Research Article https://doi.org/10.12973/eu-jer.13.3.1391

European Journal of Educational Research
Volume 13, Issue 3, 1391 - 1409.
ISSN: 2165-8714
https://www.eu-jer.com/

Developing Students' Critical Thinking: Examining the Influence of Learning Management Approaches Through Meta-Analysis and Propensity Score Matching

Supansa Surin
Chiang Mai University, THAILAND

Suntonrapot Damrongpanit*
Chiang Mai University, THAILAND

Received: December 18, 2023 ▪ Revised: February 21, 2024 ▪ Accepted: March 5, 2024

Abstract: Critical thinking is a skill that enables individuals to keep pace with changes and enhances crucial competencies for contemporary competitiveness. Many researchers have studied learning management approaches to develop students' critical thinking, resulting in a substantial body of knowledge but lacking clear systematic summaries. The researchers aimed to (a) examine the effect sizes and research characteristics influencing students' critical thinking, and (b) compare the effect sizes of learning management approaches after adjusting with propensity score matching from 108 graduate research published between 2002 and 2021. Data were collected using research characteristics recording forms and research quality assessment questionnaires. Effect sizes were calculated using Glass's method and analyzed through random effect, fixed effect, and regression meta-analysis. Findings revealed that (a) research on developing learning management approaches influences students' critical thinking at a high level ($\bar{d}$ = 1.669), with nine research characteristics, including the field of publication, courses, total duration, teacher learning process, learning media, measurement and evaluation, research design, research statistics, and research quality, statistically significantly influencing students' critical thinking, and (b) after adjustment, inquiry-based learning significantly influences students' critical thinking. Recommendations for developing students' critical thinking include learning activities that encourage problem exploration, expanding thinking through collaborative analysis, and applying diverse media and activity sheets tailored to context suitability.

Keywords: *Critical thinking, learning management approaches, meta-analysis, propensity score matching, research synthesis.*

To cite this article: Surin, S., & Damrongpanit, S. (2024). Developing students' critical thinking: Examining the influence of learning management approaches through meta-analysis and propensity score matching. *European Journal of Educational Research*, *13*(3), 1391-1409. https://doi.org/10.12973/eu-jer.13.3.1391

Introduction

In the 21st century, individuals require critical thinking, problem-solving, creativity, innovation, responsible citizenship, and collaborative skills to meet global job market demands (Kennedy & Sundberg, 2020). Scholars focus on developing students' critical thinking through learning management approaches, curriculum, and assessment strategies (Elder & Paul, 2021; Ennis, 1993, 2015, 2018; Karampelas, 2023; Paul & Elder, 2020), aiming for lasting critical thinking and improved lifelong learning (Green, 2015). Proficient critical thinkers can effectively search, analyze, and decide on information, employing reasoned thinking and logical analysis for open-mindedness, understanding, and problem-solving (Kocak et al., 2021; McPeck, 2016). They can also synthesize information, communicate, and collaborate well, contributing to societal development (Paul & Elder, 2005).

Educational institutions worldwide prioritize critical thinking in 21st century learning because it is a skill that helps individuals adapt to and access rapidly changing information. Developing this competency leads to success in work, education, and daily life in a world full of challenges and uncertainties (Dwyer et al., 2014). Thailand, like many other countries, has been actively focusing on developing students' critical thinking through various learning management approaches for over four decades. Initial surveys reveal 839 research papers published on this topic since 1976, with a focus on graduate education in science and technology teaching (39.82%) and secondary education (45.37%). The research method is the randomized control group pretest-posttest design (54.63%). Six learning management approaches have been identified (Saylor et al., 1981), including collaborative learning (5.55%), constructivism (7.41%), inquiry-based learning (45.37%), learning through techniques (22.22%), activity kits and media (7.41%), and problem-

* **Corresponding author:**
Suntonrapot Damrongpanit, Faculty of Education, Chiang Mai University, Thailand. ✉ suntonrapot.d@cmu.ac.th

based learning (12.04%). A blended teaching media approach is predominantly used (50.93%), and good-quality research (69.45%). Researchers made three observations. Firstly, differing research outcomes confuse educators. Secondly, despite positive influences on critical thinking, low scores on the Ordinary National Educational Test (O-NET) suggest misalignment with the national assessment framework, thus hindering effective student development. Lastly, research methodologies lack completeness, diminishing the credibility of findings. The findings lack systematic organization and clear conclusions. Therefore, meta-analysis is an appropriate method for deriving conclusions from those research findings, to be applied to enhance the efficiency of developing students' critical thinking.

Meta-analysis is a statistical technique that systematically synthesizes quantitative findings from research studies addressing the same topic, typically involving two or more studies. It offers insight into the impact of the studied topic, known as "effect sizes," facilitating analysis and comparison based on research characteristics. This process leads to significant, comprehensive, and reliable conclusions (Borenstein et al., 2021; Borenstein & Higgins, 2013; Cooper et al., 2019). In 2023, a review of 925 meta-analysis studies revealed its widespread use in various fields, with 78.05% in medical and health, 15.03% in humanities and social sciences, 4.32% in the environment, and 2.60% in economics. Despite the specific insights in each domain used for treatment, development, policy, or forecasting current and future changes, the influence of confounding variables persists, as found in Itsarangkul Na Ayutthaya and Damrongpanit (2022a) and Niu et al. (2013). Research characteristics such as learning processes, level, research fields, total duration, and research quality impact the effect sizes. Hence, conclusions derived from these analyses are not completely applicable until the influence of confounding variables in meta-analysis is eliminated. Propensity score matching is a suitable method for this purpose because it reduces the influence of confounding variables on effect sizes (Egger et al., 1997).

Propensity score matching reduces the variability of confounding variables by ensuring high similarity between sample groups (Badhiwala et al., 2021; Benedetto et al., 2018; Rosenbaum & Rubin, 1983). It adjusts experimental outcomes based on research characteristics to facilitate accurate treatment effect comparisons (Haukoos & Lewis, 2015; Rosenbaum, 2020), reducing publication bias and enabling objective comparisons (Bai, 2011; Heinrich et al., 2010; Staffa & Zurakowski, 2018). In the research by Itsarangkul Na Ayutthaya and Damrongpanit (2022b), propensity score matching was utilized for meta-analysis, clarifying that teaching methods emphasizing environment-based integrated knowledge construction through technology impact students’ creative thinking. This method enhances the validity of conclusions by eliminating the influence of confounding variables (Rubin, 1997).

The researchers aim to study 18 research characteristics to determine the variables that influence the effect sizes, while also controlling for confounding factors. Additionally, they aim to identify which of the six learning management approaches has the most significant impact on students' critical thinking. Therefore, their objectives are twofold: (a) To examine the effect sizes and research characteristics that influence students' critical thinking, and (b) To compare the effect sizes of learning management approaches that develop students' critical thinking after adjusting with propensity score matching.

Literature Review

Critical Thinking

Critical thinking involves a profound reasoning process leading to evidence analysis, problem-solving, and decision-making (Elder & Paul, 2021; Ennis, 2015; Kocak et al., 2021), fostering diverse perspectives and deep understanding (Paul & Elder, 2005). Critical thinkers systematically evaluate data, address problems, and communicate effectively (Ennis, 2015; Sternberg, 1986). Conversely, individuals lacking critical thinking rely on emotions, lack data awareness, and limit their viewpoints (Facione, 2023). Developing critical thinking involves personal traits and environmental factors (Indah & Kusuma, 2016; Thongnuypram & Sopheerak, 2013), crucial for effective decision-making in evolving 21st century situations.

Meta-Analysis

Meta-analysis systematically synthesizes research findings by aggregating quantitative results from studies on the same topic, providing clear and robust conclusions through the calculation of effect sizes (Borenstein et al., 2021; Card, 2012; Cooper et al., 2019). By augmenting the sample size, meta-analysis enhances statistical power, overcoming limitations inherent in small-scale studies and effectively resolving conflicting findings (Y. H. Lee, 2019; Stone & Rosopa, 2017). Moreover, meta-analysis scrutinizes variability differences based on various research characteristics, offering valuable insights into study outcomes, and generating potential hypotheses for future investigations. Consequently, meta-analysis transcends the constraints associated with narrative reviews, particularly in terms of theoretical analysis.

The factors in meta-analysis include poor study quality, heterogeneity, publication bias, small sample sizes, and methodological differences (Allen, 2020; Esterhuizen & Thabane, 2016; Y. H. Lee, 2019; Stone & Rosopa, 2017). In education, meta-analysis assesses learning management, environmental factors, testing methods, and impact on student learning (Adesope et al., 2017; Donoghue & Hattie, 2021; Lambert & Guillette, 2021).

Meta-Analysis of Learning Management Approaches for Developing Students' Critical Thinking

For studies on students' critical thinking, various research methodologies have been employed, including systematic review and meta-analysis. Interesting meta-analysis research by Abrami et al. (2008, 2015), Niu et al. (2013), and Xu et al. (2023) has been found. These studies have examined instructional practices aimed at developing students' critical thinking, alongside associated factors. Factors such as duration, student level, subject area, measuring tool, group size, and learning scaffold have been identified and analyzed for their impact on instructional effectiveness and intervention outcomes. This analysis has helped elucidate the variability in effect sizes. Key conclusions drawn from these studies inform the design of learning approaches to develop students' critical thinking, emphasizing activities such as problem-solving, questioning, discussion, collaborative learning, and engagement in finding solutions. These activities aim to develop critical thinking processes, improve reading and writing skills, and promote continuous development (Hitchcock, 2017; Lai, 2011; O'Reilly et al., 2022). However, due to the absence of a definitive learning management approach, researchers have developed diverse approaches to developing students' critical thinking. These approaches consider various factors, including curricula, teacher learning processes, use of educational media, and assessment methods. Researchers have categorized six such approaches (Saylor et al., 1981) as follows: (a) Collaborative Learning focuses on students working together in heterogeneous groups, where tasks and roles are assigned to facilitate knowledge exchange, shared responsibility, teamwork skills development, and enhancing knowledge and understanding. Popular techniques include STAD, TGT, and TAI (Jacobs & Renandya, 2019; Slavin, 2014), (b) Constructivism involves students exploring and connecting knowledge through stimulating activities such as thinking, reasoning, questioning, and problem-solving, leading to discussions, exchange of ideas, and self-generated knowledge (Amineh & Asl, 2015; Rannikmäe et al., 2020), (c) Learning through techniques promotes specific sequential thinking processes, blending questioning, comprehension, analysis, problem-solving, reasoning, and step-by-step summarization to enhance systematic thinking, decision-making, and applicability. Popular techniques include Six Thinking Hats, KWL, and KWL-Plus (Alsaleh, 2020; Saiz & Rivas, 2016), (d) Activity kits and media align learning management with learners' needs by incorporating multimedia such as images, videos, audio, and animations into learning activities to motivate, create interest, and enhance learning efficiency. This enables students to understand better the learning context (Akinbadewa & Sofowora, 2020; Lampert & Graziani, 2009), (e) Inquiry-based learning encourages students to explore and connect knowledge through deep exploration, questioning, and experimentation, enabling explanation, knowledge exchange, critical thinking, knowledge transfer, presentation, and problem-solving. Learning processes include 3E, 5E, 7E, and 9E (Eisenkraft, 2003; Nicol et al., 2020; Varoglu et al., 2023), and (f) Problem-based learning focuses on problem-solving as a core learning activity, involving problem identification, data exploration, planning, problem-solving, summarizing knowledge, and evaluation. This stimulates students to think, face real problem situations, promote collaborative learning, and develop effective understanding and applicable knowledge (Moallem et al., 2019; Moust et al., 2021).

Propensity Score Matching

Propensity score matching is a statistical technique that mimics a Randomized Controlled Trial (RCT) and reduces bias in non-randomized studies (Kane et al., 2020; Morgan, 2018), thereby enhancing the reliability of experimental results by aligning outcomes of experimental and control groups (Badhiwala et al., 2021; Benedetto et al., 2018). This method manages multiple confounding variables concurrently using propensity scores derived from logistic regression analysis for matching, handling extensive datasets, and refining experimental accuracy (Haukoos & Lewis, 2015; W.-S. Lee, 2013). However, it cannot regulate unmeasured variables and involves intricate analytical procedures, requiring caution in variable selection and thorough matching to impact resulting outcomes (Reiffel, 2020; Sainani, 2012).

Propensity score matching is widely used in medical research to assess the impacts of medications and patient treatments, adjusting for uncontrolled variables, and enabling comparisons. Its application extends to the field of education, where it complements meta-analysis to address bias from confounding variables affecting effect sizes. For example, in the research conducted by Itsarangkul Na Ayutthaya and Damrongpanit (2022b), it was found that there was a difference in research outcomes before and after using propensity score matching, with the post-research revealing reduced publication bias, as observed from the funnel plot. This leads to conclusions drawn from the actual implementation of the learning management approach. Additionally, the research conducted by Ripley (2015), was utilized in experimental studies to mitigate the influence of confounding variables among treatment groups, resulting in more accurate research outcomes.

Methodology

Research Design

This study is a meta-analysis of experimental research on learning management approaches aimed at developing students' critical thinking. Effect sizes were calculated using Glass's method (1976) based on experimental research findings. The analysis covers 18 variables related to research characteristics. These variables include two fundamental research variables, ten content substance-related variables, and six research methodology variables. They were utilized to compare effect sizes based on research characteristics.

Sample Selection

The researchers followed a four-step sample selection process: First, Identification: searching for research papers from the TDC database using the keywords "Critical Thinking" to compile all research publications. A total of 839 articles were identified. Second, Screening: selecting full-text research articles that were basic educational experimental studies focusing on learning management approaches. After screening out redundant research and non-experimental studies, 307 articles were excluded, leaving 532 articles meeting the criterion. Third, Eligibility: selecting experimental research studies with both experimental and control groups, which reported sufficient statistics to calculate effect sizes (d). Screening studies lacking experimental and control groups resulted in the exclusion of 401 articles, with 131 articles meeting the eligibility criteria. Lastly, Inclusion: selecting research studies aligned with the Basic Education Core Curriculum, from 2002 to 2021. Screening out research studies produced outside this curriculum led to the exclusion of 23 articles, leaving a total of 108 articles meeting the inclusion criteria. The process is depicted in Figure 1.

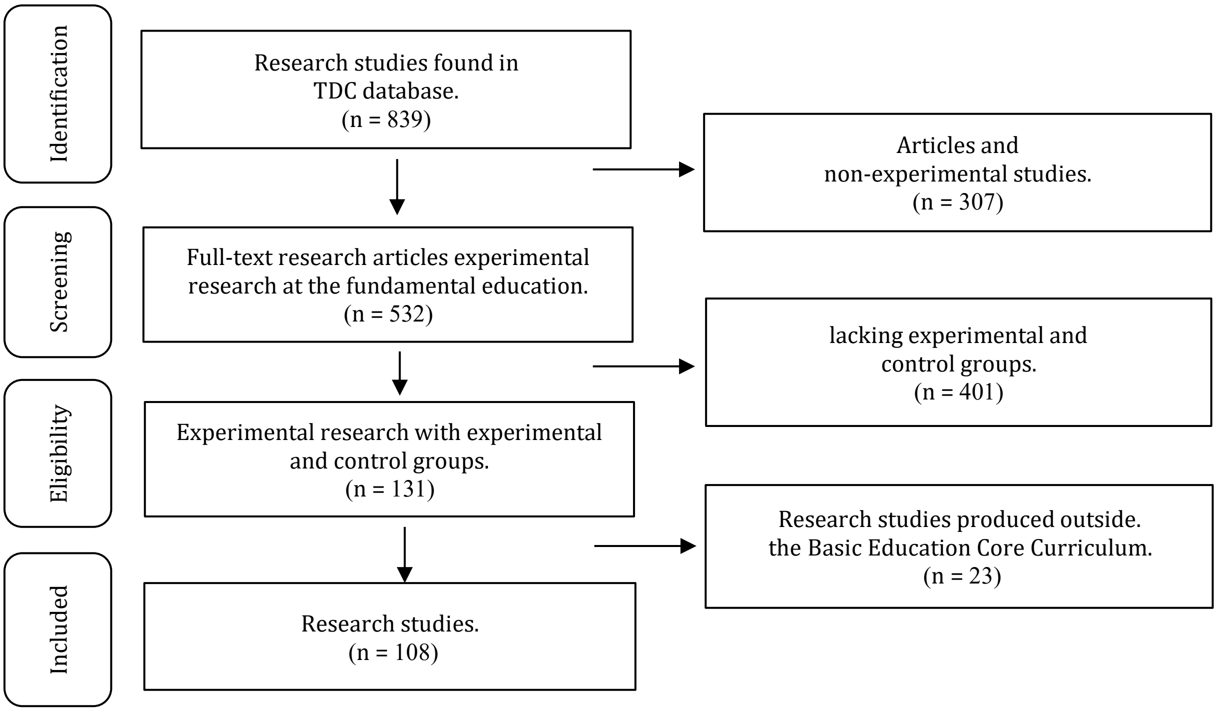


Figure 1. Flow Diagram of Selecting a Sample

Developing and Validating Research Instruments

The research utilized two tools: the Research Characteristics Recording Form and the Research Quality Assessment Questionnaire. The development and validation of these tools are outlined as follows:

The Research Characteristics Recording Form collected data across three dimensions: basic research information, the content substance of the research, and research methodology. Kappa statistics were employed to evaluate inter-rater reliability, with a reliability value of .915, and intra-rater reliability, with a reliability value of .972, indicating a high level of confidence (Czodrowski, 2014).

The Research Quality Assessment Questionnaire assessed research across seven dimensions: Title, Background, Relevant research documents, Research Methodology, Research findings, Report writing, and Research benefits. The Rubric Scoring system consisted of 5 levels, ranging from 0 indicating low research quality to 4 indicating high research quality, totaling 25 items. Kappa statistics were employed to evaluate inter-rater reliability, with a reliability value of .813, indicating relatively good reliability, and intra-rater reliability, with a reliability value of .920, indicating a high level of reliability (Czodrowski, 2014).

Analyzing of Data

In data analysis, understanding key statistical terms is essential: (a) Funnel plot, visualizes effect sizes in studies., (b) Kendall's Tau, measures dependence strength between variables., (c) Egger's Test, identifies bias from small effect sizes., (d) Tau squared (τ^2) represents the variance of effect size. If τ^2 values = 0, no variability., (e) I squared (I^2), indicates heterogeneity level between studies. If I^2 values = 25%, 50%, and 75% as low, moderate, and high variability, (f) Q statistic, evaluates heterogeneity among effect sizes., (g) Chi-square test, determines the significance of variability among effect sizes., (h) z statistic, evaluating the significance of the overall effect size and assessing the statistical significance of independent variables that are correlated with effect size. (i) Independent sample t-test: Tests difference between means of two independent groups., (j) Two-way ANOVA, analyzes the influence of two categorical independent variables on a continuous dependent variable., and (k) Logistic regression, the relationship between the binary dependent variable and independent variables. (Borenstein et al., 2021; StataCorp, 2023)

The researchers used JASP version 0.17.2 software for data analysis, which was divided into two parts of analysis as follows:

Part 1: Meta-analysis and regression meta-analysis were conducted to address research objective 1 by analyzing the effect sizes and research characteristics. Each variable was subdivided into subgroups (Borenstein & Higgins, 2013) and then transformed into dummy variables. Subsequently, a four-step analysis was employed to determine whether to use a random effect or a fixed effect:

Step 1: A check for publication bias was performed, which involved: (a) The funnel plot resembles an inverted funnel. The x-axis represents the effect sizes of each research, while the y-axis represents the standard error, which indicates sample sizes. If studies with large sample sizes are clustered around the middle of the funnel plot, while those with small sample sizes are scattered on both sides of the middle (Harrison, 2011; Sedgwick & Marston, 2015), and (b) Kendall's Tau and Egger's Test to assess statistical significance, indicating publication bias in studies (Rothstein, 2008; Thornton & Lee, 2000; Vevea et al., 2019).

Step 2: Effect sizes were analyzed by computing them according to Glass's method (1976) and evaluating their influence level. Criteria were established where d = 0.20 – 0.50 indicated low influence, d = 0.50 – 0.80 indicated moderate influence, and d > 0.80 indicated high influence. The decision between the random effect and the fixed effect was based on τ^2 and I^2 values, with high values indicating variability among studies favoring the use of the random effect (Borenstein et al., 2021; Rücker et al., 2008).

Step 3: The meta-analysis method was selected by considering: (a) Omnibus test of model coefficients to assess whether the mean effect size is zero, A significant Q statistic (p-value ≤ .05) indicated the mean effect size is different from zero., (b) Test of residual heterogeneity to determine if the residual from estimating effect sizes is zero. A significant Q statistic (p-value ≤ .05) indicated heterogeneity, favoring the use of random effect. Conversely, a non-significant Q statistic suggested homogeneity, favoring the use of a fixed effect, (c) The chi-square test was also employed to check homogeneity or heterogeneity, with a higher Q statistic than the chi-square value indicating rejection of the null hypothesis and thus heterogeneous effect sizes (Berkhout et al., 2024; Borenstein et al., 2021; Card, 2012; Higgins & Thompson, 2002).

Step 4: Regression meta-analysis utilized the z statistic to examine whether effect sizes differed from zero. The first variable served as the intercept for comparison. Statistically significant variables (p-value ≤ .05) indicated an influence on the mean effect size, while non-significant variables indicated no influence (p-value ≤ .05) (Van Houwelingen et al., 2002).

Part 2: Propensity score matching was undertaken to address research objective 2, involving six steps as follows:

Step 1: Dividing effect size groups based on the mean effect size.

Step 2: Employing independent samples t-test to assess research characteristics affecting effect size groups.

Step 3: Analyzing confounding variables through logistic regression, resulting in propensity score.

Step 4: Dividing propensity score groups.

Step 5: Conducting two-way ANOVA to explore variables still impacting effect size groups. If variables affecting the effect sizes are identified, those variables will be entered into logistic regression following step 3 is repeated until no variables affecting the effect sizes remain.

Step 6: Comparing learning management approaches aimed at developing students' critical thinking using propensity score (Bai, 2011; Harris & Horst, 2016; Staffa & Zurakowski, 2018), then employing meta-analysis and regression meta-analysis based on part 1.

Findings/Results

Results of Examining the Effect Sizes and Research Characteristics Influencing Students' Critical Thinking

The results of examining the publication bias of the samples used in this study, as assessed through the funnel plot in Figure 2, indicate that the effect sizes of the research works are predominantly from studies with large sample sizes, as evidenced by the clustering of data points towards the center of the triangular shape. A positive influence is observed, and data points appear predominantly in the upper portion, with some dispersion outside the triangular boundary. Additionally, Kendall's Tua value of 0.522 (p-value < .05) and Egger's test value of 8.964 (p-value < .05) indicate that the effect sizes are influenced by significantly different research characteristics, leading to publication bias. Consequently, these findings cannot be used to draw conclusive research conclusions.

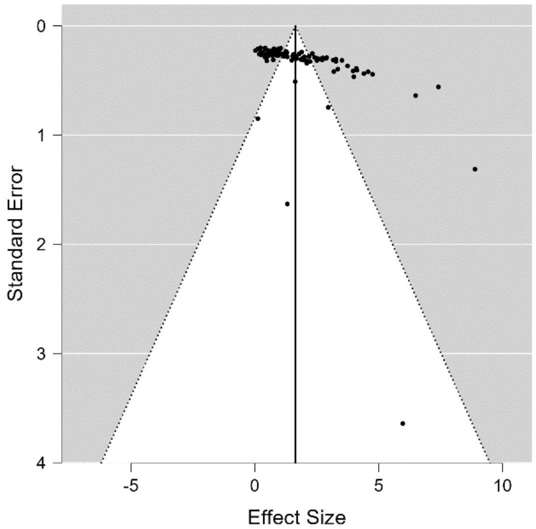


Figure 2. Funnel Plot of the Effect Sizes

The results of calculating the effect sizes of students' critical thinking, based on 108 studies, ranged from 0.024 to 8.889. Under the random effects model, it was found that $\bar{d}$ = 1.669, *SE* = 0.132, Upper 95% CI = 1.928, and Lower 95% CI = 1.409. Conversely, the fixed effects found that $\bar{d}$ = 1.291, *SE* = 0.027, Upper 95% CI = 1.343, and Lower 95% CI = 1.239. Additionally, the effect sizes of the research exhibited high variability (τ^2 = 1.029, I^2 = 93.024%), thus the random effects model was used to estimate the effect sizes.

As a result, it was found that the mean effect size was 1.669, indicating a significant influence on critical thinking at a high level (0.80 < *d*). When considering the *z*-test results, the *z*-value was equal to 12.610 (*p*-value < .05), indicating that the mean effect size significantly differed from zero and exhibited a positive influence trend, as detailed in Table 1.

The heterogeneity test resulted in a *Q* value of 1550.562 (*p*-value < .05), indicating that there is a non-zero residual from estimating the effect sizes. The chi-square values obtained from the table with *df* = 107 were $\chi 2$ = 132.144 and *df* = 109 were $\chi 2$ = 134.369, both at a significance level of .05. It's evident that the statistical *Q* value is higher than the chi-square value, thus rejecting the null hypothesis. Additionally, with τ^2 = 1.029 and I^2 = 93.024%, which are high, it indicates high variability in the effect size, signifying significant differences among each instance. Therefore, testing the effect sizes using random effects will ensure unbiased estimation results, as detailed in Table 1.

Table 1. Results of Random Effect, Fixed Effect, and Heterogeneity Analyses on the Effect Sizes.

Model	Effects Size			95% CI		Absence Hypothesis		Heterogeneity			
	k	$\bar{d}$	***SE***	**Lower**	**Upper**	***z*-value**	***p*-value**	***Q***	***df***	τ^2	I^2
Random Effects	108	1.669	0.132	1.409	1.928	12.610***	< .001	1550.562***	107	1.029	93.024%
Fixed Effects	108	1.291	0.027	1.239	1.343	48.337***	< .001				

Upon examining the effect sizes from the funnel plot (Figure 2) and estimating the effect sizes using random effects (Table 1), it was found that the effect sizes were influenced by research characteristics, rendering the data insufficient for conclusive summaries. Therefore, the influence of individual research characteristics was studied, as detailed in Table 2.

Table 2. Research Characteristics Influencing Students' Critical Thinking

Research characteristics	*k*	$\bar{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b	τ^2	I^2
Learning Management Approaches									
Collaborative Learning	6	3.021	1.305	[0.46,5.58]	5.135***	4.154	1536.587***	1.019	92.915%
Constructivism	7	1.785	0.603	[0.60,2.97]	-0.979				
Learning through Techniques	25	1.566	0.281	[1.02,2.12]	-1.909				
Activity Kits and Media	8	1.619	0.462	[0.71,2.52]	-1.283				
Inquiry-Based Learning	49	1.769	0.228	[1.32,2.22]	-1.310				
Problem-Based Learning	13	1.605	0.249	[1.12,2.09]	-1.369				
Overall	108	1.677	0.134	[1.42,1.94]	12.558***				

Table 2. Continued

Research characteristics	*k*	$\bar{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b	τ^2	I^2
Year of Publication									
2002 – 2006	21	1.640	0.342	[0.97,2.31]	6.745***	1.253	1541.177***	1.022	92.944%
2007 – 2011	51	1.686	0.207	[1.28,2.09]	0.077				
2012 – 2016	26	1.757	0.252	[1.26,2.25]	0.499				
2017 – 2021	10	2.419	0.842	[0.77,4.07]	0.957				
Overall	108	1.722	0.143	[1.44,2.00]	12.059***				
Research Field									
Curriculum and Instruction	29	1.126	0.157	[0.82,1.43]	5.836***	12.330**	1503.917***	0.996	92.776%
Research, Measurement, Evaluation, and Educational Psychology	17	2.322	0.544	[1.26,3.39]	2.079*				
Learning Management, Elementary and Secondary Education	19	1.544	0.260	[1.03,2.05]	1.249				
Science and Technology Education Teaching	43	2.066	0.261	[1.55,2.58]	3.402***				
Overall	108	1.630	0.221	[1.20,2.06]	7.361***				
Courses									
Thai Language	7	0.963	0.223	[0.53,1.40]	2.449*	5.892	1506.013***	0.997	92.762%
Mathematics	12	2.236	0.678	[0.91,3.56]	1.788				
Science	54	1.839	0.211	[1.43,2.25]	1.981*				
Social Studies, Religion, and Culture	15	1.514	0.272	[0.98,2.05]	1.148				
Career and Technology	8	1.241	0.570	[0.12.2.36]	0.520				
Health and Physical, Arts, Foreign Languages, Student Development Activities	12	2.061	0.498	[1.08,3.04]	1.413				
Overall	108	1.536	0.182	[1.18,1.89]	8.438***				
Level									
Primary School	23	1.709	0.392	[0.94,2.48]	6.624***	3.386	1.519.656***	1.007	92.862%
Junior High School	49	1.626	0.195	[1.24,2.01]	0.186				
Senior High School	36	1.981	0.278	[1.44,2.53]	1.517				
Overall	108	1.738	0.148	[1.45,2.03]	11.757***				
Duration per Plan									
1 hour	19	1.886	0.450	[1.00,2.77]	6.648***	8.241	1456.120***	0.962	92.519%
2 hours	37	1.535	0.262	[1.02,2.05]	-0.777				
3 hours	38	2.050	0.241	[1.58,2.52]	1.318				
4 hours	8	1.272	0.512	[0.27,2.28]	-0.883				
More than 4 hours	6	1.595	0.361	[0.89,2.30]	-0.089				
Overall	108	1.744	0.144	[1.46,2.03]	12.107***				
Total Duration									
0– 10 hours	44	1.913	0.230	[1.46,2.37]	11.602***	5.826	1526.481***	1.012	92.878%
11 - 15 hours	26	1.800	0.374	[1.07,2.53]	-1.478				
16 - 20 hours	30	1.718	0.281	[1.17,2.27]	-0.730				
More than 20 hours	8	0.968	0.179	[0.62,1.32]	-2.192*				
Overall	108	1.556	0.233	[1.12,1.99]	6.983***				
Learning Objectives									
Cognitive / Psychomotor Domain	11	1.947	0.777	[0.42,3.47]	4.654***	0.244	1538.271***	1.020	92.950%
Cognitive and Psychomotor Domain	64	1.796	0.196	[1.41,2.18]	0.452				
Cognitive, Psychomotor, and Affective Domain	33	1.634	0.208	[1.23,2.04]	0.246				
Overall	108	1.727	0.140	[1.45,2.00]	12.311***				

Table 2. Continued

Research characteristics	k	$\bar{d}$	SE	95%CI	z-value	Q_a	Q_b	τ^2	I^2
Teacher Learning Process									
Introduction, Teach, Practice, and Summarize	19	1.300	0.290	[0.73,1.87]	5.360***	12.469	1460.323***	0.953	92.437%
Generate Interest, Teach, Practice, Assess, and Reward	8	2.736	0.994	[0.79,4.68]	2.059*				
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	17	2.040	0.409	[1.24,2.84]	2.051*				
Encounter problems, stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	15	1.855	0.466	[0.94,2.77]	0.731				
Engage, Explore, Explain, Elaborate, and Evaluate	29	1.350	0.210	[0.94,1.76]	0.170				
Generate Interest, Explore, Plan, Practice, Summary and presentation, and Evaluate	11	1.910	0.247	[1.43,2.39]	1.523				
Explore, Define the Problem, Analyze, Presentation, Application, and Publication	9	2.335	0.638	[1.08,3.59]	2.212*				
Overall	108	1.661	0.138	[1.39,1.93]	12.055***				
Student Learning Process									
No Grouping and Discussion	7	1.390	0.400	[0.61,2.17]	3.342***	8.850*	1444.797***	0.954	92.462%
Grouping and Discussion	10	2.279	0.617	[1.07,3.49]	1.653				
Grouping, Discussion, and Presentation	79	1.790	0.160	[1.48,2.10]	0.998				
Grouping, Discussion, Presentation and Reinforcement	12	1.359	0.690	[0.01,2.71]	-0.726				
Overall	108	1.748	0.141	[1.47,2.02]	12.362***				
Learning Media									
Song/Folklore/Video (Use 1 type)	10	0.836	0.090	[0.66,1.01]	2.582*	13.407**	1478.980***	0.978	92.655%
Worksheet/Quiz (Use 1 type)	18	2.384	0.505	[1.39,3.37]	3.012**				
Blend 2 types of learning media	55	1.587	0.200	[1.20,1.98]	1.854				
Integrate learning from more than 2 types of media	25	2.069	0.291	[1.50,2.64]	3.050**				
Overall	108	1.560	0.245	[1.08,2.04]	6.361***				
Measurement and Evaluation									
Evaluating Behavior	21	1.167	0.258	[0.66,1.67]	5.088***	6.230	1485.770***	0.983	92.686%
Posttest, Checking Assignments.	54	1.957	0.214	[1.54,2.38]	2.391*				
Posttest, Checking Assignments, Evaluating Work	18	1.906	0.488	[0.95,2.86]	1.888				
Posttest, Checking Assignments, Evaluating Behavior	15	1.718	0.351	[1.03,2.41]	1.634				
Overall	108	1.663	0.178	[1.31,2.01]	9.319***				
Research Sources									
Population	38	1.437	0.186	[1.07,1.80]	8.201***	3.223	1.532.571***	1.016	92.930%
Sample	70	1.938	0.210	[1.53,2.35]	1.795				
Overall	108	1.668	0.176	[1.32,2.01]	9.472***				

Table 2. Continued

Research characteristics	k	$\bar{d}$	SE	95%CI	z-value	Q_a	Q_b	τ^2	I^2
Sampling									
Simple random sampling	18	1.813	0.368	[1.09,2.53]	6.822***	2.497	1536.545***	1.019	92.943%
Cluster/multi-stage sampling	61	1.848	0.198	[1.46,2.24]	0.101				
Stratified sampling	29	1.549	0.317	[0.93,2.17]	-1.068				
Overall	108	1.773	0.153	[1.47,2.07]	11.602***				
Research Design									
Randomized control group Pretest-Posttest Design	59	1.933	0.195	[1.55,2.32]	13.287***	4.349	1499.438***	0.993	92.759%
Nonrandomized control group Pretest-Posttest Design	30	1.545	0.344	[0.87,2.22]	-2.005				
Non-Equivalent control group Pretest-Posttest Design	19	1.572	0.304	[0.98,2.17]	-1.128				
Overall	108	1.775	0.148	[1.48,2.07]	11.984***				
Tool Quality Assessment									
Validity	8	2.218	0.527	[1.19,3.25]	5.896***	6.903*	1486.294***	0.983	92.705%
Validity and Reliability	25	2.198	0.349	[1.51,2.88]	-0.405				
Validity, Reliability, Discriminant and Difficulty	75	1.568	0.175	[1.23,1.91]	-1.808				
Overall	108	1.769	0.170	[1.44,2.10]	10.422***				
Research Statistics									
ANOVA, MANOVA	16	1.617	0.520	[0.60,2.64]	4.947***	5.195	1477.993***	0.977	92.645%
ANCOVA	23	1.721	0.373	[0.99,2.45]	1.005				
MANCOVA	19	2.117	0.278	[1.57,2.66]	2.123*				
t–test Independent Sample	50	1.692	0.205	[1.29,2.09]	0.770				
Overall	108	1.806	0.145	[1.52,2.09]	12.463***				
Research Quality									
Moderate level	14	2.895	0.417	[2.08,3.71]	9.960***	20.020***	1465.229***	0.968	92.606%
Good level	75	1.636	0.190	[1.26,2.01]	-4.343***				
Excellent level	19	1.422	0.211	[1.01,1.84]	-3.838***				
Overall	108	1.827	0.255	[1.33,2.33]	7.166***				

Note: k = Sample size, d = Mean effect size, SE = Standard error, CI = Confidence Interval, Q_a = Omnibus test of Model Coefficients, Q_b = Test of Residual Heterogeneity, *** $p < .001$, ** $p < .01$, * $p < .05$, Year of publication: 2002 – 2006 = Teacher-Center, 2007–2016 = Child-Center, 2017 – 2021= Active learning

The results of examining the effect sizes of 108 research studies showed a mean effect size of 1.669 (Table 1). When examining Table 2, it is observed that each variable in research characteristics has τ^2 values and I^2 values at a high level ($0 < \tau^2$, $75 < I^2$), indicating significant variability in the effect sizes of each study. This variability is influenced by different research characteristics. The research findings suggest nine research characteristics statistically significant at the .05 level that influenced students' critical thinking. These variables included: (a) research field, (b) courses, (c) total duration, (d) teacher learning process, (e) learning media, (f) measurement and evaluation, (g) research design, (h) research statistics, and (i) research quality. Consequently, these findings could not be used to conclude the development of students' critical thinking as they were confounded by research characteristics. Therefore, it was necessary to eliminate the influence of confounding variables first to obtain clearer conclusions. Subsequently, the researchers proposed the results of the study after adjusting the effect sizes using propensity score matching to address research objective 2.

Results of Propensity Score Matching

The application of propensity score matching requires a closely matched dataset for analysis. The researcher divided the data into two groups based on the mean effect size ($\bar{d}$ = 1.669): low effect size group ($\bar{d}$ < 1.669) and high effect size group ($\bar{d}$ > 1.669), as shown in Table 3. Subsequently, this data was analyzed using propensity score matching.

When examining the propensity score distributions of the low and high-effect size groups, it was evident that the low-effect size group had a right-skewed distribution, while the high-effect size group had a left-skewed distribution. However, both groups demonstrated a similar degree of skewness. The propensity scores for both groups ranged from 0.03883 to 0.93959, indicating a comparable distribution of data. To establish continuous score intervals, the researchers divided the data into three quartiles: Q_1 = 0.03883 – 0.33910, Q_2 = 0.33911 –0.63936, and Q_3 = 0.63937 – 0.93959, as illustrated in Figure 3. Considering the closely matched dataset, these intervals were utilized to examine agreement for two-way ANOVA preliminarily. Consequently, it was feasible to compare the samples from both groups to ensure equivalence, as detailed in Table 3.

Table 3. Means and Standard Deviations of Groups

Range of Propensity Score	Groups	*n*	Total	$\bar{d}$	*SD*
d	Low	59	108	0.709	0.044
	High	49		3.030	0.223
Q_1	Low	26	33	0.172	0.088
	High	7		0.264	0.080
Q_2	Low	25	48	0.465	0.086
	High	23		0.492	0.074
Q_3	Low	8	27	0.709	0.059
	High	19		0.740	0.077

Propensity score matching is accomplished by integrating research characteristics that impact the effect sizes into the analysis and extracting the variance using independent sample *t*-tests and two-way ANOVA. This procedure results in the effect sizes being adjusted by propensity score matching, as detailed in Table 4.

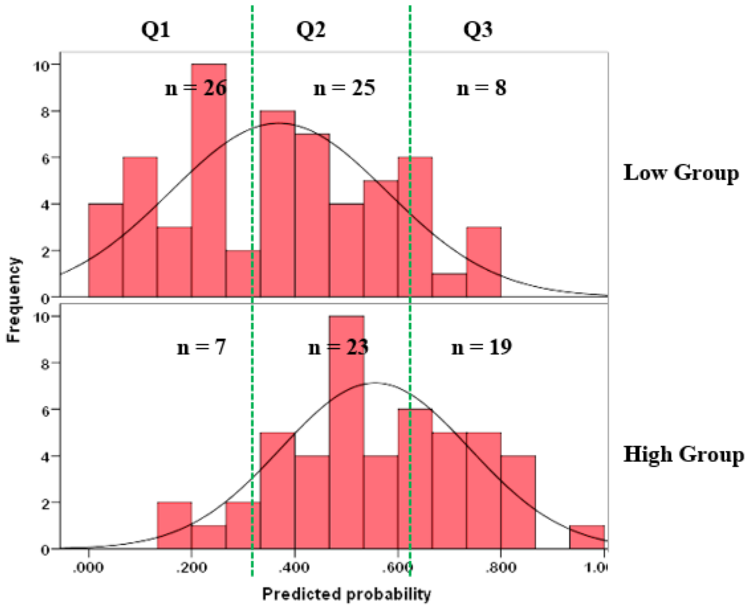


Figure 3. Comparison of the Propensity Score Distributions Between Low and High Effect Size Groups

Table 4. Basic Statistics and Comparison of Before and After-Adjustment Effect Sizes With Propensity Score Matching Between Effect Sizes Groups and Propensity Score Groups

Research Characteristics	Effect Size				Before				After	
	G	*n*	$\bar{d}$	*SD*	*t*-value	*p*-value	*F*-value	*p*-value	*F**-value	*p*-value
Learning Management Approaches	Low	59	3.288	1.378	1.000	.320	2.263	.136	0.173	.841
	High	49	3.020	1.392						
Year of Publication	Low	59	1.254	0.883	0.297	.767	0.154	.696	0.738	.460
	High	49	1.204	0.866						
Research Field	Low	59	1.525	1.278	-1.654	.101	0.013	.911	1.413	.248
	High	49	1.918	1.187						
Courses	Low	59	2.407	1.577	-0.082	.935	3.860×10^{-4}	.984	0.118	.889
	High	49	2.429	1.190						
Level	Low	59	1.085	0.746	-0.551	.583	0.029	.866	0.459	.633
	High	49	1.163	0.665						
Duration per Plan	Low	59	2.356	1.156	0.554	.580	0.382	.358	0.185	.831
	High	49	2.245	0.925						
Total Duration	Low	59	15.988	5.854	2.462*	.015	0.331	.566	0.122	.886
	High	49	13.429	4.564						
Learning Objectives	Low	59	1.169	0.647	-0.649	.518	1.484×10^{-5}	.997	1.072	.346
	High	49	1.245	0.560						

Table 4. Continued

Research Characteristics	Effect Size				Before		After			
	G	*n*	$\overline{d}$	*SD*	*t*-value	*p*-value	*F*-value	*p*-value	*F**-value	*p*-value
Teacher Learning Process	Low	59	8.729	6.395	-1.002	.319	0.156	.694	2.781	.067
	High	49	9.918	5.926						
Student Learning Process	Low	59	2.186	1.025	2.522*	.013	0.454	.502	1.893	.156
	High	49	1.776	0.654						
Learning Media	Low	59	1.695	0.951	-2.537*	.013	1.018	.315	1.296	.278
	High	49	2.102	0.714						
Measurement and Evaluation	Low	59	1.153	0.962	-1.209	.229	1.467	.229	1.546	.218
	High	49	1.367	0.883						
Research Sources	Low	59	0.576	0.498	-1.742	.084	2.721	.102	1.204	.304
	High	49	0.735	0.446						
Sampling	Low	59	1.153	0.665	0.884	.379	0.310	.597	0.520	.596
	High	49	1.041	0.644						
Research Design	Low	59	0.746	0.779	1.748	.083	1.925	.168	0.063	.939
	High	49	0.490	0.739						
Tool Quality Assessment	Low	59	1.729	0.552	1.977*	.049	9.195×10^{-4}	.976	1.595	.208
	High	49	1.490	0.681						
Research Statistics	Low	59	1.915	1.193	-0.391	.697	0.069	.793	0.147	.863
	High	49	2.000	1.061						
Research Quality	Low	59	2.633	0.227	1.307	.195	1.164	.283	1.656	.196
	High	49	2.569	0.275						

*Note: F** is the statistical value for testing the interaction effect between research characteristic variables and propensity score groups, *$p < .05$.

Based on Table 4, four research characteristic variables significantly influence the effect size groups at the .05 level of statistical significance. Subsequently, after-adjustment reveals that the *F*-value of the model with interactions from the two-way ANOVA is low and not statistically significant. Therefore, propensity score matching has effectively eliminated the differences in research characteristics affecting the effect sizes entirely. This can be observed from Figure 4, the funnel plot, after adjustment using propensity score matching (Figure 4b), where the effect sizes of all studies are predominantly clustered around the center. This indicates that they have studied large sample groups, are close to zero, and are not influenced by differing research characteristics. As a result, the conclusions drawn from these studies are likely to be true, as evidenced by the data points mostly lying within the triangular boundary.

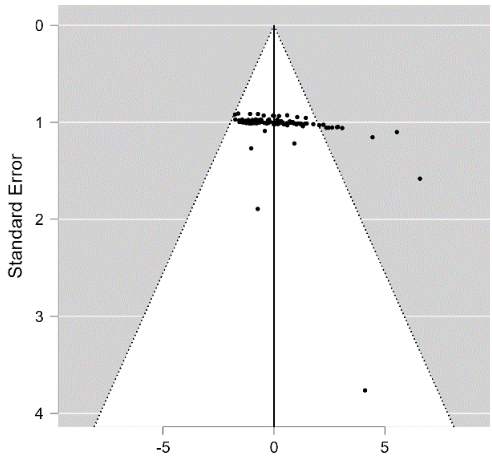


Figure 4a. Before

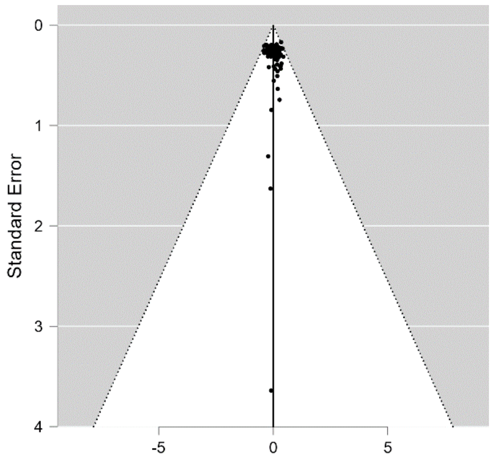


Figure 4b. After

Figure 4. The Effect Sizes of Learning Management Approaches That Developing Students' Critical Thinking Before and After Propensity Score Matching

Results of Comparing the Effect Sizes of Learning Management Approaches After Propensity Score Matching

After adjusting the effect sizes with propensity score matching and conducting tests on the types of influence, it was found that the omnibus test of model coefficients yielded a *Q* value of 6.518 ($p > .05$), indicating that the mean effect size of all research studies (Intercept) did not differences from zero. As for the test of residual heterogeneity, a *Q* value of 59.761 ($p > .05$) was obtained, indicating that the studies examined did not significantly contribute to differences in mean effect size from zero and that the effect sizes of each study did not differ from one another. Therefore, fixed effects were employed for estimation.

As a result, when estimating coefficients using collaborative learning for comparison, inquiry-based learning significantly influenced students' critical thinking at a statistically significant level of .05, as detailed in Table 5.

Table 5. Results of the Effect Sizes Based on Learning Management Approaches

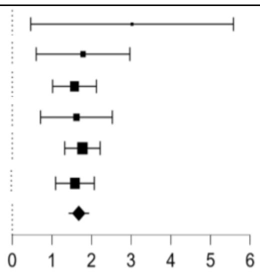
Learning Management Approaches	*k*	$\overline{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b
Collaborative Learning	6	0.242	0.113	[0.02,0.46]	1.294	6.518	59.761
Constructivism	7	0.407	0.088	[0.23,0.58]	1.193		
Learning Through Techniques	25	0.443	0.047	[0.35,0.54]	1.923		
Activity Kits and Media	8	0.428	0.092	[0.25,0.61]	1.488		
Inquiry-Based Learning	49	0.498	0.027	[0.45,0.55]	2.399*		
Problem-Based Learning	13	0.446	0.056	[0.34,0.56]	1.910		
Overall	108	0.465	0.020	[0.43,0.50]	23.103***		

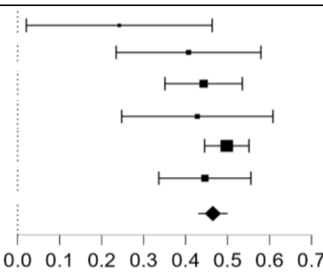
Note: *** $p < .001$, ** $p < .01$, * $p < .05$, no values of τ^2 and I^2 due to fixed effects

Before adjusting the effect sizes (Table 2), no learning management approaches were found to influence students' critical thinking. However, after propensity score matching, it was observed that inquiry-based learning influences the effect sizes at a statistical significance level of .05. This comparison is illustrated before and after propensity score matching in Table 6.

Table 6. Results of Comparing Learning Management Approaches Before and After Propensity Score Matching

Learning Management Approaches	Effect size and 95% confidence interval	
	Before Propensity Score Matching	After Propensity Score Matching
Collaborative Learning	3.02[0.46,5.58]	0.24[0.02,0.46]
Constructivism	1.78[0.60,2.97]	0.41[0.23,0.58]
Learning Through Techniques	1.57[1.02,2.12]	0.44[0.35,0.54]
Activity Kits and Media	1.62[0.71,2.52]	0.43[0.25,0.61]
Inquiry-Based Learning	1.77[1.32,2.22]	0.50[0.45,0.55]
Problem-Based Learning	1.60[1.12,2.09]	0.45[0.34,0.56]
Overall	1.68[1.42,1.94]	0.46[0.43,0.50]





When testing the types of influences on the teacher learning process, it was found that the omnibus test of model coefficients, *Q* value of 7.684 ($p > .05$), indicating that the mean effect sizes of all research studies (Intercept) did not differ from zero. As for the test of residual heterogeneity, a *Q* value of 57.211 ($p > .05$) was obtained, indicating that the studies examined did not contribute to differences in mean effect size from zero, and the effect sizes of each study did not differ from one another. Therefore, fixed effects were used for estimation.

As a result, estimating the coefficients using a learning process starting from introduction, teaching, practice, and summarizing as a basis for comparison found that a learning process starting from elicit, engage, explore, explain, elaborate, evaluate, and extend significantly influenced effect sizes at a statistical significance level of .05, as detailed in Table 7.

Table 7. Results of Effect Sizes Based on Teacher Learning Process

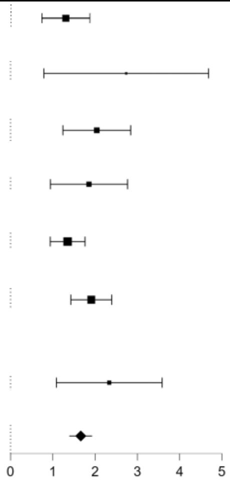
Teacher Learning Process	*k*	$\overline{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b
Introduction, Teach, Practice, and Summarize	19	0.342	0.053	[0.24,0.45]	5.374***	7.684	58.595
Generate Interest, Teach, Practice, Assess, and Reward	8	0.354	0.099	[0.16,0.55]	-0.135		
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	17	0.545	0.037	[0.47,0.62]	2.142*		
Encounter problems, Stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	15	0.415	0.053	[0.31,0.52]	0.479		
Engage, Explore, Explain, Elaborate, and Evaluate	29	0.474	0.037	[0.40,0.55]	1.564		
Generate Interest, Explore, Plan, Practice, Summary and presentation, and Evaluate	11	0.499	0.072	[0.36,0.64]	0.762		
Explore, Define Problem, Analyze, Presentation, Application, and Publication	9	0.550	0.067	[0.42,0.68]	1.708		
Overall	108	0.463	0.029	[0.41,0.52]	16.212***		

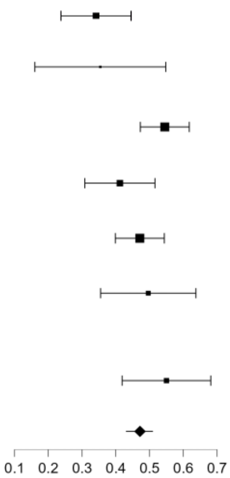
Note: *** $p < .001$, ** $p < .01$, * $p < .05$, no values of τ^2 and I^2 due to fixed effects

Before propensity score matching (Table 2), it was found that the following learning processes influenced students' critical thinking: (a) Generating interest, teaching, practicing, assessing, and rewarding, (b) Eliciting, engaging, exploring, explaining, elaborating, evaluating, and extending, and (c) Exploring, defining problems, analyzing, presenting, applying, and publishing. After propensity score matching, it was found that the process of eliciting, engaging, exploring, explaining, elaborating, evaluating, and extending influenced the effect sizes at a statistical significance level of .05. This is demonstrated by comparing before and after propensity score matching in Table 8.

Table 8. Results of Comparing Teacher Learning Process Before and After Propensity Score Matching

Teacher Learning Process	**Effect size and 95% confidence interval**	
	Before Propensity Score Matching	**After Propensity Score Matching**
Introduction, Teach, Practice, and Summarize	1.30[0.73,1.87]	0.34[0.24,0.45]
Generate Interest, Teach, Practice, Assess, and Reward	2.74[0.79,4.68]	0.35[0.16,0.55]
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	2.04[1.24,2.84]	0.54[0.47,0.62]
Encounter problems, Stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	1.85[0.94,2.77]	0.41[0.31,0.52]
Engage, Explore, Explain, Elaborate, and Evaluate	1.35[0.94,1.76]	0.47[0.40,0.55]
Generate Interest, Explore, Plan, Practice, Summary and presentation, and Evaluate	1.91[1.43,2.39]	0.50[0.36,0.64]
Explore, Define the Problem, Analyze, Presentation, Application, and Publication	2.33[1.08,3.59]	0.55[0.42,0.68]
Overall	1.66[1.39,1.93]	0.47[0.43,0.51]





Discussion

Research characteristics derived from the selection of samples from the TDC database, where students, teachers, and principals are not from the same group, do not cover qualitative research due to the inability to calculate effect sizes for meta-analysis and cannot access concealed research. The research findings thus represent only half of all studies, which found that collaborative learning has the greatest impact on students' critical thinking. However, this conclusion was influenced by various research characteristics because several institutions and fields of education develop innovative learning management, resulting in differences in research outcomes. This is due to the lack of rigor in research methodology design, sample specification, randomization, controlling confounding variables according to experimental design principles, and verifying preliminary statistical agreements, resulting in a small-study effect where small sample groups have high effect sizes and low research quality have high effect sizes. Similar to the research of Dowdy et al. (2020) and Ferguson and Brannick (2012), these reasons as significant causes of publication bias (Thornton & Lee, 2000; Vevea et al., 2019). This bias arises from budget constraints, institutional rigor, negative research result concealment, acceptance of positive research outcomes, and statistically significant results. Examining publication bias from funnel plots, which show uneven distribution and deviation from overall mean values, or detecting publication bias through other methods (e.g., Egger's test), the results are not entirely trustworthy and should be approached with caution (Egger et al., 1997; Tang & Liu, 2000). They cannot be generalized to the broader population (Allen, 2020; Y. H. Lee, 2019). Therefore, sample selection is important in conducting meta-analysis studies (StataCorp, 2023; Thornton & Lee, 2000; Vevea et al., 2019).

Research studies on teacher learning processes have demonstrated a significant positive impact on students' critical thinking at an advanced level. The teacher-learning process significantly influences the effect sizes of students' critical thinking at a statistically significant level of .05, according to Abrami et al. (2008), Mahapoonyanont (2010), Niu et al. (2013), and Oliveira et al. (2016). They explain that the teacher-learning process is a crucial part of student development. Students who learn to explore problems, expand their thinking through collaborative analysis, and assess outcomes are stimulated to think, ask questions, analyze data, express themselves, and engage in cooperative learning, thereby enhancing their critical thinking (Zhao et al., 2016). Several components contribute to this finding, including the research field, courses, total duration, teacher learning process, learning media, measurement and evaluation, research design, research statistics, and research quality. All these factors positively influence managing learning to develop students' critical thinking, as evidenced by the research of Abrami et al. (2015), Niu et al. (2013), and Xu et al. (2023). Moreover, the researchers identified two interesting issues: First, the year of publication. Despite the absence of a statistically significant correlation, the analysis indicates a positive trend in Thailand's educational management. This is due to changes in the Basic Education Core Curriculum aimed at enhancing cognitive competency and shifting from teacher-centered to active learning methods, resulting in consistent improvement in students' critical thinking skills. Second, high-impact research characteristics but small sample sizes were identified, although statistically insignificant. Such characteristics include sample selection, simple random sampling, multi-stage sampling, research design, randomized control group pretest-posttest design, duration, group-based learning processes with collaborative discussions, and post-learning assessments. These components contribute to high-quality research and tend to have a positive impact on students' critical thinking. However, increasing the sample sizes will boost the power of statistical analysis, resulting in stronger and more reliable conclusions from the meta-analysis (Borenstein & Higgins, 2013). Nevertheless, the differences in research characteristics that affect research outcomes have not yet been addressed for accurate and appropriate application (Austin, 2009; Borenstein et al., 2021). Therefore, the data balance needs to be adjusted by removing the influence of confounding variables. Propensity Score Matching can help reduce discrepancies and make comparative results more objective (Heinrich et al., 2010).

After propensity score matching, researchers observed three significant findings. First, the high-effect sizes group (n = 49) had higher propensity scores compared to the low-effect sizes group (n = 59). Upon comparing the distribution of propensity scores between both groups, the scores were balanced and similar, allowing for effective propensity score matching to mitigate disparities between the two groups (Badhiwala et al., 2021; Benedetto et al., 2018). Second, propensity score matching successfully controlled for the interference of research characteristics, with no variables significantly affecting the effect sizes. Third, after propensity score matching, the mean effect size of the research did not significantly deviate from zero (Q_a = 6.518), and the residual value from approximating zero (Q_b = 59.761) indicated that effect sizes were not significantly influenced by research characteristics statistically. Therefore, it can be concluded that inquiry-based learning positively influences students' critical thinking, serving as a teacher-learning process where students engage in exploration, problem identification, analysis, elaboration, evaluation, and extension. (Balta & Sarac, 2016). Students undertake deep analysis, synthesis, and comprehensive thinking to formulate reasoned conclusions, as per Ennis' critical thinking theory (Ennis, 1987; Kocak et al., 2021). Furthermore, the transfer of knowledge from existing to new knowledge creation, beyond mere explanation, underscores the significant role of teachers in stimulating and expanding students' thinking (Eisenkraft, 2003). This creates opportunities for students to comprehend, apply knowledge to problem-solving, evaluate data, communicate, and present through collaborative learning exchanges, fostering teamwork skills (Chu et al., 2017; Hitchcock, 2017; O'Reilly et al., 2022). Additionally, students reflect, analyze relationships, practice critical thinking, and solve problems through daily life experiences, such as communication and writing, significantly enhancing their thoughtful reasoning skills (Abrami et al., 2015; Ennis, 2015; Lai, 2011). Research by Muthma'Innah et al. (2019) and Suardana et al. (2018) demonstrated that students receiving inquiry-based learning 7E showed significantly higher development of critical thinking than control group students statistically, with improvements in information presentation and explanation and in-depth analysis of responses.

Conclusion

The conclusions drawn from this study are as follows: (a) the overall impact of research on learning management approaches aimed at developing students' critical thinking was notably high, with nine research characteristic variables showing statistically significant influence at the .05 level. These variables encompassed research field, courses, total duration, teacher learning process, learning media, measurement and evaluation, research design, research statistics, and research quality, and (b) after propensity score matching, it was found that inquiry-based learning, which emphasizes learning processes that encourage problem exploration, expanding thinking through collaborative analysis, and applying diverse media and activity sheets tailored to context suitability, can develop students' critical thinking.

Recommendations

The researchers have three recommendations for this study. First, enhancing learning management efficiency: teachers can employ inquiry-based learning focusing on problem exploration, expanding thinking through collaborative analysis, and applying, especially in science, integrating learning from multiple media sources, and conducting measurement and evaluation through posttest and assignment checks to develop students' critical thinking tailored to the context. Second, controlling confounding variables in research design: instructors should utilize conclusions regarding confounding variables affecting students' critical thinking to design research, enhancing control over these variables. Third, employing propensity score matching for meta-analysis: helps reduce publication bias and type I error in research findings, enhancing the completeness of the study.

For future research, three additional suggestions are provided. Firstly, defining the scope of study variables: specifying the characteristics and sub-variables studied facilitates data collection and analysis. Secondly, meta-analysis: can be used to gather research outcomes from various sources, summarizing diverse research findings and exploring analyses of other variables of interest, such as the impact of learning management approaches on mathematical performance or mathematical skills. Lastly, propensity score matching: to match sample groups in research to achieve comparability in uninterested variables, thereby enhancing the completeness of research outcomes.

Limitations

The researchers identified a limitation, namely, the sample size of the study. In this research, intriguing variables had small sample sizes but yielded high effect sizes. With future advancements and increased research volume, clearer conclusions can be drawn.

Acknowledgments

The researchers express their gratitude to the Faculty of Education, Chiang Mai University, for providing funding to support the dissemination of this research. We also extend our appreciation to Mr. Tanapat Itsarangkul Na Ayutthaya, an expert in assessing inter-rater reliability and providing guidance in data analysis.

Authorship Contribution Statement

Surin: Concept, design, data acquisition, data analysis, statistical analysis, drafting manuscript, critical revision of manuscript, and writing. Damrongpanit: Consulting research, data analysis, statistical analysis, editing/reviewing, technical or material support, final approval, and securing funding.

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

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*(4), 1102-1134. https://doi.org/10.3102/0034654308326084

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659-701. https://doi.org/10.3102/0034654316689306

Akinbadewa, B. O., & Sofowora, O. A. (2020). The effectiveness of multimedia instructional learning packages in enhancing secondary school students' attitudes toward Biology. *International Journal on Studies in Education*, *2*(2), 119-133. https://doi.org/10.46328/ijonse.19

Allen, M. (2020). Understanding the practice, application, and limitations of meta-analysis. *American Behavioral Scientist*, *64*(1), 74-96. https://doi.org/10.1177/0002764219859619

Alsaleh, N. J. (2020). Teaching critical thinking skills: Literature review. *Turkish Online Journal of Educational Technology*, *19*(1), 21-39. https://bit.ly/49QkEcI

Amineh, R. J., & Asl, H. D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages*, *1*(1), 9-16. http://bit.ly/3v23J8v

Austin, P. C. (2009). The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Medical Decision Making*, *29*(6), 661-677. https://doi.org/10.1177/0272989X09341755

Badhiwala, J. H., Karmur, B. S., & Wilson, J. R. (2021). Propensity score matching: A powerful tool for analyzing observational nonrandomized data. *Clinical Spine Surgery*, *34*(1), 22-24. https://doi.org/10.1097/BSD.0000000000001055

Bai, H. (2011). A comparison of propensity score matching methods for reducing selection bias. *International Journal of Research and Method in Education*, *34*(1), 81-107. https://doi.org/10.1080/1743727X.2011.552338

Balta, N., & Sarac, H. (2016). The effect of 7e learning cycle on learning in science teaching: a meta-analysis study. *European Journal of Educational Research*, *5*(2), 61-72. https://doi.org/10.12973/eu-jer.5.2.61

Benedetto, U., Head, S. J., Angelini, G. D., & Blackstone, E. H. (2018). Statistical primer: Propensity score matching and its alternatives. *European Journal of Cardio-Thoracic Surgery*, *53*(6), 1112-1117. https://doi.org/10.1093/ejcts/ezy167

Berkhout, S. W., Haaf, J. M., Gronau, Q. F., Heck, D. W., & Wagenmakers, E.-J. (2024). A tutorial on Bayesian model-averaged meta-analysis in JASP. *Behavior Research Methods, 56,* 1260-1282. https://doi.org/10.3758/s13428-023-02093-6

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (Eds.). (2021). *Introduction to meta-analysis*. John Wiley and Sons. https://bit.ly/3SGZV4M

Borenstein, M., & Higgins, J. P. T. (2013). Meta-analysis and subgroups. *Prevention Science*, *14*, 134-143. https://doi.org/10.1007/s11121-013-0377-7

Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guilford Press. https://bit.ly/3HWcMuP

Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2017). *21st century skills development through inquiry-based learning: From theory to practice.* Springer. https://doi.org/10.1007/978-981-10-2481-8

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation. https://doi.org/10.7758/9781610448864

Czodrowski, P. (2014). Count on kappa. *Journal of Computer-Aided Molecular Design*, *28*, 1049-1055. https://doi.org/10.1007/s10822-014-9759-6

Donoghue, G. M., & Hattie, J. A. C. (2021). A meta-analysis of ten learning techniques. *Frontiers in Education*, *6*, Article 581216. https://doi.org/10.3389/feduc.2021.581216

Dowdy, A., Tincani, M., & Schneider, W. J. (2020). Evaluation of publication bias in response interruption and redirection: A meta-analysis. *Journal of Applied Behavior Analysis*, *53*(4), 2151-2171. https://doi.org/10.1002/jaba.724

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking skills and Creativity*, *12*, 43-52. https://doi.org/10.1016/j.tsc.2013.12.004

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629-634. https://doi.org/10.1136/bmj.315.7109.629

Eisenkraft, A. (2003). Expanding the 5e model. *The Science Teacher*, *70*(6), 56-59. https://bit.ly/3GLas9A

Elder, L., & Paul, R. (2021). *Critical thinking: Tools for taking charge of your learning and your life* (4th ed.). Rowman & Littlefield.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). W H Freeman/Times Books/ Henry Holt & Co.

Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice*, *32*(3), 179-186. https://doi.org/10.1080/00405849309543594

Ennis, R. H. (2015). Critical thinking: A streamlined conception. In M. Davies, & R. Barnett, (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 31-47). Palgrave Macmillan. https://doi.org/10.1057/9781137378057_2

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, *37*, 165-184. https://doi.org/10.1007/s11245-016-9401-4

Esterhuizen, T. M., & Thabane, L. (2016). Con: Meta-analysis: some key limitations and potential solutions. *Nephrology Dialysis Transplantation*, *31*(6), 882-885. https://doi.org/10.1093/ndt/gfw092

Facione, P. A. (2023). *Critical thinking: What it is and why it counts*. Insight Assessment. https://bit.ly/44g6a4L

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*(1), 120-128. https://doi.org/10.1037/a0024445

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8. https://doi.org/10.3102/0013189X005010003

Green, P. (2015). Teaching critical thinking for lifelong learning. In M. Davies, & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 107-121). Palgrave Macmillan. https://doi.org/10.1057/9781137378057_7

Harris, H., & Horst, S. J. (2016). A brief guide to decisions at each step of the propensity score matching process, *Practical Assessment, Research, and Evaluation*, *21*(4), 1-11. https://doi.org/10.7275/yq7r-4820

Harrison, F. (2011). Getting started with meta-analysis. *Methods in Ecology and Evolution*, *2*(1), 1-10. https://doi.org/10.1111/j.2041-210X.2010.00056.x

Haukoos, J. S., & Lewis, R. J. (2015). The propensity score. *JAMA*, *314*(15), 1637-1638. https://doi.org/10.1001/jama.2015.13480

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). *A primer for applying propensity-score matching*. Inter-American Development Bank. https://bit.ly/3HXEuHw

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Hitchcock, D. (2017). Critical thinking as an educational ideal. In *On reasoning and argument*. (Vol. 30, pp. 477-497). Springer. https://doi.org/10.1007/978-3-319-53562-3_30

Indah, R. N., & Kusuma, A. W. (2016). Factors affecting the development of critical thinking of Indonesian learners of English language. *Journal of Humanities and Social Science*, *21*(6), 86-94. https://bit.ly/4aoSkQg

Itsarangkul Na Ayutthaya, T., & Damrongpanit, S. (2022a). A meta-analysis of instructional management models affecting creative thinking development. *European Journal of Educational Research*, *11*(4), 2069-2085. https://doi.org/10.12973/eu-jer.11.4.2069

Itsarangkul Na Ayutthaya, T., & Damrongpanit, S. (2022b). A meta-analysis of instructional management model for students' creative thinking development: An application of propensity score matching. *European Journal of Educational Research*, *11*(4), 2429-2444. https://doi.org/10.12973/eu-jer.11.4.2429

Jacobs, G. M., & Renandya, W. A. (2019). *Student centered cooperative learning: Linking concepts in education to promote student learning*. Springer. https://doi.org/10.1007/978-981-13-7213-1

Kane, L. T., Fang, T., Galetta, M. S., Goyal, D. K. C., Nicholson, K. J., Kepler, C. K., Vaccaro, A. R., & Schroeder, G. D. (2020). Propensity score matching: a statistical method. *Clinical Spine Surgery*, *33*(3), 120-122. https://doi.org/10.1097/BSD.0000000000000932

Karampelas, K. (2023). Critical thinking in national primary science curricula. *Eurasian Journal of Science and Environmental Education*, *3*(2), 51-60. https://doi.org/10.30935/ejsee/13271

Kennedy, T. J., & Sundberg, C. W. (2020). 21st century skills. In B. Akpan, & T. J. Kennedy (Eds.), *Science education in theory and practice* (pp. 479-496). Springer. https://doi.org/10.1007/978-3-030-43620-9_32

Kocak, O., Coban, M., Aydin, A., & Cakmak, N. (2021). The mediating role of critical thinking and cooperativity in the 21st century skills of higher education students. *Thinking Skills and Creativity*, *42*, Article 100967. https://doi.org/10.1016/j.tsc.2021.100967

Lai, E. R. (2011). *Critical thinking: A literature review*. Research Reports. https://bit.ly/42zDaUG

Lambert, C. T., & Guillette, L. M. (2021). The impact of environmental and social factors on learning abilities: a meta-analysis. *Biological Reviews*, *96*(6), 2871-2889. https://doi.org/10.1111/brv.12783

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal*, *109*(5), 491-509. https://doi.org/10.1086/596998

Lee, W.-S. (2013). Propensity score matching and variations on the balancing test. *Empirical Economics*, *44*, 47-80. https://doi.org/10.1007/s00181-011-0481-0

Lee, Y. H. (2019). Strengths and limitations of meta-analysis. *The Korean Journal of Medicine*, *94*(5), 391-395. https://doi.org/10.3904/kjm.2019.94.5.391

Mahapoonyanont, N. (2010). Factors related to critical thinking abilities; a meta-analysis. *Procedia Social and Behavioral Sciences*, *9*, 986-990. https://doi.org/10.1016/j.sbspro.2010.12.272

McPeck, J. E. (Ed.). (2016). *Critical thinking and education*. Routledge. https://doi.org/10.4324/9781315463698

Moallem, M., Hung, W., & Dabbagh, N. (Eds.). (2019). *The Wiley handbook of problem-based learning*. John Wiley & Sons Inc. https://doi.org/10.1002/9781119173243

Morgan, C. J. (2018). Reducing bias using propensity score matching. *Journal of Nuclear Cardiology*, *25*(2), 404-406. https://doi.org/10.1007/s12350-017-1012-y

Moust, J., Bouhuijs, P., & Schmidt, H. (2021). *Introduction to problem-based learning* (4th ed.). Routledge. https://doi.org/10.4324/9781003194187

Muthma'Innah, M., Dahlan, J. A., & Suhendra, S. (2019). Ability of mathematical critical thinking–what about learning cycle 7e model? *Journal of Physics: Conference Series*, *1157*(3), Article 032129. https://doi.org/10.1088/1742-6596/1157/3/032129

Nicol, C., Yakuba, E., & Habinshuti, G. (2020). An overview of learning cycles in science inquiry-based instruction. *African Journal of Educational Studies in Mathematics and Sciences*, *16*(2), 67-81. https://doi.org/10.4314/ajesms.v16i2.5

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, *9*, 114-128. https://doi.org/10.1016/j.edurev.2012.12.002

Oliveira, L. B., Díaz, L. J., Carbogim, F. C., Rodrigues, A. R. B., & Püschel, V. A. A. (2016). Efetividade das estratégias de ensino no desenvolvimento do pensamento crítico de graduandos de Enfermagem: uma metanálise [Effectiveness of teaching strategies on the development of critical thinking in undergraduate nursing students: a meta-analysis]. *Journal of School of Nursing – University of São Paulo/Revista da Escola de Enfermagem da USP*, *50*(2), 355-364. https://doi.org/10.1590/S0080-623420160000200023

O'Reilly, C., Devitt, A., & Hayes, N. (2022). Critical thinking in the preschool classroom-a systematic literature review. *Thinking Skills and Creativity*, *46*, Article 101110. https://doi.org/10.1016/j.tsc.2022.101110

Paul, R., & Elder, L. (2005). *A guide for educators to critical thinking competency standards*. Foundation for Critical Thinking. https://bit.ly/48gtYFY

Paul, R., & Elder, L. (2020). *The miniature guide to critical thinking concepts and tools.* Rowman and Littlefield. https://bit.ly/3OHHBHr

Rannikmäe, M., Holbrook, J., & Soobard, R. (2020). Social constructivism—Jerome Bruner. In B. Akpan, & T. J. Kennedy (Eds.), *Science education in theory and practice* (pp. 259-275). Springer. https://doi.org/10.1007/978-3-030-43620-9_18

Reiffel, J. A. (2020). Propensity score matching: the 'devil is in the details' where more may be hidden than you know. *American Journal of Medicine*, *133*(2), 178-181. https://doi.org/10.1016/j.amjmed.2019.08.055

Ripley, D. (2015). *An examination of flipped instructional method on sixth graders' mathematics learning: Utilizing propensity score matching*. University of Nevada. https://bit.ly/49cwezm

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, *7*, 143-176. https://doi.org/10.1146/annurev-statistics-031219-041058

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, *4*, 61-81. https://doi.org/10.1007/s11292-007-9046-9

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127*(8-2), 757-763. https://doi.org/29d

Rücker, G., Schwarzer, G., Carpenter, J. R., & Schumacher, M. (2008). Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology*, *8*, Article 79. https://doi.org/10.1186/1471-2288-8-79

Sainani, K. L. (2012). Propensity scores: uses and limitations. *PM&R*, *4*(9), 693-697. https://doi.org/10.1016/j.pmrj.2012.07.002

Saiz, C., & Rivas, S. F. (2016). New teaching techniques to improve critical thinking. The Diaprove methodology. *Educational Research Quarterly*, *40*(1), 3-36.

Saylor, J. G., Alexander, W. M., & Lewis, A. J. (1981). *Curriculum planning for better teaching and learning* (4th ed.). Holt, Rinehart and Winston.

Sedgwick, P., & Marston, L. (2015). How to read a funnel plot in a meta-analysis. *BMJ*, *351*, Article h4718. https://doi.org/10.1136/bmj.h4718

Slavin, R. E. (2014). Aprendizaje cooperativo y rendimiento académico: ¿por qué funciona el trabajo en grupo? [Cooperative learning and academic achievement: Why does groupwork work?]. *Annals of Psychology/Anales De Psicología, 30*(3), 785-791. https://doi.org/10.6018/analesps.30.3.201201

Staffa, S. J., & Zurakowski, D. (2018). Five steps to successfully implement and evaluate propensity score matching in clinical research studies. *Anesthesia and Analgesia*, *127*(4), 1066-1073. https://doi.org/10.1213/ANE.0000000000002787

StataCorp. (2023). *Stata 18 meta-analysis reference manual: Release 18*. Stata Press. https://www.stata.com/manuals/meta.pdf

Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. National Inst. of Education.

Stone, D. L., & Rosopa, P. J. (2017). The advantages and limitations of using meta-analysis in human resource management research. *Human Resource Management Review*, *27*(1), 1-7. https://doi.org/10.1016/j.hrmr.2016.09.001

Suardana, I. N., Redhana, I. W., Sudiatmika, A. A. I. A. R., & Selamat, I. N. (2018). Students' critical thinking skills in chemistry learning using local culture-based 7e learning cycle model. *International Journal of Instruction*, *11*(2), 399-412. https://doi.org/10.12973/iji.2018.11227a

Tang, J.-L., & Liu, J. L. (2000). Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, *53*(5), 477-484. https://doi.org/10.1016/S0895-4356(99)00204-8

Thongnuypram, C., & Sopheerak, S. (2013). Factor influencing the critical thinking of teacher students studying at the faculty of education in Suratthani Rajabhat University. *Procedia-Social and Behavioral Sciences*, *103*, 386-391. https://doi.org/10.1016/j.sbspro.2013.10.349

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207-216. https://doi.org/10.1016/S0895-4356(99)00161-4

Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, *21*(4), 589-624. https://doi.org/10.1002/sim.1040

Varoglu, L., Yilmaz, A., & Sen, S. (2023). Effect of 5E learning cycle assisted with concept maps on conceptual understanding. *Pedagogical Research*, *8*(3), Article em0161. https://doi.org/10.29333/pr/13167

Vevea, J. L., Coburn, K. M., & Sutton, A. J. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383-429). Russell Sage Foundation, https://doi.org/10.7758/9781610448864.21

Xu, E., Wang, W., & Wang, Q. (2023). The effectiveness of collaborative problem solving in promoting students' critical thinking: A meta-analysis based on empirical literature. *Humanities and Social Sciences Communications*, *10*, Article 16. https://doi.org/10.1057/s41599-023-01508-1

Zhao, C., Pandian, A., & Singh, M. K. M. (2016). Instructional strategies for developing critical thinking in EFL classrooms. *English Language Teaching*, *9*(10), 14-21. https://doi.org/10.5539/elt.v9n10p14