



The Influence of Integrated STEM Learning on Students' Critical Reasoning Skills

Muhammad Ridwan Nur Cholis¹, Bambang Subali¹, Masturi¹

¹Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang

*Corresponding Author: Muhammad Ridwan Nur Cholis

Email: cholisn214@gmail.com, bambangfisika@mail.unnes.ac.id, masturi@mail.unnes.ac.id

Article Info

Article History:

Received August 21, 2026

Revised October 7, 2026

Accepted: October 20, 2026

Keywords:

Teaching at The Right Level, STEM, Critical Thinking.

Abstract

This study aims to determine the effect of STEM-integrated Teaching at the Right Level (TaRL) on students' critical reasoning skills. The background of this study is rooted in the incomplete implementation of differentiated instruction in schools, particularly regarding process differentiation, which has failed to adequately accommodate students' varying levels of cognitive understanding. This study employed a quantitative approach with a quasi-experimental, pretest-posttest control group design. The sample consisted of two Grade X classes at SMA Negeri 1 Kendal and two Grade X classes at SMA Negeri 2 Kendal, selected through purposive sampling. The research instrument was a critical reasoning skills test developed based on the indicators proposed by Trilling and Fadel (2009). Data were analyzed using N-Gain, independent sample t-tests, and ANCOVA. The results demonstrated that the N-Gain scores of the experimental classes (0.47 and 0.33) were higher than those of the control classes (0.27 and 0.31). Independent sample t-tests yielded significance values of 0.000 and 0.006 ($p < 0.05$), indicating a significant effect of STEM-integrated TaRL on the improvement of critical reasoning skills. Furthermore, the ANCOVA results showed significance values of 0.001 and 0.000 ($p < 0.05$), confirming a significant influence of the TaRL approach on the differences in mean scores between student groups. These findings indicate that STEM-integrated TaRL is more effective than conventional instruction in developing critical reasoning skills, such as identifying, reasoning, analyzing, providing effective justifications, applying systems thinking, and drawing conclusions. This instructional approach is not only pedagogically effective but also relevant in preparing students to meet the demands of 21st-century skill development.

Introduction

Education is one of the efforts to improve human quality through the dissemination of knowledge and the development of competencies (González-Pérez et al., 2022; Gaybullaevna & Jonpulatovna, 2021; Yan et al., 2024; Zeichner, 2017). The Merdeka Curriculum was prepared as an effort to develop 21st-century skills through character strengthening through the Pancasila Student Profile (Kurniati, et al., 2022).

The results of PISA 2022 showed that Indonesia's achievements in mathematics, reading, and science declined compared to 2018 (Wijaya et al., 2024; Solihin et al., 2024; Cahyani & Setiawan, 2024; Ivanka & Setiawan, 2025). In addition, the proportion of students achieving advanced levels (Levels 5 and 6) was less than 1% to 5%, indicating that approximately 95% of students had not yet been able to demonstrate higher-order reasoning skills, argument

evaluation, and complex problem-solving. Therefore, this condition indicates the need to strengthen students' competencies, one of which is critical reasoning ability.

Students who reason critically when facing problems will be able to process the information they obtain, both quantitatively and qualitatively, relate various pieces of information, reflect, evaluate, and draw conclusions rationally (Pandu, et al., 2023). Critical reasoning enables students to think based on certain reasons and focus on making decisions about what to do. The following are indicators of students' critical reasoning. Identifying is an activity carried out by seeking information to find an appropriate problem-solving method. Casmini, et al., (2024) stated that the main component of critical thinking is that students are able to seek the truth through identification, analyze problems, and make appropriate decisions.

The ability to think in order to solve problems needs to be accompanied by good reasoning skills (Lianawati, et al., 2025; Sutiani et al., 2021; Wechsler et al., 2018; Haryani et al., 2018). This is because, in developing solutions to problems, students need to involve the process of filtering and combining information and ideas. When students are faced with a problem on a worksheet, they will learn to evaluate information in depth, make arguments based on logical reasons, and express their opinions (Jannah, et al., 2022). In-depth exploration will encourage students' cognitive processes that support them in thinking critically. Systems thinking involves various broad techniques and ways of thinking to create a useful foundation for further learning (Salim, et al., 2023). The ability to conclude is the ability of students to interpret observed events or objects based on deductive or inductive concepts and adjusted to known facts, principles, and concepts. The ability to conclude is an important basic concept in learning to build new knowledge and skills (Nurdianti et al., 2020). One learning approach that is in line with the Merdeka Curriculum is TaRL (Teaching at The Right Level). TaRL is oriented toward learning that is adjusted to students' ability levels rather than their age or grade level (Ahyar, et al., 2022). This is in line with the statement by Cahyono (2022), that TaRL is a learning approach that refers to students' abilities.

The TaRL approach begins by providing an assessment to students. The results of this assessment become the basis for teachers to divide students into smaller learning groups (Jazuli, 2022). In learning that applies TaRL, teachers must know students' initial abilities, interests, and learning styles as a basis for designing and implementing learning. Teachers must be able to serve students based on their characteristics and know the appropriate strategies, learning media, and selection of learning materials according to their needs (Muin, 2023; Bordoh, 2025; Widodo, 2018; Mahartika et al., 2023).

STEM education is currently a major focus of global learning in the 21st century, emphasizing the integration of various fields of study including science, technology, engineering, and mathematics. STEM is a form of learning approach that combines disciplines including science, technology, engineering, and mathematics that aims to increase students' interest, problem-solving abilities, critical thinking, and problem analysis in the STEM field (Fachrunnisa, et al., 2021). In learning, STEM content is integrated and studied as a single subject concept to solve real-world problems.

Based on the discussion, the integration of the TaRL learning approach with STEM provides a new learning concept and, in line with the Merdeka Curriculum, has greater potential for developing critical reasoning skills. TaRL enables teachers to conduct diagnostic assessments, identify initial abilities, group students flexibly, and provide levels of scaffolding and task complexity appropriate to the needs of each group (Wulandari et al., 2025). Meanwhile, the STEM approach provides learning experiences through contextual problems, investigation,

data analysis, design, testing, and evaluation of solutions. Thus, the integration of the two enables learning not only to be oriented toward matching students' ability levels but also to provide experiences systematically, thereby encouraging critical reasoning abilities more effectively.

Method

Research Design

This study employed a quasi-experimental method using a nonequivalent control group design to investigate the influence of Teaching at the Right Level (TaRL) integrated with Science, Technology, Engineering, and Mathematics (STEM) learning on students' critical reasoning skills. This design involved two groups: an experimental group that received STEM-integrated TaRL instruction and a control group that received conventional instruction commonly implemented in schools. Both groups completed a pretest before the instructional intervention and a posttest after the intervention. The pretest was administered to assess students' initial critical reasoning skills and establish a baseline for evaluating subsequent changes, while the posttest was used to assess students' critical reasoning skills following the instructional intervention. Because the study used a quasi-experimental design, the participants were assigned to existing classes rather than randomly assigned individually to the experimental and control conditions. Consequently, the pretest results were considered in interpreting differences between the groups and evaluating the potential influence of the intervention (Sugiyono, 2023).

Research Setting and Participants

The research was conducted at SMA Negeri 1 Kendal, located at Jl. Soekarno Hatta Barat Km. 03, Purwokerto, Patebon, Kendal Regency, and SMA Negeri 2 Kendal, located in Kendal Regency. The participants were Grade X students during the 2024/2025 academic year. The study involved four classes, comprising two experimental classes and two control classes, with a total sample of 144 students. The classes were selected using purposive sampling based on predefined inclusion criteria. These criteria included participation in a uniform curriculum, enrollment at the same grade level, and comparable class characteristics without formal differentiation into superior or selective classes. The selection process was intended to identify classes with reasonably comparable educational characteristics and to support a meaningful comparison between STEM-integrated TaRL instruction and conventional instruction. However, because purposive sampling was used, the findings should be interpreted in relation to the participating classes and schools rather than assumed to be representative of all secondary school students.

Instructional Procedures

The research procedures consisted of a pre-intervention assessment, an instructional intervention, and a post-intervention assessment. Before the intervention, students in both the experimental and control groups completed a pretest measuring their critical reasoning skills. In the experimental group, the pretest results were also used to classify students into low, medium, and high initial-ability categories in accordance with the TaRL principle of adapting instruction to students' learning levels. The experimental group subsequently received STEM-integrated TaRL instruction, in which STEM-related learning activities were organized in accordance with students' identified initial ability levels. The control group received conventional instruction as ordinarily implemented in the participating schools. After the instructional period, both groups completed the posttest to measure their critical reasoning skills following the respective instructional approaches. This procedure provided the basis for

examining within-group changes over time and differences in post-intervention outcomes between the experimental and control groups. To support replication, the final manuscript should specify the duration and number of instructional sessions, the subject matter taught, the detailed learning activities in each condition, the procedures used to implement STEM integration, and the criteria and procedures for assigning students to the three TaRL ability categories.

Research Instrument

The primary research instrument was a descriptive test designed to measure students' critical reasoning skills before and after the instructional intervention. The same assessment framework was used for the pretest and posttest to enable comparison of students' performance over time. The test items were developed with reference to five critical reasoning indicators: identifying, reasoning and analyzing, providing effective reasons, using systems thinking, and concluding. These indicators served as the basis for assessing students' ability to identify relevant information, analyze and interpret information, provide reasoned justifications, understand relationships within a system, and formulate conclusions. The pretest scores represented students' initial performance, whereas the posttest scores represented their performance after the instructional intervention. As research instruments must provide valid and reliable measurements, the manuscript should also report the number of test items, the scoring rubric, the maximum possible score, the procedures used to establish content validity, and the reliability coefficient or other relevant evidence of measurement consistency. If expert judgment, a pilot study, or a statistical validation procedure was conducted, the specific procedure and its results should be reported (Sundayana, 2018). These details are necessary to establish the quality and reproducibility of the instrument.

Data Collection Procedures

Data were collected through the administration of pretests and posttests to students in the experimental and control groups at both participating schools. The pretest was administered before the learning intervention to establish students' initial critical reasoning performance and, in the experimental group, to support classification into low, medium, and high initial-ability categories. Following the pretest, the experimental classes participated in STEM-integrated TaRL learning, while the control classes received conventional instruction. The posttest was administered after the instructional period using the established critical reasoning assessment framework. Students' responses were scored according to the assessment criteria, and the resulting scores were compiled for statistical analysis. The data were then examined to compare pretest and posttest performance within groups, posttest outcomes between instructional conditions, and differences among the initial-ability categories. The final report should specify the test administration conditions, intervention schedule, scoring procedures, and measures taken to ensure consistent implementation across the two schools.

Data Analysis

Data analysis was conducted using descriptive and inferential statistical procedures to examine changes in students' critical reasoning skills and differences associated with the instructional approaches and initial-ability categories. Before conducting inferential analyses, the dataset should be checked for completeness, scoring accuracy, outliers, and the assumptions relevant to the selected statistical tests. Where applicable, normality and homogeneity of variance should be examined for the independent-samples and paired-samples *t*-tests, while the

assumptions of analysis of covariance (ANCOVA), including the homogeneity of regression slopes, should be evaluated before interpreting the ANCOVA results.

Descriptive Statistics

Descriptive statistics were used to summarize students' pretest and posttest scores in the experimental and control groups at SMA Negeri 1 Kendal and SMA Negeri 2 Kendal. The analysis should include the number of students, mean scores, and standard deviations for each group and assessment occasion. These statistics provide an overview of students' initial performance, post-intervention performance, and the observed patterns of change. Descriptive results also help contextualize the inferential analyses by showing the direction and magnitude of the score differences before statistical significance is assessed.

Independent-Samples t-Test

The independent-samples *t*-test was used to examine differences in critical reasoning scores between the experimental and control groups. In this study, the test was intended to determine whether students who received STEM-integrated TaRL instruction differed in their performance from students who received conventional instruction. The analysis should clearly identify the score variable being compared, such as posttest scores, and report the group means, *t*-statistic, degrees of freedom, *p*-value, and, where possible, an effect-size estimate with a confidence interval. If the purpose is to estimate differences in post-intervention performance while accounting for baseline differences, the interpretation of the independent-samples *t*-test should be supplemented by an appropriate baseline-adjusted analysis. Statistical significance alone does not establish the practical magnitude of an intervention effect.

Paired-Samples t-Test

The paired-samples *t*-test was used to determine whether students' critical reasoning scores changed significantly between the pretest and posttest. The analysis compared the two sets of scores obtained from the same students within each relevant group. Separate analyses were conducted for the participating schools or groups, as appropriate, to examine whether students' performance changed over the instructional period. The results should report the mean pretest and posttest scores, the mean difference, the *t*-statistic, degrees of freedom, and the *p*-value. An effect-size estimate for paired observations should also be provided where possible. A statistically significant pretest–posttest difference indicates that students' scores changed over time; however, by itself, it does not establish that the change was caused exclusively by the instructional intervention. The comparison with the control group is therefore important when interpreting the potential contribution of STEM-integrated TaRL learning.

Analysis of Covariance (ANCOVA)

Analysis of covariance (ANCOVA) was used to examine differences in critical reasoning outcomes among students classified into low, medium, and high initial-ability categories, while statistically adjusting for pretest scores. In this analysis, the posttest score served as the dependent variable, the initial-ability category served as the categorical independent variable, and the pretest score served as the covariate. The purpose was to determine whether adjusted posttest performance differed significantly among the initial-ability categories after accounting for students' baseline scores. Where the overall ANCOVA indicated a statistically significant difference among categories, a suitable post hoc procedure was used to identify which pairs of categories differed significantly. The results should include the relevant *F*-statistics, degrees of freedom, *p*-values, adjusted means where available, and an effect-size measure. The

assumptions of ANCOVA, particularly the homogeneity of regression slopes and the appropriateness of the covariate, should be checked and reported. Importantly, this model evaluates differences among initial-ability categories; it does not independently establish the effect of the instructional treatment unless the instructional condition is also included as a factor in the model. If the primary research objective is to estimate the effect of STEM-integrated TaRL relative to conventional instruction while controlling for pretest scores, instructional condition should be included explicitly in the ANCOVA model.

Post Hoc Analysis

Post hoc comparisons were conducted following a statistically significant overall ANCOVA result to determine which initial-ability categories differed from one another in their adjusted critical reasoning outcomes. The comparisons involved the low- versus medium-ability groups, low- versus high-ability groups, and medium- versus high-ability groups. The significance values were used to determine whether each pairwise difference was statistically significant, with the significance threshold set at $p < .05$. The manuscript should identify the post hoc procedure used, such as Bonferroni adjustment or another appropriate multiple-comparison method, and clarify whether the reported mean differences were based on raw or adjusted scores. Reporting the method is important because multiple pairwise comparisons increase the risk of false-positive findings if the significance level is not appropriately controlled.

Normalized Gain Analysis

Normalized gain (*N*-gain) was calculated to describe the extent of improvement in students' critical reasoning scores from pretest to posttest relative to the maximum possible improvement. The *N*-gain values were used to characterize the degree of observed improvement in the experimental and control groups. Group-level values should be calculated and reported using a consistent procedure, with an explanation of whether the reported value represents the mean of individual students' normalized gains or a normalized gain calculated from group mean scores. The method used to interpret the resulting values should also be stated explicitly. Because normalized gain describes the relative change in scores rather than directly testing a treatment effect, it should be interpreted alongside the between-group comparisons and other inferential results.

Statistical Significance and Interpretation

A significance level of .05 was used as the reference threshold for the inferential analyses. Results with $p < .05$ were interpreted as statistically significant, whereas results with $p \geq .05$ were interpreted as not statistically significant. For each analysis, statistical significance was considered together with the direction and magnitude of the observed differences. Where feasible, effect sizes and confidence intervals should be reported to clarify the practical importance and precision of the findings. The conclusions concerning the influence of STEM-integrated TaRL learning should be based primarily on analyses that directly compare the experimental and control conditions, taking account of the pretest scores and the quasi-experimental design. Changes within the experimental group and differences among initial-ability categories provide complementary information but should not, on their own, be treated as conclusive evidence of a causal treatment effect.

Result and Discussion

The analysis included an *independent sample t-test* and a *paired sample t-test*, as well as normality and homogeneity tests to ensure that the statistical assumptions were met.

Subsequently, the ANCOVA test was used to evaluate the effectiveness of the learning process by controlling for pretest scores. The N-Gain test was then conducted to assess the improvement in learning outcomes. In this study, a difference test was conducted to determine whether there was a statistically significant difference between the experimental class and the control class. The following are the mean scores of critical reasoning abilities in the experimental and control classes at two schools, namely SMA N 1 Kendal and SMA N 2 Kendal, presented in Figures 1 and 2.

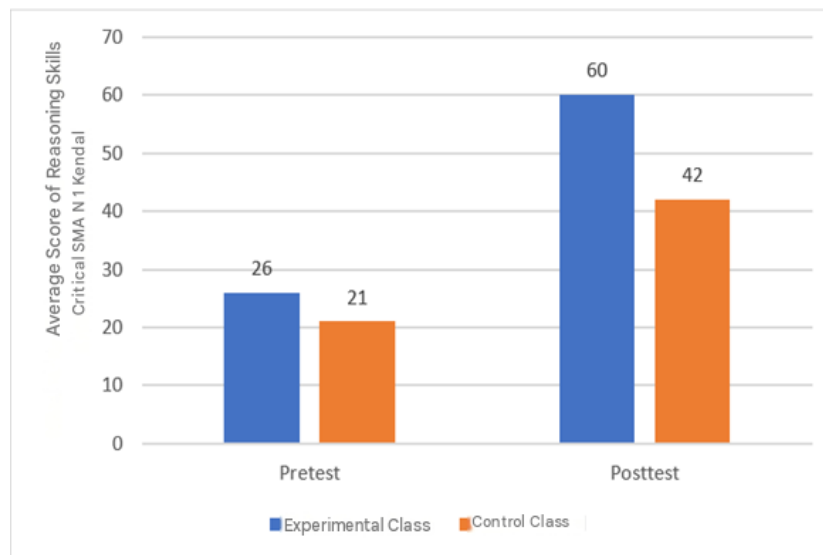


Figure 1. Average Pretest and Posttest Scores of Critical Reasoning Skills at SMA Negeri 1 Kendal

Figure 1. The improvement in students' critical reasoning skills in the experimental and control classes at SMA Negeri 1 Kendal. The experimental class showed an increase in the average score from 26 in the pretest to 60 in the posttest, representing an improvement of 34 points. Meanwhile, the control class increased from 21 to 42, representing an improvement of 21 points. These results indicate that the experimental class achieved greater progress than the control class. The higher improvement in the experimental class suggests that the integration of Teaching at the Right Level (TaRL) and STEM learning may be more effective in developing students' critical reasoning skills than conventional instruction. This improvement may be associated with learning activities that accommodate students' initial ability levels and engage them in contextual problem-solving, investigation, and analytical reasoning.



Figure 2. Average Pretest and Posttest Scores of Critical Reasoning Skills at SMA Negeri 2 Kendal

Figure 2. The improvement in students' critical writing skills at SMA Negeri 2 Kendal in both the experimental and control classes. The experimental class showed an increase in the average score from 36 in the pretest to 57 in the posttest, representing an improvement of 21 points. Meanwhile, the control class increased from 22 to 46, representing an improvement of 24 points. Although both classes demonstrated improved performance, the control class showed a slightly greater increase in absolute score points. However, the experimental class achieved a higher posttest score of 57 compared with 46 in the control class. These findings suggest that STEM-integrated Teaching at the Right Level (TaRL) may support the development of students' critical writing skills, although further statistical analysis is needed to determine whether the differences in improvement between the two classes are statistically significant.

Table 1. Results of the Test of Differences in Average Critical Thinking Skills

School	Group	n	Mean	df	t-value	Sig. (2-tailed)
SMA Negeri 1 Kendal	Experimental	36	59.97	70	4.087	0.000
	Control	36	44.53	70		
SMA Negeri 2 Kendal	Experimental	36	57.03	70	2.837	0.006
	Control	36	46.42	66.54		

Table 1. presents the results of the independent-samples t-test comparing students' critical thinking skills in the experimental and control classes at SMA Negeri 1 Kendal and SMA Negeri 2 Kendal. At SMA Negeri 1 Kendal, the experimental class achieved a higher mean score (59.97) than the control class (44.53), with a t-value of 4.087 and a significance value of 0.000 ($p < 0.05$). Similarly, at SMA Negeri 2 Kendal, the experimental class obtained a higher mean score (57.03) than the control class (46.42), with a t-value of 2.837 and a significance value of 0.006 ($p < 0.05$). These results indicate statistically significant differences in critical thinking skills between the experimental and control classes at both schools, with the experimental classes achieving higher mean scores. The findings support the potential effectiveness of STEM-integrated Teaching at the Right Level (TaRL) in improving students' critical thinking skills compared with conventional instruction.

Table 2. Results of the Difference Test for Average Improvement in Critical Reasoning Skills

School	n	df	t-value	Sig. (2-tailed)
SMA Negeri 1 Kendal	36	35	-11.185	0.000
SMA Negeri 2 Kendal	36	35	-12.070	0.000

Table 2. presents the results of the paired-samples t-test examining the improvement in students' critical reasoning skills at SMA Negeri 1 Kendal and SMA Negeri 2 Kendal. At SMA Negeri 1 Kendal, the analysis yielded a t-value of -11.185 with a significance value of 0.000 ($p < 0.05$), while at SMA Negeri 2 Kendal, the t-value was -12.070 with a significance value of 0.000 ($p < 0.05$). These results indicate statistically significant differences between students' pretest and posttest scores at both schools, suggesting that students' critical reasoning skills improved following the learning intervention. The larger absolute t-value at SMA Negeri 2 Kendal indicates stronger statistical evidence of a pretest–posttest difference, although the magnitude of improvement should be assessed using the actual score differences or effect sizes. Overall, the findings suggest that STEM-integrated Teaching at the Right Level (TaRL) may contribute to improving students' critical reasoning skills.

Table 3. Summary of Ancova Test Results of SMA N 1 Kendal

Source of Variation	F	Sig.	Decision
Pretest (Covariate)	3.546	0.069	Not Significant
Initial Ability Group	8.346	0.001	Significant
Overall Model	14.369	0.000	Significant
Adjusted R ²	0.534	—	The model explains 53.4% of the variance in posttest scores.

Table 3. presents the results of the ANCOVA analysis for SMA Negeri 1 Kendal to examine the effects of students' initial ability levels on posttest scores while considering pretest performance as a covariate. The results show that the pretest covariate was not statistically significant ($F = 3.546$, $p = 0.069$), indicating that pretest scores did not have a significant effect on posttest scores in this model. In contrast, the initial ability group showed a statistically significant effect ($F = 8.346$, $p = 0.001$), suggesting that posttest performance differed significantly across students' initial ability levels. Furthermore, the overall model was statistically significant ($F = 14.369$, $p < 0.001$), with an adjusted R² of 0.534 , indicating that the model explained 53.4% of the variance in posttest scores. These findings suggest that students' initial ability levels played an important role in explaining differences in critical reasoning performance at SMA Negeri 1 Kendal. However, the significant overall model alone does not establish the independent causal effect of STEM-integrated TaRL without a clearly specified treatment-group effect in the ANCOVA results.

Table 4. Summary of Ancova Test Results of SMA N 2 Kendal

Source of Variation	F	Sig.	Decision
Pretest (Covariate)	0.122	0.730	Not Significant
Initial Ability Group	18.003	0.000	Significant
Overall Model	88.558	0.000	Significant
Adjusted R ²	0.882	–	The model explains 88.2% of the variance in posttest scores.

Table 4. presents the results of the ANCOVA analysis for SMA Negeri 2 Kendal, examining the effects of students' initial ability levels on posttest scores while controlling for pretest performance. The results indicate that the pretest covariate was not statistically significant ($F = 0.122$, $p = 0.730$), suggesting that pretest scores did not significantly contribute to explaining posttest performance in this model. In contrast, the initial ability group showed a statistically significant effect ($F = 18.003$, $p < 0.001$), indicating significant differences in posttest scores among students with different initial ability levels. Furthermore, the overall model was statistically significant ($F = 88.558$, $p < 0.001$), with an adjusted R^2 of 0.882, meaning that the model explained 88.2% of the variance in posttest scores. These findings highlight the importance of students' initial ability levels in explaining differences in critical reasoning performance at SMA Negeri 2 Kendal. However, the independent effect of STEM-integrated Teaching at the Right Level (TaRL) cannot be established from these results alone unless the treatment-group effect is also reported in the ANCOVA model.

Table 5. Results of Advanced Test of SMA N 1 Kendal

Group Comparison	Mean Difference	Sig.	Interpretation
Low vs. Medium	-18.610	0.001	Significantly different
Low vs. High	-12.450	1.000	Not significantly different
Medium vs. High	6.160	1.000	Not significantly different

Table 5. presents the results of the post hoc comparisons of students' critical reasoning skills across different initial ability groups at SMA Negeri 1 Kendal. The results show a statistically significant difference between the low- and medium-ability groups, with a mean difference of -18.610 and a significance value of 0.001 ($p < 0.05$), indicating that the two groups differed significantly in their posttest performance. In contrast, the comparison between the low- and high-ability groups showed no statistically significant difference (mean difference = -12.450 , $p = 1.000$). Similarly, no statistically significant difference was found between the medium- and high-ability groups (mean difference = 6.160 , $p = 1.000$). These findings suggest that, at SMA Negeri 1 Kendal, the most pronounced statistically significant difference in critical reasoning performance occurred between students in the low- and medium-ability groups, whereas the differences involving the high-ability group were not statistically significant.

Table 6. Results of Advanced Test of SMA N 2 Kendal

Group Comparison	Mean Difference	Sig.	Interpretation
Low vs. Medium	-23.609	0.000	Significantly different
Low vs. High	-46.533	0.000	Significantly different
Medium vs. High	-22.924	0.000	Significantly different

Table 6. presents the results of the post hoc analysis comparing students' critical reasoning skills across the low, medium, and high initial ability groups at SMA Negeri 2 Kendal. The results indicate statistically significant differences among all three pairwise comparisons, with a significance value of 0.000 ($p < 0.001$) for each comparison. Specifically, the mean difference between the low and medium groups is -23.609 , between the low and high groups is -46.533 , and between the medium and high groups is -22.924 . These findings suggest that students' critical reasoning scores differ significantly according to their initial ability levels, with the largest mean difference observed between the low and high groups. The negative mean differences indicate the direction of the comparisons as reported in the table, suggesting that the scores of the first group in each comparison are lower than those of the second group. Overall, the findings highlight the importance of considering students' initial ability levels when implementing integrated STEM learning using the Teaching at the Right Level (TaRL) approach.

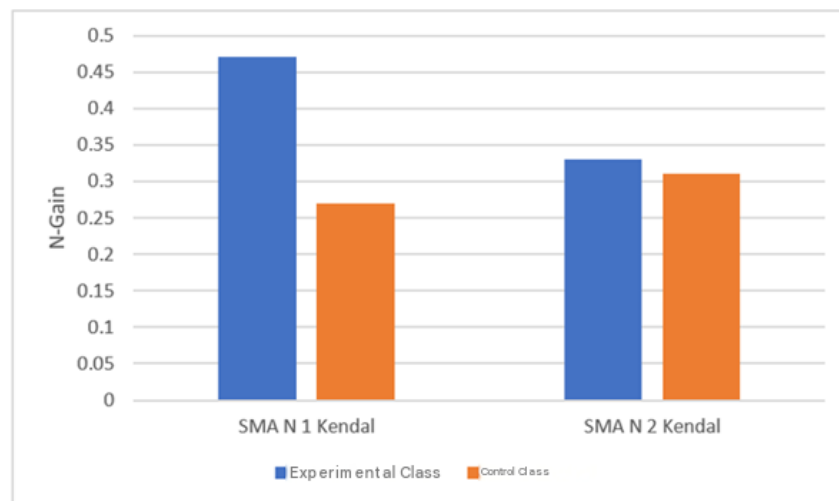


Figure 3. Comparison of N-Gain Scores for Students' Critical Reasoning Skills between the Experimental and Control Classes at SMA N 1 Kendal

Apart from carrying out an overall N-Gain test analysis, this research also carried out an N-Gain test analysis on each critical reasoning indicator. This aims to determine the results of increasing each indicator of students' critical reasoning skills. The results of the N-Gain test analysis for each indicator are presented in Figure 3. for SMA N 1 Kendal.

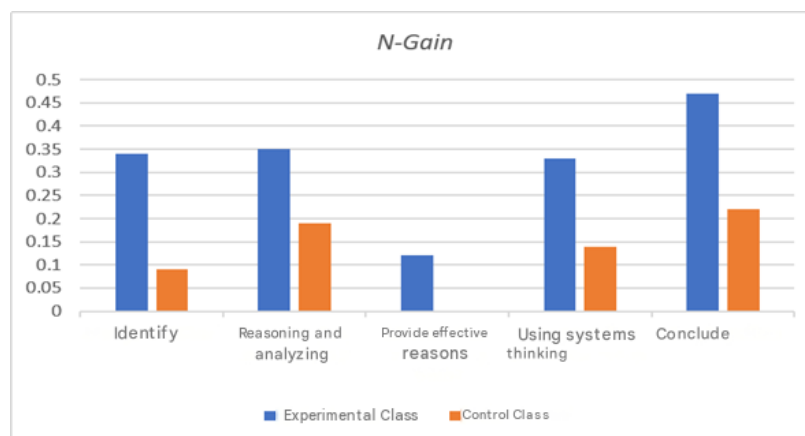


Figure 4. N-Gain of Each SMA N 1 Kendal Indicator

Apart from carrying out an overall N-Gain test analysis, this research also carried out an N-Gain test analysis on each critical reasoning indicator. This aims to determine the results of increasing each indicator of students' critical reasoning skills. The results of the N-Gain test analysis for each indicator are presented in Figure 4. for SMA N 1 Kendal.

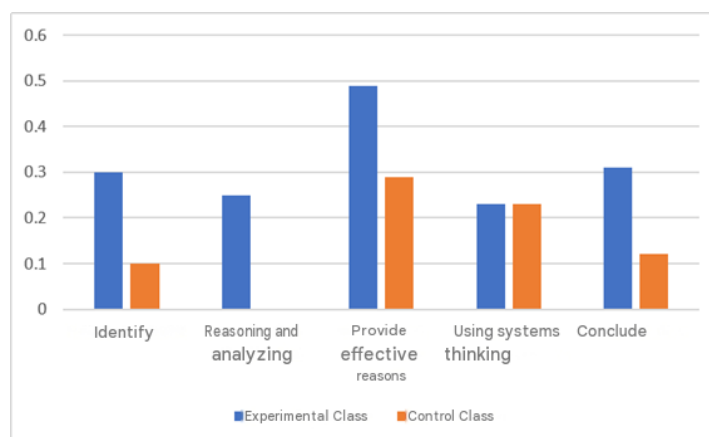


Figure 5. N-Gain of Each SMA N 2 Kendal Indicator

Based on the analysis results presented in Figure 5, the implementation of physics learning using the TaRL approach integrated with STEM was able to produce different levels of critical reasoning abilities in the experimental and control classes. The TaRL approach enables teachers to adjust learning strategies based on students' ability levels so that each group receives a learning experience that suits its needs. This is in line with research conducted by Ningrum (2023), which found that through learning using the TaRL approach, students can be mapped into several groups, making it easier for teachers to provide treatment according to their characteristics and learning needs.

The Integration of STEM and TaRL in Developing Students' Critical Reasoning Skills

The integration of Science, Technology, Engineering, and Mathematics (STEM) in learning provides opportunities for students to connect knowledge and skills across four interrelated disciplines. Through this integration, students are encouraged to engage in problem-solving activities that require them to analyze situations, make decisions, formulate and test hypotheses, evaluate possible solutions, and conduct investigations. These activities create meaningful learning experiences in which students actively construct knowledge rather than simply memorize concepts. As a result, STEM-integrated learning can support the development of students' critical reasoning abilities by encouraging them to apply logical and systematic thinking when addressing problems (Davidi, et al., 2021).

The effectiveness of STEM-integrated learning can be further supported by the Teaching at the Right Level (TaRL) approach, which considers differences in students' initial abilities and learning needs. By adjusting learning activities to students' levels of understanding, teachers can provide appropriate guidance and opportunities for students to develop their reasoning skills gradually. When combined with STEM, TaRL allows students to engage in contextual problem-solving activities that correspond to their learning readiness while encouraging them to apply concepts to real-world situations. This combination may create a more supportive

learning environment in which students can develop conceptual understanding and critical reasoning through active participation.

In this learning process, students are encouraged to examine problems from multiple perspectives and establish connections between theoretical knowledge and practical applications. For example, solving contextual problems may require students to identify relevant information, interpret data, consider alternative solutions, and determine the most appropriate course of action. These processes encourage students to move beyond superficial understanding toward deeper analysis and evidence-based reasoning. Therefore, the integration of STEM and TaRL can provide a structured learning experience that supports students in developing the ability to think critically, systematically, and independently.

In this study, students' critical reasoning abilities were examined through five indicators: identifying, reasoning and analyzing, providing effective reasons, using systematic thinking, and concluding, as proposed by Trilling & Fadel (2009). These indicators represent interconnected dimensions of critical reasoning that can be developed through meaningful learning experiences. The integration of STEM and TaRL provides opportunities for students to practice these skills through contextual investigations and structured problem-solving activities. Thus, this instructional approach offers a relevant framework for supporting the development of critical reasoning, particularly when learning activities are designed to accommodate students' abilities and encourage active engagement with real-world problems.

Developing Analytical Reasoning and Evidence-Based Explanations

The ability to identify relationships among different types of information, concepts, data, and opinions is an essential component of critical reasoning. Students need to recognize relevant information, distinguish between important and less relevant details, and connect existing knowledge with new information to understand a problem comprehensively. Learning activities that incorporate contextual questions can facilitate this process because they require students to interpret information within meaningful situations rather than rely solely on memorized knowledge. This is consistent with the findings of Nikmah (2024), who reported that contextual learning questions enable students to identify patterns, reflect on previously acquired knowledge, and establish connections with new knowledge.

Beyond identifying information, students also need to develop the ability to reason and analyze by examining how and why particular events or phenomena occur. Instead of focusing exclusively on questions about what happened, learning activities should encourage students to investigate the underlying causes, processes, and relationships involved. This approach allows students to consider alternative explanations, evaluate the relevance of available information, and construct logical arguments. Through analytical activities, students can broaden their perspectives and deepen their understanding of complex issues. Mollick (2022) emphasizes the importance of analytical abilities in enabling individuals to examine subjects from different perspectives and enrich their thinking processes.

The ability to provide effective reasons is another important aspect of critical reasoning. Students must be able to support their opinions and proposed solutions with relevant facts, logical explanations, and appropriate evidence. In STEM-integrated learning, this ability can be developed through activities that require students to explain the procedures they use, justify their decisions, and evaluate whether their proposed solutions are supported by the results of their investigations (Bhakti et al., 2020; Stevenson et al., 2026). Such activities also encourage students to communicate their ideas clearly and respond to alternative arguments. Critical

reasoning therefore serves as an important foundation for constructing convincing explanations and developing meaningful academic communication (Sapia, et al., 2022).

The relationship between analytical reasoning and evidence-based explanations is particularly important because students must evaluate information before using it to support a conclusion. They need to determine whether the available evidence is relevant, whether their reasoning is logically consistent, and whether their explanations adequately address the problem. Through structured discussions, investigations, and problem-solving tasks, students can practice these processes and become more aware of the importance of evidence in developing arguments. Consequently, learning activities that emphasize analysis, justification, and evaluation can provide opportunities for students to strengthen their critical reasoning abilities. The extent of improvement in each indicator, however, should be interpreted using the specific empirical results obtained in this study.

Developing Systematic Thinking, Decision-Making, and Conclusion-Drawing Skills

Systematic thinking is an important component of critical reasoning because it enables students to organize information, examine relationships among variables, and evaluate possible solutions in a logical sequence. In the context of STEM-integrated learning, students may be required to observe a problem, collect relevant data, analyze the information, and determine the relationships between different components. These activities encourage students to approach problems methodically rather than make decisions based on assumptions or incomplete information. When supported by the TaRL approach, systematic problem-solving activities can be adapted to students' learning readiness, allowing them to develop these skills progressively (Andriani & Trisnowati, 2025; Ardana et al., 2026).

Decision-making is closely related to systematic thinking because it requires students to evaluate available information and select the most appropriate course of action. Effective decision-making involves several interconnected processes, including observing data, verifying information, considering alternatives, drawing conclusions, and determining appropriate actions. Students who engage in these processes learn to evaluate the possible consequences of their choices and provide logical justifications for their decisions. Karahan et al. (2023) highlight decision-making as an important thinking ability that students need to develop. Therefore, learning activities that require students to compare alternatives and justify their choices can provide opportunities to strengthen both systematic thinking and critical reasoning.

Drawing conclusions is another essential indicator because it requires students to synthesize information and determine whether their conclusions are supported by the evidence they have examined. Rather than simply repeating information, students need to interpret findings, identify relevant relationships, and formulate conclusions that logically follow from their analysis. This process involves evaluating the quality of information and ensuring that the conclusions correspond to the evidence available. Critical reasoning encompasses the ability to analyze information, evaluate its relevance and reliability, and draw conclusions based on the information possessed (Ariadila, et al., 2023). Consequently, conclusion-drawing should be understood as the outcome of a systematic process of information analysis and evaluation.

The five indicators of critical reasoning are therefore interconnected rather than independent abilities. Identifying relevant information provides the basis for analysis, analytical reasoning supports the development of effective explanations, and systematic thinking helps students evaluate alternatives before making decisions and drawing conclusions. The integration of

STEM and TaRL can provide opportunities for students to practice these processes through contextual investigations and structured problem-solving activities. Nevertheless, claims regarding improvements in specific indicators should be supported by the corresponding findings from the study. By connecting the theoretical framework with empirical results, the discussion can provide a more comprehensive explanation of how integrated STEM learning may contribute to the development of students' critical reasoning abilities.

Conclusion

This study shows that STEM-integrated TaRL learning significantly improves students' critical reasoning abilities. This is evidenced by the results of the independent sample t-test analysis, with significance values of 0.000 and 0.006 (<0.05), indicating that TaRL learning has a significant effect on improving critical reasoning abilities. This demonstrates that TaRL learning is more effective than conventional learning in schools in developing critical reasoning abilities. This improvement is reflected in students' abilities to identify, reason and analyze, provide effective reasons, use systems thinking, and draw conclusions. The STEM-integrated TaRL learning approach demonstrates learning through experiences adjusted to students' levels of learning readiness based on their initial abilities. The learning process involves contextual problem-solving, understanding the *engineering design process*, utilizing technology, applying mathematical concepts, and engaging in collaborative discussions, which can effectively develop critical reasoning abilities. This learning approach is not only pedagogically effective but also relevant for preparing students to face the development of 21st-century skills.

References

- Ahyar., Nurhidayah, N., & Saputra A. (2022). Implementasi Model Pembelajaran TaRL dalam Meningkatkan Kemampuan Literasi Dasar Membaca Peserta Didik di Sekolah Dasar Kelas Awal. *Jurnal Ilmiah Ilmu Pendidikan*, 5(11), 5241-5246. DOI: [10.54371/jiip.v5i11.1242](https://doi.org/10.54371/jiip.v5i11.1242)
- Andriani, A., & Trisnowati, E. (2025). Integrating problem-based learning and the teaching at the right level approach to improve students' collaboration skills. *International Journal of Education and Teaching Zone*, 4(3), 406-422.
- Ardana, D. A., Mawarsari, V. D., & Sulistyaningsih, D. (2026). Systematic literature review: Integrating problem-based learning with teaching at the right level to enhance problem solving skills. *Jurnal THEOREMS (The Original Research of Mathematics)*, 11(1), 1-15. <https://doi.org/10.31949/th.v11i1.16985>
- Ariadila, S. N., Silalahi, Y. F. N., Fadiyah, F. H., Jamaludin, U., & Setiawan, S. (2023). Analisis Pentingnya Keterampilan Berpikir Kritis Bagi Siswa. *Jurnal Ilmiah Wahana Pendidikan*, 9(20), 664-669. DOI: <https://doi.org/10.5281/zenodo.8436970>
- Bhakti, Y. B., Astuti, I. A. D., Okyranida, I. Y., Asih, D. A. S., Marhento, G., Leonard, L., & Yusro, A. C. (2020, February). Integrated STEM project based learning implementation to improve student science process skills. In *Journal of Physics: Conference Series* (Vol. 1464, No. 1, p. 012016). IOP Publishing.
- Bordoh, A. (2025, May). Social studies teachers' knowledge and criteria for selecting and utilizing instructional materials in teaching and learning. In *Frontiers in Education* (Vol. 10, p. 1560679). Frontiers Media SA. <https://doi.org/10.3389/feduc.2025.1560679>
- Cahyani, C. P. N., & Setiawan, E. P. (2024). How motivation mediated science achievement

- of Indonesian students: A path analysis on PISA 2018 data. *Jurnal Penelitian Pendidikan IPA*, 10(12), 10493-10501. <https://doi.org/10.29303/jppipa.v10i12.10012>
- Cahyono, S. D. (2022). Melalui Model Teaching at Right Level (TARL) Metode Pemberian Tugas untuk Meningkatkan Motivasi dan Hasil Belajar Peserta Didik Perencanaan Usaha Pengolahan Makanan Awetan dari Bahan Pangan Nabati di Kelas X . MIA . 3 MAN 2 Payakumbuh Semester. *Jurnal Pendidikan Tambusai*: 6(2) 12407 12418.
- Casmini, N. L., Redhana, I. W. ., & Suma, I. K. (2024). Exploring the Impact of Teaching Strategies on Students' Critical Thinking Skills in Chemistry. *Mimbar Ilmu*, 29(3), 422–431. <https://doi.org/10.23887/mi.v29i3.90701>
- Davidi, E. I. N., Senen, E., & Supardi, K. (2021). Integrasi Pendekatan STEM (Science, Technology, Enggeenering, and Mathematic) untuk Peningkatan Keterampilan Berpikir Kritis Siswa Sekolah Dasar. *Jurnal Pendidikan dan Kebudayaan*, 11(1), 11-22. <https://doi.org/10.24246/j.js.2021.v11.i1.p11-22>
- Fachrunnisa, R., Suwono, H., Yuenong, C., Sutaphan, S., & Praipayom, N. (2021). Eco-friendly fashion: A STEM Sandpit Project in Indonesia Senior High School. *Journal of Physics: Conference Series* 1835. doi:10.1088/1742-6596/1835/1/012046
- Gaybullaevna, D. N., & Jonpulatovna, S. M. (2021). Innovative approach to education is a factor for developing new knowledge, competence and personal qualities. *Asian Journal of Multidimensional Research (AJMR)*, 10(1), 148-153. <http://dx.doi.org/10.5958/2278-4853.2021.00014.8>
- González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of Education 4.0 in 21st century skills frameworks: systematic review. *Sustainability*, 14(3), 1493. <https://doi.org/10.3390/su14031493>
- Haryani, S., Masfufah, Wijayati, N., & Kurniawan, C. (2018, March). Improvement of metacognitive skills and students' reasoning ability through problem-based learning. In *Journal of Physics: Conference Series* (Vol. 983, No. 1, p. 012174). IOP Publishing. <https://doi.org/10.1088/1742-6596/983/1/012174>
- Ivanka, T. A., & Setiawan, E. P. (2025). Factors Affecting Indonesian Students' Science Achievement: A Multilevel Analysis of the PISA Dataset. *Jurnal Penelitian Pendidikan IPA*, 11(12), 249-260. <https://doi.org/10.29303/jppipa.v11i12.11602>
- Jannah, E. U., Fathani, A. H., & Fuady A. (2022). Analisis Kemampuan Berpikir Kritis Peserta Didik dalam Memecahkan Masalah Matematis ditinjau dari *Mathematical Belief*. *JPM UIN Antasari*, 09(2).
- Jazuli, L. (2022). Teaching At The Right Level (Tarl) Through The All Smart Children Approach (Sac) Improves Student's Literature Ability. *Progres Pendidikan*, 3(3), 156–165. <https://doi.org/10.29303/prospek.v3i3.269>
- Karahan, E., Bozan, M. A., Akcay, A. O., & Akcay, I. M. (2023). An Investigation of Primery School Students' Critical Thinking Dispositions and Decision-Making Skills. *International Journal of Educational Research Review*, 8(2), 137-150. <https://doi.org/10.24331/ijere.1205285>
- Kurniati, P., Kelmaskouw, A. L., Deing, A., Bonin, B., & Haryanto, B. A. (2022). Model proses inovasi kurikulum merdeka implikasinya bagi peserta didik dan guru abad 21. *Jurnal Citizenship Virtues*, 2(2), 408–423.

- Lianawati, D., Utami, B., & Rahmawati, I. (2025). Analisis Keterampilan Berpikir Kritis Siswa Kelas X MIPA 3 SMAN 4 Kediri. *Prosiding Seminar Nasional Hayati*, 3(1), 211-215. <https://doi.org/10.29407/fngx7z61>
- Mahartika, I., Aisyah, S. N. I., Meisyalla, L. N., & Ilhami, A. (2023). What are the Characteristics of Learners and the Variations of Non-Electronic Learning Media?. *JPI (Jurnal Pendidikan Indonesia)*, 12(2), 305-316. <https://doi.org/10.23887/jpiundiksha.v12i2.51325>
- Mollick, E. (2022). Embracing Generative AI for Authentic Learning. *Creative Education*, 15(1), 1-6.
- Muin, F. (2023). Adapting Teaching at The Right Level (TaRL) in English Instruction. *The Proceedings of International Conference: Ronggolawe English Teaching Conference (RETCO)*, 11(1), 1-10. Retrieved from <http://prosiding.unirow.ac.id/index.php/retco/article/view/1712>
- Nikmah, N., Musbikhin, & Mandala, A. S. (2024). Development of STEM-Collaborative Learning Based Arithmetic Teaching Material for Enhancing Students' Critical Thinking. *Jurnal Pengembangan Pembelajaran Matematika (JPPM)*, 6(2), 94-107. <https://doi.org/10.14421/jppm.2024.62.94-107>
- Ningrum, M. C., Juwono, B., & Sucahyo, I. (2023). Implementasi Pendekatan TaRL untuk Meningkatkan Motivasi Belajar Peserta Didik pada Pembelajaran Fisika. *PENDIPA Journal of Science Education*, 7(1), 94-99. DOI: <https://doi.org/10.33369/pendipa.7.2.94-99>
- Nurdianti, Rizky., Atmojo, I.R.W., Sularmi. (2020). Penerapan Metode Edutainment berbasis Guided Discovery Learning (GDL) untuk Meningkatkan Keterampilan Menyimpulkan pada Pembelajaran IPA Kelas III Sekolah Dasar. *Didaktika Dwija Indria*, 8(4), 47-53. DOI: <https://doi.org/10.20961/ddi.v8i04.42294>
- Pandu, R., Purnamasari, I., & Nuvitalia, D. (2023). Pengaruh Pertanyaan Pemantik Terhadap Kemampuan Bernalar Kritis dan Hasil Belajar Peserta Didik. *Pena Edukasia*, 1 (2): 127-134.
- Salim, Rodia., Adam, Adiyana., Silawane, Nurjannah., Ali, R.R., Mayabubun, Yayan., & Dahlan, Albar. (2023). Tingkat Keberhasilan Pembelajaran di Perguruan Tinggi: (Analisis Metode Diskusi untuk Meningkatkan Keterampilan Berpikir Kritis). *JUANGA: Jurnal Agama dan Ilmu Pengetahuan*, 9(2). DOI : 10.59115/juanga.v9i0
- Sapia, P., Napoli, F., & Bozzo, G. (2022). The Lawson's test for scientific reasoning as a predictor for University formative success: A prospective study. *Education Sciences*, 12(11), 1-15. <https://doi.org/10.3390/educsci12110814>
- Solihin, R. R., Susanto, T. T. D., Fauziah, E. P., Yanti, N. V. I., & Ramadhania, A. P. (2024). The efforts of Indonesian government in increasing teacher quality based on PISA result in 2022: A literature review. *Perspektif Ilmu Pendidikan*, 38(1), 57-65.
- Stevenson, E., van Driel, J., & Millar, V. (2026). How to support teacher learning of integrated STEM curriculum design. *Journal for STEM Education Research*, 9(1), 59-84. <https://doi.org/10.1007/s41979-024-00133-0>
- Sugiyono. (2023). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: ALFABETA.

- Sundayana, H Rostina. (2018). *Statistika Penelitian Pendidikan*. Bandung: Penerbit Alfabeta.
- Sutiani, A., Situmorang, M., & Silalahi, A. (2021). Implementation of an inquiry learning model with science literacy to improve student critical thinking skills. *International Journal of Instruction*, 14(2), 117-138.
- Trilling, B., and Fadel, C. (2009). *21st Century Skills Learning for Life in Our Times*. San Fransisco: Jossey-Bass A Wiley Imprint.
- Wechsler, S. M., Saiz, C., Rivas, S. F., Vendramini, C. M. M., Almeida, L. S., Mundim, M. C., & Franco, A. (2018). Creative and critical thinking: Independent or overlapping components?. *Thinking skills and creativity*, 27, 114-122.
- Widodo, S. A. (2018). Selection of Learning Media Mathematics for Junior School Students. *Turkish Online Journal of Educational Technology-TOJET*, 17(1), 154-160.
- Wijaya, T. T., Hidayat, W., Hermita, N., Alim, J. A., & Talib, C. A. (2024). Exploring contributing factors to PISA 2022 mathematics achievement: Insights from Indonesian teachers. *Infinity Journal*, 13(1), 139-156. <https://doi.org/10.22460/infinity.v13i1.p139-156>
- Wulandari, Mursali, S., & Dewi, I. N. (2025). Pengaruh Model Pembelajaran *Guided Inquiry* Berbantuan *Virtual Laboratory* Terhadap Keterampilan Berpikir Kritis Peserta Didik. *Biocaster: Jurnal Kajian Biologi*, 5(3), 202-216. <https://doi.org/10.36312/biocaster.v5i3.441>
- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature human behaviour*, 8(10), 1839-1850.
- Zeichner, K. M. (2017). Competition, economic rationalization, increased surveillance, and attacks on diversity: Neo-liberalism and the transformation of teacher education in the US 1. In *The struggle for the soul of teacher education* (pp. 40-62). Routledge. <https://doi.org/10.4324/9781315098074>