Available online at: http://journal.uny.ac.id/index.php/jpip

Jurnal Penelitian Ilmu Pendidikan

Volume 19, Nomor 1, 23—36, 2026

A counterbalanced quasi-experiment comparing problem-based learning with zep quiz, zep quiz alone, and traditional lecture on critical thinking in high school mechanics

Fikriyah Luthfiyani[1] and Iwan Permana Suwarna[1,*]

[1]Faculty of Tarbiyah and Teacher Training, Syarif Hidayatullah State Islamic University, Jakarta, Indonesia

*Corresponding Author. e-mail: iwan.permana.suwarna@uinjkt.ac.id

Abstract

Developing critical thinking in physics is a priority in 21st-century education, yet evidence comparing the effectiveness of different instructional approaches, particularly those integrating problem-based learning (PBL) with digital quiz platforms, remains limited. This study compared the effects of three learning models, PBL assisted by Zep Quiz, Zep Quiz alone, and traditional lecture, on 11th-grade students' critical thinking skills in mechanics. A counterbalanced within-subject quasi-experimental design was employed with 102 students from SMAN 4 Tangerang Selatan. Each student experienced all three models across three topics (Newton's Laws, Momentum & Impulse, Rotational Motion) in a rotated order. Critical thinking was measured using a validated 18-item test based on Facione's six indicators (reliability: 0.78–0.79 for two topics, 0.45 for Rotational Motion). Data were analyzed using N-gain, repeated-measures ANOVA, and ANCOVA with pretest scores as covariate. All models improved critical thinking with moderate-to-high N-gain (0.73–0.78). Repeated-measures ANOVA showed significant differences between models ($F(2,202)=4.920$, $p=.008$, $\eta^2=.046$). However, after controlling for pretest scores, ANCOVA revealed no statistically significant differences between models ($F(2,202)=2.229$, $p=.109$, $\eta^2=.015$, small effect). Topic characteristics explained substantially more variance ($\eta^2=.642$) than model variations ($\eta^2=.015$). The largest gains occurred in Momentum & Impulse (N-gain=0.89), the lowest in Rotational Motion (N-gain=0.55). The three learning models are comparably effective in improving critical thinking in mechanics. Topic characteristics, particularly conceptual complexity, outweigh model variations in determining learning gains. Physics educators should prioritize alignment between instructional design and topic demands rather than focusing solely on model selection. Future research should extend to other physics topics and investigate longer-term retention.

Keywords: Critical Thinking, Physics Learning, Problem-Based Learning, Interactive Quizzes, Quasi-Experiment

How to Cite (APA): Luthfiyani, F., & Suwarna, I. P. (2026). A counterbalanced quasi-experiment comparing problem-based learning with zep quiz, zep quiz alone, and traditional lecture on critical thinking in high school mechanics. *Jurnal Penelitian Ilmu Pendidikan, 19*(1), 23—36. https://doi.org/10.21831/jpip.v19i1.97460

Received 8-4-2026; Received in revised from 12-04-2026; Accepted 18-04-2026

This is an open-access article under the CC–BY-SA license.

INTRODUCTION

Critical thinking skills are one of the key competencies in the 21st century that must be developed through science education. Critical thinking includes the ability to analyze, evaluate, draw logical conclusions, and reflect on the thinking process (Abrami et al., 2015). Critical thinking skills in physics learning not only emphasize mastery of concepts, but also develop scientific reasoning and complex conceptual problem solving (Moeis & Maulana, 2021). The Programme for International Student Assessment (PISA) report shows that Indonesian students

still have difficulty completing questions based on contextual reasoning and analysis, indicating that the learning process is not yet optimal in supporting the development of critical thinking (OECD, 2019). A teacher-centered learning approach is one of the factors contributing to low critical thinking skills among students. In this pattern, teachers act as the primary source of information, while students function mainly as passive recipients. while students are merely passive recipients (Goodwin, 2024).

A similar condition was found in physics learning at SMAN 4 Tangerang Selatan. Weaknesses in the conventional physics learning approach also caused students to experience conceptual difficulties and misconceptions that hindered critical thinking. Many students tend to memorize formulas without understanding the relationships between concepts, so their knowledge is procedural rather than conceptual (Aulia et al., 2025). This can be seen in their inability to distinguish between concepts such as force and resultant force, as well as momentum and impulse.

Students often experience difficulties when they have to represent physics problems in the form of diagrams or graphs, such as free-body diagrams or force-time graphs, even though these representations are important components in the physics modeling process (Jannah et al., 2022; Takaoğlu, 2024). As a result, students are less able to translate real phenomena into a logical scientific framework. Students' perception of physics as an abstract subject also hinders the development of critical thinking. Learning that focuses on solving routine problems makes it difficult for students to relate physics concepts to real life, such as vehicle technology or sports equipment (Kotsis, 2025). As a result, learning motivation decreases, and critical thinking skills weaken. To overcome this problem, various alternative learning approaches have begun to be implemented by physics teachers. The emergence of awareness of the importance of innovation in learning has encouraged teachers to adopt models such as Problem-Based Learning (PBL) or digital media such as Zep Quiz (Adhelacahya et al., 2023).

Teachers generally determine learning models based on their teaching experience and habits, so the effectiveness of each model has not been empirically tested (Susanto & Airlanda, 2023). Previous studies have shown that PBL is effective in improving critical thinking and problem-solving skills (Dewi et al., 2024). The use of digital quizzes such as Zep Quiz can increase student motivation and engagement in learning (Ahsan et al., 2026). Zep Quiz is a digital game-based learning evaluation platform that integrates interactive quiz concepts with a virtual environment (metaverse classroom). This platform allows teachers to design various types of questions, such as multiple choice, true-false, and stimulus-based questions, which students can access online through a specific class code.

Previous studies comparing Indonesian and Singaporean education systems mostly focused on general educational structures or policy descriptions. Limited studies have specifically compared curriculum policy and learning approaches within the context of improving educational quality using recent policy developments and international evidence. Therefore, this study seeks to comparatively analyze curriculum policies and learning methods in Indonesia and Singapore to identify strengths, challenges, and potential implications for improving educational quality in Indonesia such as Peter A. Facione's framework as discussed by Al Kandari and Al Qattan (2020).

This study aims to analyze the differences in effectiveness between the PBL assisted Zep Quiz, the use of Zep Quiz without PBL integration, and the traditional lecture method in improving students' critical thinking skills in physics and mechanics material. This study also aims to examine the contribution of each learning model to improving critical thinking skills by controlling students' initial abilities through Repeated Measures ANOVA and ANCOVA analysis. Thus, this study seeks to identify the most effective learning model in significantly developing students' critical thinking skills.

This study is expected to provide a more controlled and methodological empirical contribution to the development of physics learning strategies that are relevant to the demands of 21st-century competencies. Unlike previous studies, which generally tested learning models

separately and did not adequately control for the influence of class characteristics and material, this study used a *counterbalanced* design to minimize the effects of treatment order and class variation. This methodological approach allowed for a more objective and systematic comparison of the effectiveness of learning models in examining the integration of Problem-Based *Learning* and digital quizzes in physics learning.

Specifically, this study answers the following questions:

1. Is there a difference in the improvement of students' critical thinking skills between the three learning approaches?
2. Do the characteristics of mechanical materials affect the level of improvement in students' critical thinking skills?
3. How do students' achievements on each critical thinking indicator based on Facione's framework differ after the implementation of the three approaches?

METHODS

This study employed a qualitative comparative descriptive approach to analyze curriculum policies and learning methods in Indonesia and Singapore. A documentation study was used as the primary data collection method. Data sources consisted of official policy documents, government reports, international assessment reports, and scientific publications published between 2015 and 2025.

In line with the problems in the introduction, which show that the effectiveness of various learning models on students' critical thinking skills has not been comprehensively described and tends to be studied separately, this study uses a quasi-experimental approach with a counterbalanced repeated measures design, in which each student experiences the three learning models on different materials. This design was chosen to control *for the order effect* and minimize bias due to differences in class characteristics, so that comparisons of the effectiveness of learning models could be made more objectively (Cresswell & Speelman, 2020). The *counterbalanced* design in this study is an extension of the conventional quasi-experimental design, which generally applies one treatment to one class. The application of this design allows each learning model to be applied alternately to different physics materials, so that the variation in students' critical thinking skills that is produced better reflects the influence of the learning model rather than the influence of specific material or class characteristics. This is in line with the research objective of comparing the effectiveness of the PBL assisted Zep Quiz model, the Zep Quiz model, and the traditional lecture method model in physics learning.

Subjects and Sample

This research was conducted in three 11th-grade classes at SMAN 4 Tangerang Selatan in the odd semester of the 2025/2026 academic year, with a total of 102 students, consisting of 34 students in each class, namely classes XI-1, XI-2, and XI-3. Classes were selected based on student interests, where in the current curriculum, high school students are grouped into three fields of interest, namely engineering, health, and social studies. Therefore, the learning in this study focused on material relevant to the field of engineering. The learning material used included Newton's laws of motion, momentum, and impulse, and rotational motion in the topic of motion dynamics. Of the initial 110 students, 8 students were not included in the data analysis because they were no longer actively attending school or had low attendance rates during the learning process. Thus, the composition of the research subjects analyzed it was consisted of 72 male students and 30 female students.

Treatment and Counterbalanced Design

The counterbalanced design scheme was arranged by applying the three learning models alternately in each class and material, so that each class received all treatments in different material contexts. The distribution of the application of learning models in each class and material is presented in Table 1.

Table 1. Counterbalanced Design in 3 Classes

Class	Subject		
	Force and Newton's Laws	Momentum and Impulse	Rotational Motion

XI-1	PBL assisted Zep Quiz	Zep Quiz	Traditional Lecture Method
XI-2	Zep Quiz	Traditional Lecture Method	PBL assisted Zep Quiz
XI-3	Traditional Lecture Method	PBL assisted Zep Quiz	Zep Quiz

Based on Table 1, the research results are not influenced by specific subject matter factors or class characteristics alone, but rather reflect the influence of the learning model in a more methodological manner (Valentine et al., 2018). All learning in the three classes was conducted by the same physics teacher as the main instructor. Another physics teacher acted as an observer to ensure consistency in the implementation of the learning model according to the designed scenario. This arrangement was made to minimize the teacher effect and increase the internal validity of the research. This step was taken to improve the reliability of the treatment and strengthen the internal validity of the research.

Research Instruments

Critical thinking skills were measured using a contextual multiple-choice test developed based on six critical thinking indicators according to Peter A. Facione (Al Kandari & Al Qattan, 2020), namely interpretation, analysis, evaluation, inference, explanation, and self-regulation. The instrument consisted of 18 items covering the three topics of mechanics. The distribution of items for each indicator and topic was arranged based on the instrument development blueprint. The content validity of the instrument was assessed through expert judgment by two physics education lecturers and one experienced physics teacher. Item quality analysis was conducted using the Classical Test Theory (CTT) approach through ANATES software, including internal reliability, difficulty level, and discrimination power tests. The reliability coefficient was calculated using the Spearman-Brown formula. Reliability in the topics of Momentum and Impulse, as well as Force and Newton's Laws, was in the high category, while in the topic of Rotational Motion, it was in the low-medium category. These limitations were considered in the interpretation of the research results.

Implementation Stage

The research procedure consisted of the following stages.

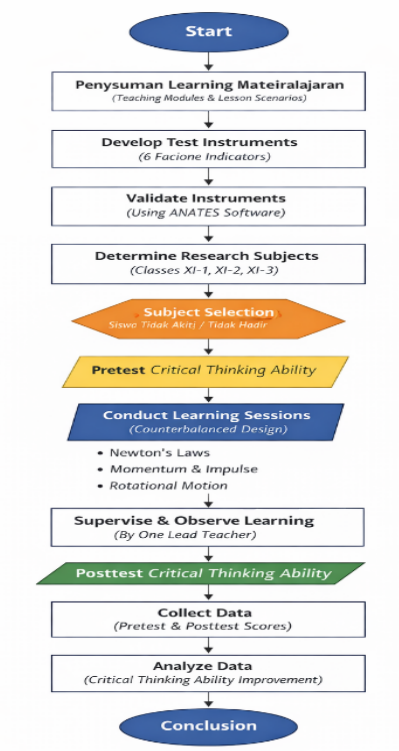


Figure 1. Research Implementation Flow

(1) Preparation Stage

In the preparation stage, the researcher developed learning tools, teaching modules, and critical thinking ability test instruments. The test instruments were developed based on six indicators of critical thinking ability according to Peter A. Facione (Al Kandari & Al Qattan, 2020), namely *interpretation, analysis, evaluation, inference, explanation,* and *self-regulation*. All research instruments were then validated using Anates software to ensure the feasibility and consistency of the measuring instruments used.

(2) Implementation Stage

This research was conducted from October 21 to November 21, 2025. Before learning, all students were given a pretest to measure their initial critical thinking skills. Learning was then carried out using a *counterbalanced* design with a duration of (5 x 90 minutes) with the following distribution:

a) 2 sessions on Newton's Laws of Motion,
b) 2 sessions on Momentum and Impulse,
c) 1 session for Rotational Motion

The difference in the allocation of learning time for each subject is adjusted according to the level of complexity of the concept and the depth of discussion. The topics of Force and Newton's Laws, as well as Momentum and Impulse, were allocated two sessions because they had a higher level of misconception and mathematical complexity, while the topic of Rotational Motion was given in one session because it was more focused and analogous to translational motion. This variation in learning duration was one of the limitations of the study that needed to be considered when interpreting the differences in improvement between topics.

(3) Data Collection Stage

Research data were collected through critical thinking tests administered in the form of pre-tests and post-tests. The test instrument consisted of 18 questions with a maximum score of 100, which measured students' critical thinking skills based on six Facione indicators. The data obtained were then analyzed to determine changes in students' critical thinking skills after the learning treatment.

The documents were selected purposively based on relevance to curriculum systems, learning methods, and educational quality indicators. Key sources included curriculum policy documents from the Indonesian Ministry of Education and the Singapore Ministry of Education, PISA reports, and peer-reviewed journal articles.

Data Processing and Analysis

The critical thinking ability pretest and posttest scores were first converted to a scale of 0– 100, and then the normalized gain (N-gain) value was calculated to describe the increase in student ability. The N-gain calculation refers to a formula commonly used in science education research to measure learning effectiveness. This study used a counterbalanced repeated measures design, in which each student experienced all three learning models on different materials. Thus, the learning model was analyzed as a within-subject factor, while the variation in the order of treatment was controlled through class settings.

Data analysis was conducted using content analysis and comparative analysis techniques. The analysis process involved three stages: data identification, data categorization, and comparative interpretation. The findings were then interpreted to identify similarities, differences, strengths, and challenges in both educational systems.

Data analysis began with descriptive statistics to determine the mean and distribution of scores in each learning condition. Next, prerequisite tests were conducted, namely the normality of residuals test and the homogeneity of variance test, to ensure that the assumptions of parametric analysis were met (Valentine et al., 2018). To test the differences in critical thinking skills between learning models in the same subjects, Repeated Measures ANOVA was used with the help of IBM SPSS Statistics Version 25 software. This analysis considered the correlation between measurements because each student received all treatments.

Although the research design was within-subject, an additional analysis using Analysis of Covariance (ANCOVA) was performed with pretest scores as covariates and posttest scores as dependent variables. In a repeated measures design, ANOVA considers the correlation between measurements in the same subject but does not explicitly control for the linear contribution of initial ability to the final score in each learning condition. Given that each model was applied to different materials, variations in initial ability between topics could potentially affect posttest scores even though the subjects were the same. Therefore, the use of ANCOVA aims to statistically control for variation in initial ability and provide additional confirmation of the Repeated Measures ANOVA findings. Thus, ANCOVA is used as a complementary analysis to strengthen the validity of the inference, not as a substitute for the main within-subject analysis (Rutherford, 2001).

RESULTS AND DISCUSSION

Results

This section presents the results of the study on the improvement of students' critical thinking skills after the application of the PBL assisted Zep Quiz, Zep Quiz, and the Traditional lecture method with a counterbalanced quasi-experimental design. The analysis focused on pretest scores, posttest scores, and N-gain scores on the topics of Newton's Laws and Forces, Momentum and Impulse, and Rotational Motion to ensure that the improvements obtained were not influenced by differences in material or class characteristics. The comparison of critical thinking skill improvement between learning models was analyzed using descriptive statistics and ANOVA tests, followed by ANCOVA tests to control for the influence of students' initial abilities. In addition, student achievement was also analyzed based on six critical thinking indicators according to Facione (Al Kandari & Al Qattan, 2020), which were measured using instruments developed by researchers based on the conceptual framework. This analysis aims to provide a comprehensive empirical picture of the relative effectiveness of each learning model in physics education.

Empirical Validity Test Results using Anatest

The critical thinking ability instrument in this study was developed independently based on six indicators according to Peter A. Facione (Al Kandari & Al Qattan, 2020), namely interpretation, analysis, evaluation, inference, explanation, and self-regulation. The instrument was compiled in the form of a contextual multiple-choice test adapted to the material on Newton's Forces and Laws, Momentum and Impulse, and Rotational Motion. The quality of the instrument was analyzed using the Classical Test Theory (CTT) approach through the Anatest program to test internal consistency, difficulty level, and item discrimination power. The r_xy value was reported as the average correlation coefficient between the item score and the total score, which was used as an indicator of the empirical validity of the item.

Table 2. Results of Instrument Validity and Reliability Analysis

Material	M	SD	r_{xy}	Reliability Spearman-Brown	Category
Momentum and Impulse	16.29	2.03	0.64	0.78	High
Newton's Laws of Motion	15.05	3.13	0.66	0.79	High
Rotational Motion	10.81	1.88	0.29	0.45	Moderate

Based on Table 2, the instruments on Momentum and Impulse (r = 0.78) and Force and Newton's Laws (r = 0.79) showed high reliability. In the Rotational Motion material, the reliability obtained was r = 0.45, which is in the moderate category. In the Classical Test Theory framework, reliability estimates are influenced by the number of items and the heterogeneity of respondent scores. With a relatively limited number of items, this reliability value is still acceptable in the context of educational research, although the interpretation of the results in this material needs to be done carefully.

Analysis of Critical Thinking Skill Improvement Based on N-Gain

The improvement in students' critical thinking skills was analyzed using *normalized gain* (N-gain) to describe the effectiveness of learning based on a comparison of *pretest* and *posttest* scores. The following are the average N-Gain per material and model.

N-gain calculations were performed using the following equation:

$$g = \frac{S_{post} - S_{pre}}{S_{maks} - S_{pre}} \quad (1)$$

The distribution of the average N-gain based on material and learning model is presented in Figure 2.

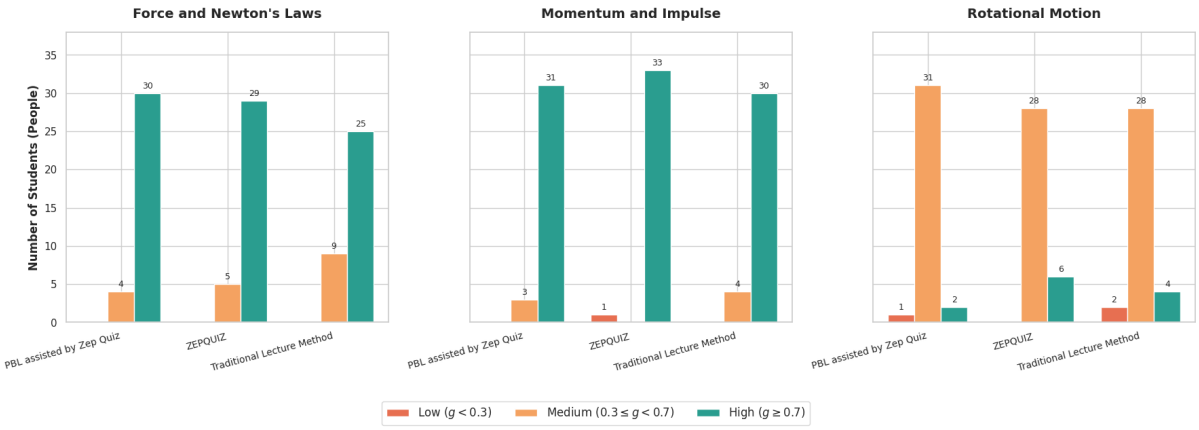


Figure 2. Graph of N-Gain values based on Learning Material and Model

Based on Figure 2 and referring to the N-gain classification criteria (high: g > 0.7; medium: 0.3 ≤ g ≤ 0.7; low: g < 0.3), most students showed an increase in critical thinking skills in the high category, especially in the Force and Newton's Laws and Momentum and Impulse materials in all learning models. Conversely, in the Rotational Motion material, the increase in critical thinking skills tended to be in the moderate category. To ensure the significance of these differences, inferential statistical analysis was performed using Repeated Measures ANOVA and ANCOVA after the data met the prerequisite tests of normality and homogeneity.

Normality and Homogeneity Prerequisite Tests

Prerequisite tests were conducted prior to inferential analysis to ensure that the data met the statistical assumptions required in repeated-measures ANOVA and ANCOVA analyses. Normality tests were conducted on the distribution of scores in each learning condition as a preliminary check for distribution deviations. The data analyzed included pretest and posttest scores of students' critical thinking skills in each learning model, which were treated as within-subject factors in a counterbalanced design.

Normality Test

Normality tests were performed on model residuals using the Shapiro–Wilk test as part of the Repeated Measures ANOVA assumption test. A normality test was performed on the model residuals using the Shapiro–Wilk test as part of the Repeated Measures ANOVA assumption test. The test results showed that all residuals had a significance value < 0.05, which statistically indicated a deviation from the normal distribution. However, visual inspection through a Q-Q plot showed that the data distribution relatively followed the diagonal line and did not show extreme deviations. In addition, the large sample size (N = 102) allowed for the application of the Central Limit Theorem, so that the Repeated Measures ANOVA analysis could still be continued because this method is robust against minor violations of normality (Rencher, 2002). The percentage of Q-Q plots for normality testing is shown in Figure 3.

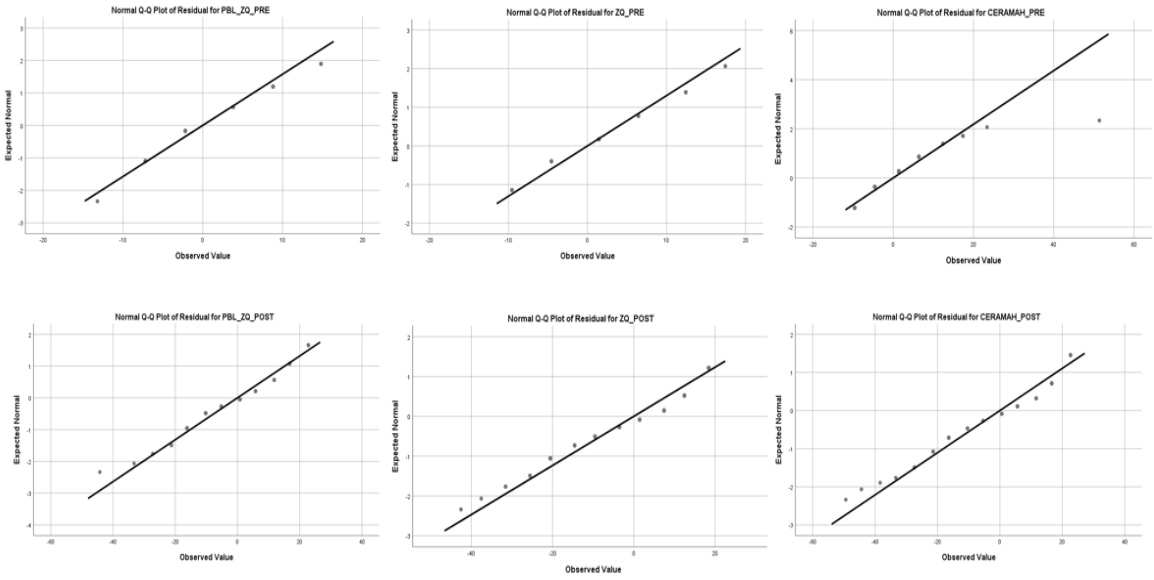


Figure 3. Normality Q-Q Plots

Homogeneity Test

The homogeneity of variance test was performed using Levene's test as a prerequisite for ANCOVA analysis to ensure the equality of variances between treatment conditions. The results of the homogeneity of variance test are presented in Table 3.

Table 3. Homogeneity Test Results

	Levene Statistic	df1	df2	Sig.
CTT Pretest Score	2.676	2	303	.070
Posttest CTT Score	2,719	2	303	.068

Based on the table 3, the Levene test results, the significance values for the *pretest* scores (p = 0.070) and *posttest* scores (p = 0.068) were obtained. The test results show significance values greater than 0.05 for all variables, so the assumption of homogeneity is fulfilled.

Improvement in students' critical thinking skills in each learning material

The results of the descriptive statistical test of students' critical thinking ability scores based on *pretest*, *posttest*, and N-gain categories for each learning material are presented in Table 4.

Table 4. Descriptive Statistics of Critical Thinking Ability (CCTT) Scores Based on Learning

Materials	Pretest (Mean ± SD)	Posttest (Mean ± SD)	N-Gain
Newton's Laws of Motion	14.72 ± 7.45	83.91 ± 12.33	High
Momentum and Impulse	14.02 ± 8.36	90.34 ± 12.19	High
Rotational Motion	15.64 ± 7.76	61.94 ± 9.14	Moderate

Based on Table 4, all learning materials demonstrated improvements in students' critical thinking skills. The greatest improvement was observed in the Momentum and Impulse topic. This was reflected by the higher average posttest scores and the high N-gain category. In contrast, the Rotational Motion topic showed a lower level of improvement, which fell within the moderate N-gain category. This difference may be attributed to the conceptual characteristics of rotational motion. These characteristics tend to be more abstract and mathematically demanding. Often, they require stronger visual representations and more gradual conceptual development. Such learning processes may not have been fully accommodated within a single instructional session. These results suggest that variations in learning gains across topics may be influenced by differences in conceptual complexity and representational demands. To

examine whether these differences were statistically significant, a repeated-measures ANOVA was conducted on the pretest, posttest, and N-gain scores. The results are showed in the table 5 below.

Table 5. Results of the Repeated Measures ANOVA test of Critical Thinking Skills for Each Topic

Variable	GH (Mean ± SD)	MI (Mean ± SD)	GR (Mean ± SD)	F (2, 202)	p	η^2
Pretest	14.72 ± 7.45	14.02 ± 8.36	15.64 ± 7.76	1.128	.326	.011
Posttest	83.91 ± 12.33	90.34 ± 12.19	61.94 ± 9.14	180.280	< .001	.641
N-gain	0.81 ± 0.14	0.89 ± 0.15	0.55 ± 0.11	181.018	< .001	.642

Mauchly's Test of Sphericity[a]							
					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
MATERIAL	.995	.523	2	.770	.995	1.000	.500

Based on the table 5, the results of the Repeated Measures ANOVA test, there was no significant difference in initial ability between materials before treatment was given. $F(2, 202) = 1.128$, $p = .326$, $\eta^2 = .011$, indicating that the initial ability of students could be considered relatively equal. In the posttest scores, a significant difference was found between materials. $F(2, 202) = 180.280$, $p < .001$, $\eta^2 = .641$. A significant difference was also found in the N-gain value, $F(2, 202) = 181.018$, $p < .001$, $\eta^2 = .642$. The very large effect size values were likely influenced by differences in learning time allocation and differences in instrument reliability between materials. Lower reliability in the Rotational Motion material had the potential to increase measurement error, so the interpretation of differences between materials needed to be done carefully. The Mauchly test results showed that the sphericity assumption was met ($W = .995$, $p = .770$), and the analysis was conducted without Greenhouse–Geisser correction.

Improvement in students' critical thinking skills based on the learning model

The comparison of the improvement in students' critical thinking skills based on *the* PBL assisted Zep Quiz, Zep Quiz, and Traditional lecture method was analyzed using descriptive and inferential statistics. Descriptive statistics of the *pretest*, *posttest*, and N-gain scores of students' critical thinking skills are presented in Table 6.

Table 6. Descriptive Statistics of Critical Thinking Skills Scores Based on Learning Models

Learning Model	Pretest (Mean ± SD)	Posttest (Mean ± SD)	N-Gain
PBL assisted Zep Quiz	13.20 ± 6.34	77.24 ± 15.17	0.74 ± 0.179
Zep Quiz	15.56 ± 7.68	81.57 ± 16.21	0.778 ± 0.20
Traditional Lecture Method	15.62 ± 9.16	77.39 ± 18.09	0.73 ± 0.01

The table 6 analyzed students' learning outcomes revealed improvements in all groups across pretest, posttest, and N-gain scores. Initial pretest scores were relatively low, with the PBL assisted Zep Quiz group obtaining a mean score of 13.20 ± 6.34, followed by the Zep Quiz group (15.56 ± 7.68) and the Traditional Lecture Method group (15.62 ± 9.16). After the intervention, posttest scores increased substantially in all groups, with the Zep Quiz group achieving the highest mean score (81.57 ± 16.21), followed by the Traditional Lecture Method (77.39 ± 18.09) and PBL assisted Zep Quiz (77.24 ± 15.17). Similarly, the Zep Quiz group demonstrated the highest N-gain value (0.778 ± 0.20), slightly exceeding the PBL assisted Zep Quiz (0.74 ± 0.179) and Traditional Lecture Method groups (0.73 ± 0.01), indicating that all learning models effectively improved students' learning outcomes. Therefore, a Repeated Measures ANOVA test was conducted, as presented in Table 7.

Table 7. Results of the Repeated Measures ANOVA test on critical thinking skills for the Learning Model

Variable	PBL assisted Zep Quiz (Mean ± SD)	Zep Quiz (Mean ± SD)	Traditional Lecture Method (Mean ± SD)	$F(2,202)$	p	η^2
Pretest	13.20 ± 6.34	15.56 ± 7.68	15.62 ± 9.16	3.029	.051	.029
Posttest	77.24 ± 15.17	81.57 ± 16.21	77.39 ± 18.09	4.920	.008	.046
N-gain	0.74 ± 0.18	0.78 ± 0.20	0.73 ± 0.21	3.647	.028	.035

Mauchly's Test of Sphericity[a]							
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
MODEL	.984	1.570	2	.456	.985	1,000	.500

Table 7 stated that the results of the Repeated Measures ANOVA test showed that there was no significant difference in initial ability between measurement conditions before the treatment was given, $F(2,202)= 3.029$, $p = .051$, $\eta^2 = .029$, so that the initial ability of students could be considered relatively equal. In the posttest scores, a significant difference was found between learning models. $F(2,202) = 4.920$, $p = .008$, $\eta^2 = .046$. A significant difference was also found in the N-gain value, $F(2, 202) = 3.647$, $p= .028$, $\eta^2 = .035$. However, the effect size values were in the small category, indicating that the contribution of the learning model to the improvement of critical thinking skills was relatively limited. The Mauchly test results indicated that the sphericity assumption was met (W = 0.984, p = 0.456), so the analysis was conducted without Greenhouse–Geisser correction. Further analysis using ANCOVA with pretest scores as a control showed that the differences between learning models were not statistically significant. The improvement in critical thinking skills was more influenced by factors other than learning model variations. The ANCOVA test result showed in table 8 below.

Table 8. ANCOVA Test Results for Critical Thinking Skills Against Learning Models

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1232.789a	3	410.930	1.499	.215	.015
Intercept	409,990.761	1	409,990.761	1,495.318	.000	.832
Pretest	.449	1	.449	.002	.968	.000
Model	1,222.362	2	611,181	2,229	.109	.015
Error	82,803.237	302	274,183			
Total	1980848.000	306				

The table 8 explained the ANCOVA was conducted to statistically control for students' initial abilities when comparing the effectiveness of the learning models. The results indicated that the pretest scores did not have a significant covariate effect, suggesting that students' initial abilities did not substantially influence the posttest outcomes. Furthermore, the ANCOVA results showed that differences in posttest critical thinking scores among the learning models were not statistically significant ($F(2, 202) = 2.229$, $p = .109$, $\eta^2 = .015$). According to Cohen's benchmarks, this effect size is considered small, indicating that variations in learning models explained only a limited proportion of the variance in students' learning outcomes. The discrepancy between the ANOVA and ANCOVA results suggests that some variation in posttest scores may be associated with initial differences among students; however, this influence was not statistically significant after adjustment.

Results of the analysis of students' critical thinking skills based on six Facione indicators

To provide an initial overview of students' critical thinking performance, descriptive

statistics of the average scores based on Facione's critical thinking indicators were analyzed across the three learning groups: PBL assisted Zep Quiz, Zep Quiz, and Traditional Lecture. The indicators consisted of interpretation, analysis, evaluation, inference, explanation, and self-regulation. Mean scores and standard deviations were used to describe students' performance and score variability across each indicator. The results are presented in Table 9.

Table 9. Average Critical Thinking Ability Scores Based on Facione's Indicators

Critical Thinking Indicator	PBL assisted Zep Quiz	Zep Quiz	Traditional Lecture
Interpretation (Identification)	2.35 ± 0.65	2.56 ± 0.57	2.32 ± 0.68
Analysis (Induction)	2.61 ± 0.63	2.78 ± 0.50	2.62 ± 0.63
Evaluation	1.79 ± 1.19	2.19 ± 1.01	2.00 ± 1.18
Inference (Deduction)	2.41 ± 0.71	2.53 ± 0.66	2.44 ± 0.69
Explanation (Credibility)	2.28 ± 0.74	2.36 ± 0.69	2.30 ± 0.72
Self-Regulation (Observation)	2.12 ± 0.77	2.21 ± 0.73	2.15 ± 0.75

Table 9 showed that the lowest scores were recorded in the evaluation and self-regulation indicators. This finding indicates that students' self-reflection skills in physics learning still require greater attention. The findings of Yaumie et al. (2025) support this result, confirming that the self-reflection aspect is rarely investigated in physics education research in Indonesia and tends to show relatively low improvement. Their study also highlights that reflective components are often overlooked in learning design. The limited integration of reflective activities in instructional design, combined with the lack of assessment instruments capable of authentically measuring metacognitive processes, appears to be a major factor contributing to the low improvement in self-reflection skills. Self-regulation skills, which involve monitoring and regulating students' own thinking processes, are also often not optimally developed in many test instruments. Most instruments tend to emphasize routine problem-solving rather than encouraging deeper critical reflection.

Discussion

These findings are consistent with literature on digital game-based learning and augmented reality (CAP), suggesting that improvements in critical thinking are not determined solely by the presence of game elements but by how these elements are pedagogically integrated with learning content (Rizki et al., 2024). Previous studies have shown that gamification design influences the depth of cognitive processing and learning engagement. The slightly higher average scores observed in several indicators suggest that interactive quiz activities may stimulate student participation and immediate cognitive responses. This interpretation is supported by Ahsan et al. (2026), who reported that digital quiz applications such as Quizizz Mobile significantly enhance students' motivation through interactive and gamified features promoting active engagement. Similar findings were also reported by Aulia et al. (2025), demonstrating that digital quiz media significantly increase students' active participation during learning activities.

However, despite their potential to improve engagement, quiz-based environments alone may not sufficiently support the development of higher-order critical thinking processes. The present findings are corresponded by Docktor et al. (2016), who argued that analytical and evaluative thinking in physics learning requires the integration of multiple representations, conceptual reasoning, and argumentative discussions. This perspective becomes particularly relevant considering that the evaluation and self-regulation indicators consistently showed lower average scores across learning conditions. The relatively weak performance in these dimensions suggests that students may still experience difficulties in reflective judgment and monitoring their own thinking processes. Similarly, Yaumie et al. 2025) found that self-reflection skills remain underdeveloped in Indonesian physics education because reflective components are frequently underrepresented in instructional practices and assessment systems.

The relatively small differences observed in analysis and inference indicators, along with the balanced performance in explanation and self-regulation, further suggest that critical

thinking development is a gradual process that depends on sustained exposure to deeper learning experiences rather than on short-term instructional variation. This interpretation supports previous findings emphasizing that critical thinking in science education develops through continuous problem-based learning involving intensive conceptual discussion and active reasoning processes (Al Kandari & Al Qattan, 2020; Tiruneh et al., 2017). Furthermore, the moderate improvement observed in Rotational Motion compared with Newton's Laws and Momentum topics suggests that conceptual complexity and representational demands may influence students' critical thinking performance. Physics concepts involving abstract visualization and mathematical reasoning likely require richer instructional support beyond interactive quiz activities alone.

Therefore, in the context of physics education, particularly for conceptually demanding topics, an integrated multimodal learning design appears necessary. Such an approach should combine problem-based learning, digital interactive activities, visual simulations, reflective exercises, and argumentative discussions. The integration of these components may create more opportunities for students to engage not only in surface-level interaction but also in deeper conceptual and metacognitive processing, thereby supporting more comprehensive development across all dimensions of critical thinking skills.

CONCLUSION

This study shows that PBL Assisted Zep Quiz, Zep Quiz, and Traditional lecture method all contribute to improving students' critical thinking skills in mechanics, as indicated by an increase in posttest scores and N-gain scores in the moderate to high categories. Descriptively, the greatest improvement occurred in the Momentum and Impulse material, while Rotational Motion showed the lowest achievement. However, the results of the repeated-measure ANOVA and ANCOVA tests, controlling for pretest scores, showed that the differences between learning models were not statistically significant with a small effect size. The use of a counterbalanced repeated measures design in this study provided stronger control over interclass variability, allowing for a more objective comparison of the effectiveness of learning models than conventional quasi-experimental designs. These results indicate that the three learning models are relatively comparable in their effectiveness in improving students' critical thinking skills.

These findings indicate that the improvement in critical thinking skills is not solely determined by variations in learning models but is also related to the characteristics of the material and learning design applied. Physics teachers can consider integrating interactive media such as PBL Assisted Zep Quiz approach to support student engagement and conceptual understanding. However, the generalization of these research results needs to be done carefully because the research was conducted in one school, limited to mechanical material, and took place in a relatively short intervention period. In addition, the reliability of the instrument on Rotational Motion material was in the moderate category, which had the potential to increase measurement error and reduce the sensitivity of the analysis in detecting actual differences. Further research with a broader sample coverage, longer treatment duration, and instruments with stronger internal consistency is needed to strengthen external validity and causal inference. Academically, this study contributes comparative controlled evidence on the relative effectiveness of several learning models in developing critical thinking skills in physics learning.

ACKNOWLEDGEMENT

The author would like to express his gratitude to the Head of the Study Program at UIN Syarif Hidayatullah Jakarta, the supervising lecturer who provided guidance, direction, and input during the research process and the writing of this article. The author also thanks his family, especially his mother, who has always provided support, encouragement, and prayers during the research. In addition, appreciation is expressed to Febriana Eka as a research partner in data collection who has contributed to the smooth running and success of this research.

REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for Teaching Students to Think Critically: A Meta-Analysis. *Review of Educational Research*, *85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Adhelacahya, K., Sukarmin, S., & Sarwanto, S. (2023). Impact of Problem-Based Learning Electronics Module Integrated with STEM on Students' Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 4869–4878. https://doi.org/10.29303/jppipa.v9i7.3931

Ahsan, F. A., Nasution, N. A., Dermawan, T. D., & Alda, M. (2026). Pengaruh Penggunaan Aplikasi Quizizz Mobile terhadap Motivasi Belajar Siswa di Era Pembelajaran Digital. *Jurnal Publikasi Teknik Informatika*, *5*(1), 192–200. https://doi.org/10.55606/jupti.v5i1.6413

AlKandari, A. M., & AlQattan, M. M. (2020). E-Task-Based Learning Approach to Enhancing 21st-Century Learning Outcomes. *International Journal of Instruction*, *13*(1), 551–566. https://doi.org/10.29333/iji.2020.13136a

Aulia, H., Rafi, M. A., Permatasari, R. P., Jaufillaili, J., & Apandi, A. (2025). How Does Digital Quiz Motivate Students in EFL Setting? *International Journal of Linguistics and Translation Studies*, *6*(3), 104–121. https://doi.org/10.36892/ijlts.v6i3.604

Cresswell, C., & Speelman, C. P. (2020). Does mathematics training lead to better logical thinking and reasoning? A cross-sectional assessment from students to professors. *PLOS ONE*, *15*(7), e0236153. https://doi.org/10.1371/journal.pone.0236153

Dewi, S., Dwikoranto, & Setiani, R. (2024). Analisis Efektivitas dan Respon Peserta Didik terhadap Penerapan Model Problem Based Learning (PBL) pada Materi Usaha dan Energi. *IPF: Inovasi Pendidikan Fisika*, *13*(2), 143–151. https://doi.org/10.26740/ipf.v13n2.p143-151

Docktor, J. L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., Mason, A., Ryan, Q. X., & Yang, J. (2016). Assessing student written problem solutions: A problem-solving rubric with application to introductory physics. *Physical Review Physics Education Research*, *12*(1), 10130. https://doi.org/10.1103/PhysRevPhysEducRes.12.010130

Goodwin, J. R. (2024). What's the Difference? A Comparison of Student-Centered Teaching Methods. *Education Sciences*, *14*(7), 736. https://doi.org/10.3390/educsci14070736

Jannah, M., Nasir, M., Siahaan, D. S., & Soewarno, S. (2022). Analysis of Students' Difficulties in Solving Physics Problems with Multiple Representation Using What's Another Way Method. *AL-ISHLAH: Jurnal Pendidikan*, *14*(2), 2479–2488. https://doi.org/10.35445/alishlah.v14i2.1008

Kotsis, K. T. (2025). From Theory to Practice: Exploring Students' Everyday Use of Physics Knowledge. *European Journal of Contemporary Education and E-Learning*, *3*(5), 59–76. https://doi.org/10.59324/ejceel.2025.3(5).05

Moeis, M. R., & Maulana, M. F. (2021). Improving plastic degradation by increasing the thermostability of a whole cell biocatalyst with LC-cutinase activity. *Journal of Physics: Conference Series*, *1764*(1), 12029. https://doi.org/10.1088/1742-6596/1764/1/012029

OECD. (2019). PISA 2018 Results (Volume I): What Students Know and Can Do. In *PISA*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd Edition). John Wiley & Sons, Inc.

Rizki, I. A., Suprapto, N., Saphira, H. V, Alfarizy, Y., Ramadani, R., Saputri, A. D., & Suryani, D. (2024). Cooperative model, digital game, and augmented reality-based learning to enhance students critical thinking skills and learning motivation. *Journal of Pedagogical Research*, 1. https://doi.org/10.33902/JPR.202423825

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM Approach* (1st Edition). SAGE Publications.

Susanto, F. S., & Airlanda, G. S. (2023). Efektivitas Model Problem Based Learning dan Problem Solving untuk Meningkatkan Kemampuan Berpikir Kritis Siswa dalam Pembelajaran IPAS. *Jurnal Basicedu*, *7*(6), 3646–3653. https://doi.org/10.31004/basicedu.v7i6.6353

Takaoğlu, Z. B. (2024). High School Students' Multiple Representation Translation Skills on One-Dimensional Motion: A Cross-Grade Study. *Journal of Science Learning*, *7*(1), 47–55.

https://doi.org/10.17509/jsl.v7i1.61099

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, *15*(4), 663–682. https://doi.org/10.1007/s10763-016-9723-0

Valentine, K. D., Kopcha, T. J., & Vagle, M. D. (2018). Phenomenological Methodologies in the Field of Educational Communications and Technology. *TechTrends*, *62*(5), 462–472. https://doi.org/10.1007/s11528-018-0317-2

Yaumie, S. N., Jatmiko, B., Satriawan, M., Supardi, Z. A. I., & Arivi, M. A. (2025). The Role of Guided Inquiry to Improve Critical Thinking Skills in Physics Education: A Systematic Review of Indonesian Studies. *Jurnal Pendidikan MIPA*, *26*(3), 1893–1908. https://doi.org/10.23960/jpmipa.v26i3.pp1893-1908