fully linked to the context of the problem	3
	Writes a conclusion that is not quite contextually accurate	2
	Writes only the final result without a concluding statement	1
	Does not write a conclusion at all	0
Situation	Creates a complete and accurate solution model	4
	Creates an accurate solution model but does not complete it	3
	Creates an accurate solution model but it is incomplete	2
	Creates an incorrect solution model	1
	Provides no answer	0
Clarity	Clarity of symbols and terms: uses standard symbols precisely, accompanied by explanations of symbol meaning so they are unambiguous	4
	Uses correct symbols but without an explanation of meaning, or the explanation is incomplete	3
	Uses symbols that are non-standard or inconsistent, but the calculation intent can still be understood	2
	Does not use mathematical symbols; the answer is written in ordinary sentences without probability notation	1
	Does not write an answer	0
Overview	Verifies values and checks the consistency between solution steps thoroughly	4
	Performs checking, but does not check consistency thoroughly	3
	The checking performed contains errors	2
	No evaluation of the answer's correctness	1
	No checking or verification at all	0

The test results are calculated using the formula:

$$Score = \frac{score\ of\ all\ indicators\ obtained}{maximum\ score\ of\ all\ indicator} \times 100$$

The data were analyzed descriptively and inferentially using IBM SPSS Statistics version 25. Descriptive statistics (mean, median, standard deviation, minimum, and maximum) were calculated for each of the three tests to describe the overall distribution of scores. Before conducting inferential tests, we examined the assumption of normality using the Shapiro-Wilk test, which is appropriate given the sample size of fewer than 50 students (Anantasia & Rindrayani, 2023). Although the Shapiro-Wilk test for the difference in scores between sessions (Test 2 − Test 1: $p = 0.32$; Test3−Test2: $p = 0.40$) did not indicate a significant violation of normality, the nonparametric Friedman test was still chosen as a more conservative approach, given that the marginal distributions of Test 2 and Test 3 individually failed the normality test and the sample size was relatively small ($n = 35$) (Inayah et al., 2025). This approach also accounts for the nature of the rubric score data, which is a mixed ordinal-interval scale. If the Friedman test indicates a significant overall difference, paired comparisons follow with the Wilcoxon Signed-Rank test, the nonparametric version of the paired t-test used for two dependent samples (Widiana et al., 2025). Decisions are based on standard criteria for comparing p-values against a significance level of 0.05: a $p - value\ below\ 0.05\ indicates$ a statistically significant difference.

RESULT

Descriptive Analysis of Mathematics Critical Thinking Scores

Data on students' mathematical critical thinking were obtained from three assessments administered to 35 students in Class X-8. Table 3 summarizes the descriptive statistics for each test.

Table 3 Descriptive Statistics of the Three Tests

Tes	N	Rata-rata	Median	SD	Min	Maks
Tes 1	35	59,86	60,20	7,44	48,60	72,20
Uji 2	35	61,61	61,00	10,06	44,70	75,00
Tes 3	35	69,04	68,20	4,83	62,00	83,90

The average score rose from 59.86 on Test 1 to 61.61 on Test 2 and reached 69.04 on Test 3, indicating a steady upward trend across all sessions. The standard deviation widened from 7.44 on Test 1 to 10.06 on Test 2 before narrowing sharply to 4.83 on Test 3, indicating that student performance became much more homogeneous in the final session.

Students were also grouped into performance categories using two complementary categorization schemes. The first scheme classifies students relative to the sample mean and standard deviation for each test (High, Medium, Low); the second scheme, adapted from (Zahra & Hakim, 2022)classifies students based on fixed score ranges (Very Good, Good, Fair, Poor). Table 4 presents the distribution based on the mean and standard deviation scheme, while Table 5 presents the distribution across the three sessions based on the fixed-score-range scheme.

Table 4. Distribution of Students by Score Category

Category	Test 1	Test 2	Test 3
High	7/20,0%	8/22,9%	4/11,4%
Moderate	21/60,0%	19/54,3%	26/74,3%
Low	7/20%	8/22,9%	5/14,3%

Table 5. Distribution of Students by Score Category Across All Sessions

Score Range	Category	Test 1	Test 2	Test 3
≥ 75	Very Good	0 / 0.0%	1 / 2.9%	4 / 11,4%
$56 \leq score < 74$	Good	23 / 65.7%	21 / 60.0%	31 / 88.6%
$40 \leq score < 55$	Fair	12 / 34.3%	13 / 37.1%	0 / 0.0%
< 40	Poor	0 / 0.0%	0 / 0.0%	0 / 0.0%

Normality Test

We used the Shapiro-Wilk test to assess the normality of each test's score distribution before hypothesis testing. As shown in Table 6, only the distribution for Test 1 meets the assumption of normality ($Sig. = 0.055 > 0.05$). In contrast, the distributions for Test 2 and Test 3 deviate significantly from normality ($Sig. = 0.006$ and 0.013, respectively). Consequently, the nonparametric Friedman and Wilcoxon Signed-Rank tests were used for hypothesis testing, rather than repeated-measures ANOVA.

Table 6. Shapiro-Wilk Normality Test Results

Test	Statistic	df	Sig.
Test 1	0.940	35	0.055
Test 2	0.907	35	0.006
Test 3	0.918	35	0.013

Hypothesis Testing

A Friedman test was conducted to determine whether students' critical thinking scores differed overall across the three sessions. This test yielded a chi-square value of 23.543 ($df = 2, Asymp.Sig. = 0.000$), which is below the 0.05 threshold. Therefore, the null hypothesis of no difference was rejected, indicating a statistically significant difference in students' mathematical critical thinking scores across Test 1, Test 2, and Test 3 during PjBL implementation.

Because the Friedman test indicated an overall significant difference, the Wilcoxon Signed-Rank test was used to identify which specific session pairs differed significantly. To control for Type I error inflation resulting from the three paired comparisons derived from the same Friedman test, a Bonferroni correction was applied by multiplying each p-value by three ($\alpha = 0.05/3 = 0.0167$ as the effective threshold). Table 7 summarizes the results.

Table 7. Wilcoxon Signed-Rank Test (Pairwise Comparisons)

Comparison	Z	Sig. (2-tailed)	Sig. (Bonferroni correction)	Students Improving	Decision
Test 1 → Test 2	$-1{,}720$	$0{,}085$	0,256	20 of 35	Not significant
Test 2 → Test 3	$-3{,}448$	$0{,}001$	0,002	24 of 35	Significant
Test 1 → Test 3	$-4{,}824$	$0{,}000$	<0,001	33 of 35	Significant

The comparison between Test 1 and Test 2 did not reach statistical significance (Z = -1.720, Sig. = 0.085), although 20 of the 35 students increased their scores. In contrast, the comparison between Test 2 and Test 3 was statistically significant (Z = -3.448, Sig. = 0.001), with 24 out of 35 students showing improvement and the total positive score (525.50) far exceeding the total negative score (104.50). The comparison between Exam 1 and Exam 3 yielded the strongest results among the three (Z = -4.824, Sig. = 0.000), with 33 out of 35 students showing improvement (sum of positive ranks 69.50 vs sum of negative ranks 20.50 in the corresponding rank-based calculation). These patterns remained significant after the Bonferroni correction, confirming that improvement from the first session to the last was consistent and statistically robust.

FRISCO Indicator Development

In addition to the overall scores, student responses were analyzed by indicator to track how each of the six FRISCO components evolved over the course of the session. Table 7 and Figure 1 present the average scores (on a scale of 0–4) obtained for each indicator on each test.

Table 8. Mean Scores by FRISCO Indicator Across the Three Tests

Indicator	Test 1	Test 2	Test 3
Focus	3,13	3,00	3,94
Reason	3,99	3,93	3,94
Inference	1,00	1,02	1,14
Situation	3,13	3,00	3,94
Clarity	2,11	2,75	2,45
Overview	1,01	1,09	1,15

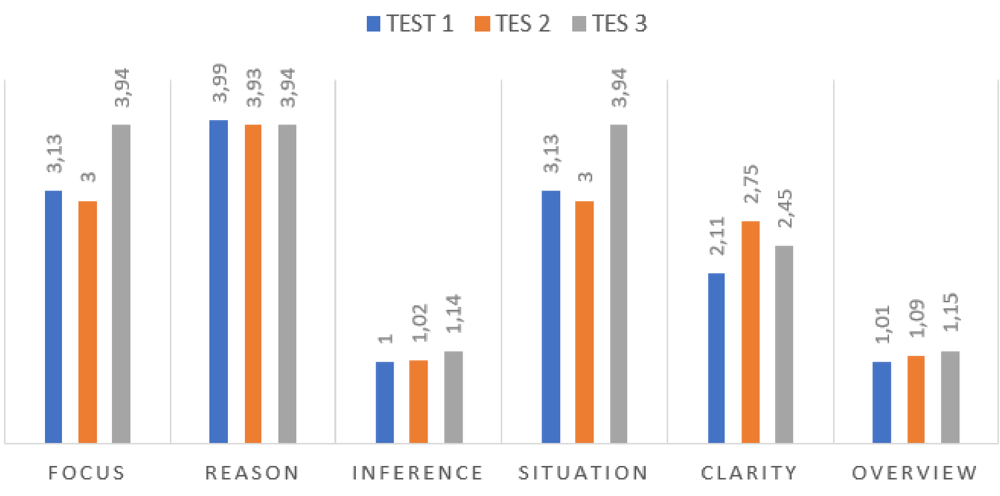

Figure 1. Average Scores by FRISCO Indicator Across the Three PJBL Sessions

The Reasoning indicator consistently recorded the highest average scores across all sessions (3.90–3.99), suggesting that students possessed strong step-by-step logical calculation skills from the outset. The Focus and Situation indicators stagnated on Test 2, then rose to 3.94 on Test 3, suggesting that proficiency in identifying core problems and constructing accurate solution models required more than one session. Clarity displayed a variable pattern, increasing from 2.11 to 2.75 between Test 1 and Test 2, then stabilizing at 2.45 in Test 3, reflecting increased experimentation with mathematical notation during presentations. Inference and Overview remained the lowest-scoring indicators, though both improved slightly on Test 3, indicating that drawing contextual conclusions and verifying final answers continued to pose significant challenges.

The findings presented must be substantiated by adequate data and should directly address the research hypothesis outlined in the introduction. The results section should summarize key scientific findings, emphasizing how these outcomes answer the research questions. Highlight differences between the present findings and those of previous studies. Organize detailed explanations into subsections, and present tables and figures centrally, with appropriate references in the manuscript.

DISCUSSION

The Trajectory of Mathematical Critical Thinking Development

Overall, the results indicate that implementing PjBL is associated with gradual, cumulative improvement in students' mathematical critical thinking. These findings are consistent with those of (Kristiyanto, 2020), who reported that PjBL can foster meaningful learning and develop critical thinking in a measurable, gradual manner rather than all at once. The increase in the average score from 59.86 on Test 1 to 69.04 on Test 3 is also consistent with the pattern reported by (Sherly et al., 2020) in their study on the effect of PjBL on students' ability to solve math word problems.

One notable finding is the absence of a statistically significant increase between Test 1 and Test 2, despite a numerical rise in the average score. This pattern is best understood as an adaptation phase, during which students are still adjusting to the demands of project-based independent inquiry. (Lapase, 2021) Also argues that project-based learning is not merely a series of problem-solving exercises, but rather a process that challenges students to seek answers on their own initiative. This adjustment inevitably takes time before its impact can be measured. The increase in the standard deviation on Test 2 (10.06, up from 7.44 on Test 1) further supports this interpretation. Some students appeared to adjust their scores faster than others during the second session, resulting in a more heterogeneous distribution of scores during this transitional phase. However, by Test 3, the standard deviation had narrowed sharply to 4.83, indicating that score improvements were beginning to distribute more evenly across the class. It was no longer concentrated among a specific group of students.

Indicator-Level Patterns and Comparison with Prior Research

An analysis of each FRISCO indicator reveals a more detailed picture than the aggregate score alone. The sharp increase in the "Focus" and "Situation" indicators in the third session aligns with the findings in Lestari & Lubis (2023)who also found that PjBL's emphasis on realistic and fundamental questions strengthens students' ability to filter relevant information and construct accurate situational models. Consistently strong performance on the "Reasoning" indicator is consistent with the notion that constructive inquiry in PjBL sessions provides a solid foundation for logical reasoning, even before students have fully mastered the other indicators.

Instead, "Inference" and "Overview" remained the weakest indicators throughout this study, a finding that differs from that of Lestari & Lubis 2023), who reported a relatively greater improvement in both indicators in their own PjBL intervention. One plausible explanation, consistent with classroom observations during this study, is that students tend to stop working once they obtain numerical results, without pausing to draw contextual conclusions or verify the validity of their answers. This pattern remains consistent with the broader theoretical claim put forward by Fisher et al., (2020), who argue that PjBL has great potential to foster higher-order critical thinking through reflection and authentic inquiry, a potential that, in this study, appears to require more sustained guidance during the inference and evaluation stages before it can be fully realized

Implications for Solving Word Problems

The overall pattern of score improvement observed in the context of probability word problems aligns with the conclusions of (Khasanah et al., 2024), who found that systematically implementing PjBL strengthens students' ability to formulate problem-solving strategies and carry them through to completion. Specifically, the PjBL stages of identifying driving questions and designing projects appear to train students to connect abstract formulas with real-world reasoning, so that word problems are gradually perceived as something easier to tackle rather than a source of anxiety.

CONCLUSION AND SUGGESTIONS

This study examines how students' mathematical critical thinking develops over three Project-Based Learning sessions on Probability, using the FRISCO framework as an analytical lens. Students' scores showed a gradual, rather than instantaneous, trajectory of improvement: the increase from the first to the second session was not statistically significant, while the increase in the third session was significant with a large effect size, and the Friedman test confirmed a significant overall difference among the three sessions with a moderate-to-large effect size. At the indicator level, Focus and Situation improved sharply in the third session; Reasoning remained consistently the strongest indicator, while Inference and Big Picture remained the weakest, indicating that formulating contextual conclusions and verifying final answers remained the most persistent challenges for students throughout this PjBL sequence.

These findings should be interpreted with an awareness of several study limitations. The single-group design with repeated measurements, without a control group or a pre-intervention baseline, means the observed trajectory of improvement cannot be interpreted as causal evidence of PjBL's effectiveness. The mechanisms proposed to explain this pattern, such as students' adaptation to independent inquiry or engagement in authentic contexts, are theoretically plausible explanations but are not yet supported by observational data, project artefacts, or direct student reflections. This study also did not document the equivalence of question formats across sessions or inter-rater reliability in rubric scoring. Nevertheless, these results still offer practical implications: mathematics teachers may consider PjBL for topics with strong real-world relevance, provided they provide explicit guidance during the inference and answer verification stages. Future research should use a design with a comparison group, document instrument equivalence across sessions, and report scoring reliability to strengthen the evidence base for claims about PjBL's effectiveness.

REFERENCES

Aini, M., Ridianingsih, D. S., & Yunitasari, I. (2022). Efektivitas model pembelajaran project based learning (PJBL) berbasis stem terhadap keterampilan berpikir kritis siswa. [The effectiveness of stem-based project based learning (PJBL) learning model on critical thinking]. *Jurnal Kiprah Pendidikan, 1(4)*, 247–253. https://doi.org/10.33578/kpd.v1i4.118

Aizah, D. L., & Dewi, H. L. (2024). Efektivitas project based learning dalam kurikulum merdeka terhadap kemampuan berpikir kritis matematis siswa smp negeri 1 karangdadap. [The effectiveness of project-based learning in the merdeka curriculum on the mathematical critical thinking skills of students at karangdadap public junior high school 1]. *Circle: Jurnal Pendidikan Matematika*, *4* (2), 175–185. https://doi.org/10.28918/circle.v4i2.8714

Amir, A., & Solida, A. (2022). Penerapan pembelajaran berbasis proyek dan hasil belajar mahasiswa [The implementation of project-based learning and student learning outcomes]. *Jurnal Ilmiah Universitas Batanghari Jambi*, *22*(3), 2112. https://doi.org/10.33087/jiubj.v22i3.3018

Anantasia, G., & Rindrayani, S. R. (2025). Metodologi penelitian quasi eksperimen [Quasi-experimental research methodology]. *Adiba: Journal of Education*, *5*(2), 183–192. https://jutepe-joln.net/index.php/JURPERU/article/view/811/800

Anisa, W. N. & Prasetyo, A. K. (2023). Peningkatan kemampuan menyelesaikan soal cerita melalui model problem based learning pada pembelajaran matematika di kelas v sd negeri blotongan 02 [Improving the ability to solve workshops through the problem-based learning model in mathematics instruction in grade 5 at blotongan 02 public elementary school]. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, *9*(4), 2550–2562. https://doi.org/10.36989/didaktik.v9i04.1832

Ansya, Y. A., & Salsabilla, T. (2025). Keterampilan guru abad 21 dalam pembelajaran ipa di sekolah dasar [21st-century teacher skills in science instruction in elementary school]. *Jurnal Dunia Pendidikan,* 6(2), 498–512. https://doi.org/10.55081/jurdip.v6i2.4579

Azizah, R. (2022). Project based learning dalam pembelajaran matematika [Project-based learning in mathematics education]. *J-PiMat Jurnal Pendidikan Matematika*, *4*(2), 539–550. https://doi.org/10.31932/j-pimat.v4i2.2026

Difinubun, F. A., Makmuri, M., & Flavia Aurelia, H. (2022). Analisis kebutuhan modul ajar matematika untuk meningkatkan kemampuan pemecahan masalah siswa smk kelas x [Analysis of the needs for mathematics instructional modules to improve problem-solving skills among 10th-grade vocational high school students]. *Griya Journal of Mathematics Education and Application*, *2*(4), 853–864. https://doi.org/10.29303/griya.v2i4.244

Fisher, D., Kusumah, Y. S., & Dahlan, J. A. (2020). Project-based learning in mathematics: A literatur review. *Journal of Physics: Conference Series*, *1657*(1). https://doi.org/10.1088/1742-6596/1657/1/012032

Inayah, S., Safitri, Y., Zurweni, Sarmada, N., Andriani, N., Maison, Anindita, M., Mardhotillah, B., Multahadah, C., Gusmanely, Z., Muhammad, D., & Karmanto, B. (2025). *Statistika non parametrik*. U ME Publishing. https://umepublishing.com/statistika-non-parametrik/

Khasanah, U. N., Rejeki, S., & Sugiyanti, S. (2024). Efektivitas pembelajaran project-based learning berbasis steam terhadap kemampuan berpikir kritis pada siswa smp [The effectiveness of steam-based project-based learning on critical thinking skills among middle school students]. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *8*(3), 2225–2242. https://doi.org/10.31004/cendekia.v8i3.3529

Kristiyanto, D. (2020). Peningkatan kemampuan berpikir kritis dan hasil belajar matematika dengan model project based learning (PJBL) [Improving critical thinking skills and mathematics learning outcomes using the project-based learning (PJBL) model]. *Mimbar Ilmu*, *25*(1), 1-10. https://doi.org/10.23887/mi.v25i1.24468

Lapase, M. H. (2021). Implementasi pembelajaran berbasis proyek untuk meningkatkan hasil belajar siswa pada mata pelajaran matematika di sd negeri pinedapa [Implementation of project-based learning to improve student learning outcomes in mathematics at pinedapa public elementary school]. *Jurnal Paedagogy*, *8*(2), 134. https://doi.org/https://doi.org/10.33394/jp.v8i2.3492

Lestari, R., & Lubis, A. (2023). Pengaruh model project based learning dengan media visual terhadap keterampilan berpikir kritis [The effect of project-based learning with visual media on critical thinking skills]. *JMPM: Jurnal Matematika dan Pendidikan Matematika*, *8*(1), 15–27.

https://doi.org/https://doi.org/10.26594/jmpm.v8i1.3631
Maesaroh, S. (2021). Kemampuan berpikir kritis matematis siswa smp pada materi bangun ruang sisi datar [Middle school students' mathematical critical thinking skills in the topology of flat-sided three-dimensional shapes]. *Jurnal Riset Intervensi Pendidikan (JRIP)*, *3*(2), 99–105. https://journal.rekarta.co.id/index.php/jrip/article/view/164
Muhajirin, Risnita, & Asrulla. (2024). Pendekatan penelitian kuantitatif dan kualitatif serta tahapan penelitian [Quantitative and qualitative research approaches and research phases]. *Journal Genta Mulia*, *15*(1), 82–92. https://ejournal.uncm.ac.id/index.php/gm/article/view/903
Rafik M., Febrianti V. P., Nurhasanah A., & Muhajir S. N. (2022). Telaah literatur: pengaruh model pembelajaran project based learning (pjbl) terhadap kreativitas siswa guna mendukung pembelajaran abad 21 [Literature review: the effect of the project-based learning (PJBL) model on student creativity to support 21st-century learning]. *Jurnal Pembelajaran Inovatif*, *5*(1), 80–85. https://doi.org/https://doi.org/10.21009/JPI.051.10
OECD. (2023). *PISA 2022 results (volume I and ii) – country notes: Indonesia*. OECD Publishing. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html
Benyamin, B., Qohar, A., & Sulandra, I. M. (2021). Analisis kemampuan berpikir kritis siswa sma kelas x dalam memecahkan masalah spltv [An analysis of the critical thinking skills of 10th-grade high school students in solving spltv problems]. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 5(2),* 909–922. https://doi.org/https://doi.org/10.31004/cendekia.v5i2.574
Rosa, E., Destian, R., Agustian, A., & Wahyudin, W. (2024). Inovasi model dan strategi pembelajaran dalam implementasi kurikulum merdeka [Innovations in learning models and strategies in the implementation of the merdeka curriculum]. *Journal of Education Research*, *5*(3), 2608–2617. https://doi.org/10.37985/jer.v5i3.1153
Setiana, D. S., Nuryadi, N., & Santosa, R. H. (2020). Analisis kemampuan berpikir kritis matematis ditinjau dari aspek overview [Analysis of mathematical critical thinking skills from an overview perspective]. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 1-12. https://doi.org/10.30998/jkpm.v6i1.6483
Sherly, D. F., Hadi, F. Rahmatika, & Pradana, L. N. (2020). Efektivitas model pembelajaran project based learning (PJBL) terhadap keterampilan menyelesaikan soal cerita matematika [The effectiveness of the project-based learning (PJBL) model on students' skills in solving mathematics workshops]. *Prosiding Konferensi Ilmiah Dasar*, *4(1)*. http://prosiding.unipma.ac.id/index.php/KID
Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Bandung). Alfabeta.
Triani Putri, M., Sagala, V., & Listiana, Y. (2022). Profil berpikir kritis siswa dalam memecahkan soal higher order thinking berdasarkan gaya berpikir sternberg menurut fungsinya [Profile of students' critical thinking in solving higher-order thinking problems based on sternberg's thinking styles according to their functions]. *GAUSS: Jurnal Pendidikan Matematika*, *5*(1), 50–59. https://doi.org/10.30656/gauss.v5i1.4741
Tunu, D. J. I., Daniel, F., & Gella, N. J. M. (2022). Analisis kemampuan menyelesaikan soal cerita matematika siswa ditinjau dari gender [An analysis of students' ability to solve math word problems by gender]. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *6*(2), 1499–1510. https://doi.org/10.31004/cendekia.v6i2.1366
Widiana, D. R., Syafiuddin, Sriwijayasih, I., Aju, I. R., Praharsi, Y., & Novianarenty, E. (2025). The use of the Wilcoxon signed-rank test in analyzing the difference in test scores before and after digital marketing training. *Jurnal Teknologi Maritim, 8*(2), 23-32. https://doi.org/10.35991/jtm.v8i2.73

Zahra, F. A., & Hakim, D. L. (2022). Kesulitan siswa kelas vii smp dalam berpikir kritis matematis pada materi garis dan sudut [Difficulties faced by seventh-grade middle school students in mathematical critical thinking regarding lines and angles]. *PHI: Jurnal Pendidikan Matematika*, *6*(2), 208. https://doi.org/10.33087/phi.v6i2.244