European Journal of Educational Research
Volume 13, Issue 4, 1535 - 1553.
ISSN: 2165-8714
https://www.eu-jer.com/

Quantifying Influence: Propensity Score Matching Unravels the True Effect Sizes of Learning Management Models on Students' Analytical Thinking

Supansa Surin
Chiang Mai University, THAILAND

Suntonrapot Damrongpanit*
Chiang Mai University, THAILAND

Received: January 3, 2024 ▪ Revised: April 17, 2024 ▪ Accepted: May 10, 2024

Abstract: Analytical thinking is crucial for developing problem-solving, decision-making, and higher-order thinking skills. Many researchers have consistently developed learning management models to enhance students' analytical thinking, resulting in extensive knowledge but lacking clear systematic summaries. This study aims to: (a) explore the effect sizes and research characteristics influencing students' analytical thinking, and (b) compare the effect sizes of learning management models after adjusting the effect sizes by propensity score matching. In exploring 131 graduate research papers published between 2002 and 2021. The research utilized research characteristics recording forms and research quality assessment forms for data collection. Effect sizes were calculated using Glass's method, while data analysis employed random effects, fixed effects, and regression meta-analysis methods. The findings indicate that (a) research on learning management models significantly impacts students' analytical thinking at a high level ($\bar{d}$ = 1.428), Seven research characteristics, including year of publication, field of research, level, duration per plan, learning management process, measurement and evaluation, and research quality, statistically influence students' analytical thinking, and (b) after propensity score matching, learning through techniques such as KWL, KWL-plus, Six Thinking Hats, 4MAT, and Mind Mapping had the highest influence on students' analytical thinking. Recommendations for developing students' analytical thinking involve creating a learning management process that fosters understanding, systematic practical training, expanding thinking through collaborative exchanges, and assessments using learning materials and tests to stimulate increased analytical thinking.

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

To cite this article: Surin, S., & Damrongpanit, S. (2024). Quantifying influence: Propensity score matching unravels the true effect sizes of learning management models on students' analytical thinking. *European Journal of Educational Research*, *13*(4), 1535-1553. https://doi.org/10.12973/eu-jer.13.4.1535

Introduction

The current goal of learner development is to cultivate individual characteristics that align with the demands of the labor market in the 21st century. This requires everyone to have analytical thinking, critical thinking, problem-solving thinking, creative thinking, collaboration, responsible citizenship, and lifelong learning skills (González-Pérez & Ramírez-Montoya, 2022; Kennedy & Sundberg, 2020; World Economic Forum, 2023). Many scholars are thus interested in developing analytical thinking in students. This involves the continuous development of curriculum, learning management models, and assessment methods (Cornell, 2024; Few, 2015; Thaneerananon et al., 2016) to promote higher-order thinking in students according to Bloom's Taxonomy (Krathwohl, 2002). Individuals with analytical thinking can identify, categorize, and break down complex problems into subcomponents to explore relationships, significance, and systemic principles (Bloom, 1956; Krathwohl, 2002). They can critically evaluate problems or events and thoroughly filter information, enabling precise data processing, efficient problem-solving, sound decision-making, and accurate fact-checking (Amer, 2005; Rasheva-Yordanova et al., 2018; Robbins, 2011). This contributes to the development of critical thinking, creative thinking, and innovation for the benefit of society, the environment, and the world (Kao, 2014).

Analytical thinking holds significant importance in education, especially in the context of preparing students for the demands of the 21st century. It is a skill that enables students to analyze complex problems, discern patterns, and solve problems logically when faced with diverse challenges (Binkley et al., 2012). Analytical thinking in education promotes deep understanding by encouraging students to explore connections and draw conclusions from evidence. This not only

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

enhances learning effectiveness and confidence but also cultivates a lifelong learning attitude, which is crucial for learning success in the 21st century (Wagner, 2008). Educational institutions have implemented diverse approaches to develop students' analytical thinking, akin to those in Thailand. Several educational institutions have recognized the importance of developing students' analytical thinking through continuous graduate-level research from 1995 to the present (n = 1,912). Preliminary surveys revealed that most of the research about learning management models (n = 1,272), serves as frameworks or structures used for planning, designing, developing, implementing, and evaluating through students, teachers, and stakeholders according to contextual appropriateness. This framework encompasses curriculum, learning objectives, the learning management process, learning media, and assessment methods to facilitate student learning and achieve desired goals systematically (Morrison et al., 2019; Smith & Ragan, 2005). It serves as a guideline for teachers to understand the learning management process, which sequentially organizes experiences for students, explaining what teachers practice and what students will receive, enabling teachers to effectively develop students' analytical thinking. This is evident from the research of Sartika (2018) and Suyatman et al. (2021). When considering the learning management models that have been developed, they can be categorized into six learning management models (Saylor et al., 1981), as follows: (a) Collaborative learning (12.21%) involves group learning, where each member possesses different abilities, roles are assigned to promote exchange within the group, fostering knowledge and teamwork skills. This leads to effective problem-solving and communication. Examples include Teams-Games-Tournament (TGT), Student Teams Achievement Division (STAD), and Student Teams Achievement Division (TAI) (Gillies, 2014; Jacobs & Renandya, 2019; Slavin, 2014), (b) Constructivism (14.50%) involves organizing experiences to enable learners to generate questions, inquire, analyze, connect, debate, and exchange knowledge, leading to the creation of conclusions applicable to problem-solving on one's own (Clark, 2018; Rannikmäe et al., 2020; Triantafyllou, 2022), (c) Learning through techniques (19.09%) involves targeted and specialized learning techniques integrating questioning, component analysis practice, differentiation, fostering understanding, reasoning, and systematic thinking such as know-want-learn strategy (KWL), know-want-learn plus strategy (KWL-plus), Six Thinking Hats, 4MAT, and Mind Mapping (Sartika, 2018; Spaska et al., 2021), (d) Activity kits and media (9.16%) involve a learning approach aligning with learners' needs by designing step-by-step activities and carrying them out comprehensively. It combines the use of multimedia to generate interest and efficient learning tools (Adelana et al., 2021; Akinbadewa & Sofowora, 2020; Lampert & Graziani, 2009), (e) Inquiry-based learning (29.77%) involves exploration, observation, investigation, and experimentation, allowing students to describe, exchange learning, expand thinking, present ideas, and draw conclusions applicable to problem-solving in various situations. Examples include 3E, 5E, and 7E (Balta & Sarac, 2016; Khairani et al., 2021; Nicol et al., 2020; Varoglu et al., 2023), and (f) Problem-based learning (15.27%) emphasizes the centrality of problems, allowing students to identify, explore, investigate, plan, solve problems, summarize, and evaluate results. It involves learners facing problems, exchanging problem-solving methods, and applying them effectively (Moallem et al., 2019; Moust et al., 2021; Tan, 2021).

In addition, significant research characteristics have been identified in studies concerning learning management models that develop students' analytical thinking. These studies predominantly originate from the curriculum and instruction field, utilizing a randomized control group pretest-posttest design for research design, employing independent sample t-tests for data analysis, and maintaining research quality at a good level. Upon scrutinizing these studies, researchers noted two noteworthy observations. Firstly, despite employing the same learning management models, there are variations in research outcomes, leading to ambiguous conclusions. Consequently, teachers may find themselves perplexed regarding the earnest application of research findings. Secondly, while the research quality is low, the effect size is high. This discrepancy arises from incomplete research procedures, such as vague descriptions of the randomization process, incomplete evaluation of tool quality, and failure to verify basic assumptions in statistical data analysis. These research characteristics render the research results unreliable (Akkerman et al., 2008; Patton, 1999). Thus, concluding this research will help reduce confusion and clarify previous research findings. Meta-analysis can systematically generate these conclusions and effectively utilize the gathered information to enhance the development of students' analytical thinking appropriately.

Meta-analysis is a statistical technique that systematically compiles quantitative research findings on the same topic (Glass, 1976). It systematically analyzes the collected research, organizing the results into "effect sizes" to facilitate the study and comparison of variables of interest. This methodology aims to generate clear and comprehensive conclusions, providing a reliable explanation for differences in research findings (Borenstein et al., 2021; Card, 2012; Cooper et al., 2019). Currently, meta-analysis is utilized in the field of education, employing it for three main purposes. Firstly, it evaluates the effectiveness of different learning management models, learning media, and technologies on student skills and learning outcomes (Donoghue & Hattie, 2021; Sailer & Homner, 2020). Secondly, it identifies factors influencing student skills and learning outcomes through the study of the relationship between the environment and student learning. This includes investigating how environmental factors affect student learning and the influence of teacher characteristics on student learning (Burch et al., 2019; Lambert & Guillette, 2021). Lastly, it examines the outcomes of testing methodologies (Adesope et al., 2017; Rowland, 2014). Meta-analysis contributes to enhancing the power of statistics by leveraging large sample sizes, helping address issues related to research with small sample sizes that yield interesting results, leading to more accurate conclusions. Additionally, meta-analysis examines the variability in research findings to understand the influence of confounding variables, resulting in precise and comprehensive conclusions. These conclusions can guide treatment, development, policymaking, or forecasting of present and future changes (Lee, 2019; Stone & Rosopa, 2017). However, to conduct an effective meta-analysis, it is crucial to select samples that encompass all aspects and ensure equal distribution. Selection of previous research samples has indicated limitations in accessing comprehensive research due to biases favoring positive outcomes, lack of direct reflection of results, and the inability to verify distorted research findings against reality (Ahmed et al., 2012; Esterhuizen & Thabane, 2016). Consequently, the effect sizes are influenced by confounding variables, as exemplified by the study of Itsarangkul Na Ayutthaya and Damrongpanit (2022a), Niu et al. (2013), and Xu et al. (2023). These research characteristics encompass learning management processes, levels, field of research, total duration, and research quality, all of which impact the effect sizes. Therefore, the conclusions drawn from the meta-analysis cannot be applied directly. Propensity score matching must be employed to control the influence of confounding variables to achieve clearer conclusions (Egger et al., 1997).

Propensity score matching is a statistical method that simulates a randomized controlled trial to reduce bias in non-randomized samples (Kane et al., 2020; Morgan, 2018). It aims to decrease the variability of confounding variables (Rosenbaum & Rubin, 2023) by ensuring that the comparison groups are as similar as possible (Benedetto et al., 2018). The technique adjusts experimental outcomes based on research characteristics serving as confounding variables, allowing for a controlled comparison of treatment effects (Haukoos & Lewis, 2015; Thoemmes, 2012). It aids in minimizing the occurrence of publication bias and enhances the comparability of results for specific objectives (Bai, 2011; Staffa & Zurakowski, 2018). According to research conducted by Itsarangkul Na Ayutthaya and Damrongpanit (2022b), where propensity score matching was utilized for meta-analysis, a significant difference was identified. Before propensity score matching, the study revealed that instructional designs emphasizing self-directed knowledge creation alongside the use of technology, as well as instructional designs focusing on activity-based learning and creative learning environments, had a notable impact on creative thinking. After propensity score matching, it was observed that instructional designs emphasizing integrated knowledge creation within an environment fostering creativity using technology influenced creative thinking. Consequently, propensity score matching contributes to drawing clearer conclusions about the learning management model that develops students' analytical thinking (Rubin, 1997).

Based on the reasons mentioned above, research on learning management models that develop students' analytical thinking has yielded conflicting results, which may stem from the characteristics of the research itself. Therefore, researchers want to know how the effect sizes of each research vary. Among the fourteen variables of the research, which variables influence the research results differently? If the influence of the research characteristics is eliminated, which of the six learning management models has the greatest impact on students' analytical thinking? Therefore, the research objectives are defined as follows: (a) explore the effect sizes and research characteristics influencing students' analytical thinking, and (b) compare the effect sizes of the learning management models that develop students' analytical thinking after adjusting the effect sizes using propensity score matching.

Methodology

Research Design

This study is a meta-analysis of experimental studies focusing on learning management models that develop students' analytical thinking. The research utilizes Glass's method (1976) to calculate the effect sizes based on the research findings. Subsequently, the mean effect size is analyzed concerning fourteen research characteristics, encompassing three dimensions: (a) basic information with two variables, (b) research content with nine variables, and (c) research methodology with three variables. These findings are then compared to evaluate the effect sizes across different research characteristics.

Sample Selection

The researcher employed a four-step research selection process as follows:

Step 1: Identification. The researcher searched for research papers in the Thai Digital Collection (TDC), a database containing research papers and articles from various universities and journals across the country. The search was conducted using only titles and the keyword 'analytical thinking.' This resulted in the identification of 1,912 research papers.

Step 2: Screening. The researcher screened experimental research papers that were full-text and had the 'learning management model' as the independent variable and 'analytical thinking' as the dependent variable. This process was based on the titles; out of 1,272 papers that met the criteria, 640 were excluded.

Step 3: Eligibility. The researcher further refined the selection to include only experimental research papers that had both experimental and control groups, along with reported statistical values essential for calculating effect sizes (*d*). This eligibility assessment was based on abstracts and results, and out of 152 papers that met the criteria, 1,120 were excluded.

Step 4: Inclusion. The researcher selected research papers that aligned with the Basic Education Core Curriculum, emphasizing child-centered and active learning between 2002 and 2021. This selection was based on instrument development, and out of 131 papers that met the criteria, 21 were excluded. This selection process is depicted in Figure 1.

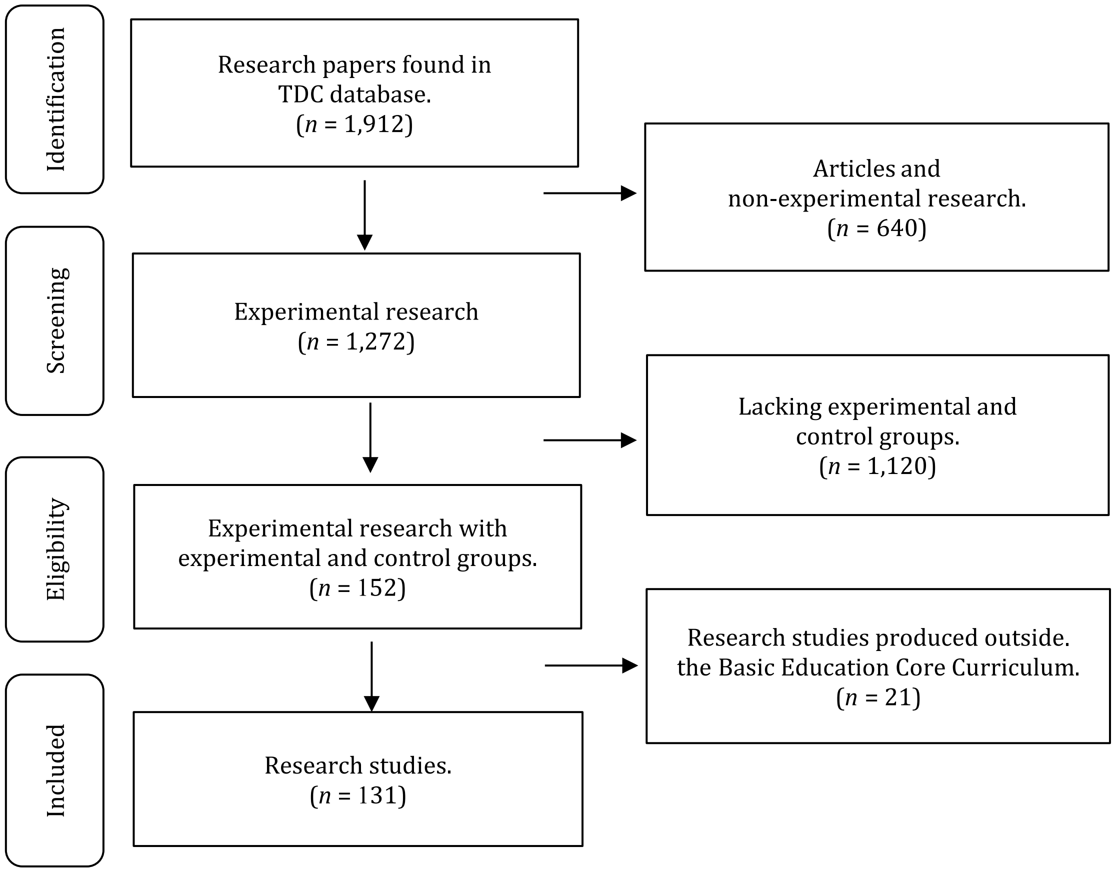


Figure 1. Flow Diagram of Selecting a Sample

Developing and Validating Research Instruments

The researchers developed research characteristics recording forms and research quality assessment forms for data collection. The details of development and quality assurance are as follows:

1. Research characteristics recording forms and code sheets were used to collect basic information, content, and methodology. Content validity was assessed by three experts, revealing an Index of Item-Objective Congruence (IOC) ranging from 0.67 to 1.00 for all items. Kappa statistics were employed to calculate inter-rater reliability from two assessors who evaluated nine research papers, resulting in a value of .915. Intra-rater reliability, assessed by one assessor conducting measurements repeated twice with a one-week interval, yielded a value of .972, indicating a high level of confidence (Czodrowski, 2014).

2. Research quality assessment forms and criteria were used to evaluate titles, backgrounds, literature reviews, methodology, research findings, report writing, and research utility (Ariyasinsomboon, 2001; Kmet et al., 2004). These scoring rubrics consist of five levels, ranging from 0 for low quality to 4 for high quality, comprising 25 items. Content validity was confirmed by three experts, showing an IOC between 0.67 and 1.00 for all items. Kappa statistics were employed to calculate inter-rater reliability from two assessors who evaluated nine research works, yielding a coefficient of .813. Intra-rater reliability, assessed by one assessor conducting measurements repeated twice with a one-week interval, yielded a coefficient of .920, indicating a high level of confidence (Czodrowski, 2014).

Analyzing of Data

In the realm of data analysis, grasping essential statistical concepts is paramount (Borenstein et al., 2021; StataCorp, 2023). These include:

1. Funnel plot: Offers a visual representation of effect sizes across studies.

2. Kendall's Tau: Quantifies the strength of dependence between variables.

3. Egger's Test: Pinpoints bias stemming from small effect sizes.

4. Tau squared (τ^2): Reflects the variance of effect size. When τ^2 equals 0, there is no variability.

5. I squared (I^2): Indicates the level of heterogeneity among studies. I^2 values of 25%, 50%, and 75% signify low, moderate, and high variability respectively.

6. *Q* statistic: Assesses heterogeneity among effect sizes.

7. Chi-square test: Determines the significance of variability among effect sizes.

8. *z* statistic: Gauges the significance of the overall effect size and evaluates the statistical significance of independent variables correlated with the effect size.

9. Independent sample *t*-test: Compares means between two independent groups.

10. Two-way ANOVA: Analyze the impact of two categorical independent variables on a continuous dependent variable.

11. Logistic regression: Examines the relationship between a binary dependent variable and independent variables.

The researcher used JASP version 0.17.2 software for data analysis, conducting it in two parts as follows:

In the first part, meta-analysis and regression meta-analysis were employed to address objective 1, involving examining effect sizes and research characteristics. The research characteristics were categorized into subgroups and converted into dummy variables. This analysis comprised four steps:

Step 1: Publication bias was assessed using a funnel plot, a graph resembling an inverted funnel. The x-axis depicted the effect sizes of each study, while the y-axis represented standard error. Studies with low standard error (indicating a large sample size) clustered in the center of the funnel. Additionally, Kendall's Tau and Egger's Test were performed, with statistically significant results indicating publication bias in the studied research (Harrison et al., 2017; Nakagawa et al., 2022; Sedgwick & Marston, 2015).

Step 2: Effect sizes analysis: Glass's method (1976) was employed to calculate the effect sizes. The values of d = 0.20 – 0.50 were considered small, d = 0.50 – 0.80 as moderate, and d > 0.80 as large. The choice between employing the random effects and fixed effects was determined by assessing the τ^2 and I^2 values. Higher values of these parameters, signifying greater variability among studies, leaned towards opting for the random effects (Borenstein et al., 2021; Rücker et al., 2008).

Step 3: Test selection was based on the omnibus test of model coefficients to examine if the mean effect size are zero, the test of residual heterogeneity to scrutinize if residuals differ from zero (employing random effects estimation if p-value < .05, and fixed effects estimation if p-value > .05), and chi-square to check for heterogeneity by comparing Q values (Berkhout et al., 2024; Borenstein et al., 2021).

Step 4: Regression meta-analysis: z-statistics were used to test if the effect sizes differ from zero. The first variable served as an intercept for comparison. After testing, variables with statistically significant results (p-value < .05) indicated an influence on the effect sizes (Van Houwelingen et al., 2002).

In the second part, Propensity score matching was employed to address objective 2, involving a six-step data analysis process (Bai, 2011; Harris & Horst, 2016; Staffa & Zurakowski, 2018):

Step 1: Researchers use mean effect size as a criterion to divide the effect size group into two categories: the group with effect sizes lower than the mean effect size is termed the low group, and the group with effect sizes higher than the mean effect size is referred to as the high group.

Step 2: Analyze the research characteristic variables that impact the effect size group differently using an independent sample *t*-test.

Step 3: Calculate propensity scores by incorporating the research characteristic variables that impact the effect size group from Step 2 into logistic regression analysis.

Step 4: Stratifying groups, based on propensity scores.

Step 5: Checking covariate balance, using two-way ANOVA. If imbalances persist, step 3.

Step 6: Utilizing propensity scores in meta-analysis and regression meta-analysis, like the approach undertaken for objective 1.

Results

The Results of Exploring Publication Bias

When analyzing the funnel plot (Figure 2), it is found that the effect sizes of each study mostly have a positive influence and are clustered around the center of the triangular funnel, indicating that most studies use a large sample size. Upon examining the effect sizes, it is noted that they are distributed outside the triangular boundary, especially with two studies deviating significantly from the boundary, indicating a likelihood of publication bias occurring. This suspicion finds support in Kendall's Tau value of 0.437 (p-value < .05) and Egger's Test result of 6.915 (p-value < .05). This suggests that the effect sizes have been influenced by publication bias. Therefore, it is not possible to draw conclusions regarding the development of students' analytical thinking.

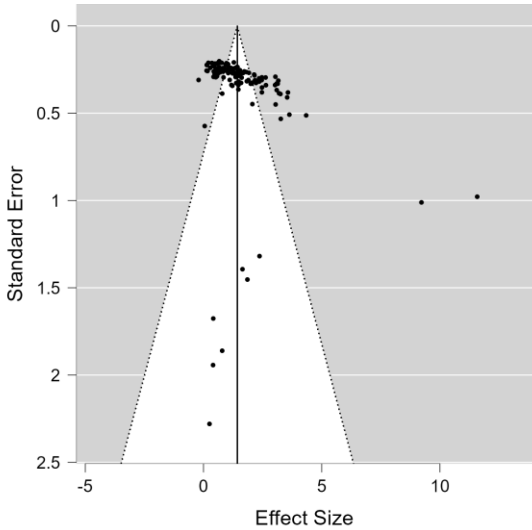


Figure 2. Funnel Plot of the Effect Sizes

The Results of Testing the Effect Sizes

The results of calculating the effect sizes of students' analytical thinking, with a total of 131 values ranging from -0.208 to 11.575, revealed that in the random effects test, $\bar{d}$ = 1.428, SE = 0.077, Upper 95% CI = 1.578 and Lower 95% CI = 1.278. Meanwhile, in the fixed effects test, $\bar{d}$ = 1.192, SE = 0.024, Upper 95% CI = 1.240, and Lower 95% CI = 1.144. The effect sizes of the research exhibited high variability (τ^2 = 0.645, I^2 = 89.136%), leading to the utilization of random effects. After testing the random effects, the mean effect size was found to be 1.428, indicating a significant influence on students' analytical thinking at a high level (d > 0.80). Upon examining the z-test results, signifying a statistically significant deviation from zero and a positive trend, as detailed in Table 1.

Heterogeneity

The heterogeneity test resulted in a Q value of 1207.424 (p-value < .05), indicating that the residuals from estimating the effect sizes are significantly non-zero. As for the chi-square values obtained from the table, with df = 130, the value is χ^2= 157.610, and with df = 132, the value is χ^2 = 159.814. At a significant level of .05, it is evident that the Q statistic is higher than the chi-square value, leading to the rejection of the null hypothesis of zero heterogeneity. With τ^2 = 0.645, and I^2 = 89.136, signifying high variability in the effect sizes, it is apparent that the effect sizes of each study differ significantly. Therefore, random effects should be utilized to ensure unbiased estimation, as detailed in Table 1.

Table 1. Results of Random Effects, Fixed Effects, 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	131	1.428	0.077	1.278	1.578	18.642***	< .001	1207.424***	130	0.645	89.136%
Fixed Effects	131	1.192	0.024	1.144	1.240	48.692***	< .001				

The Results of Testing the Effect Sizes Based on Research Characteristics

When considering the funnel plot and the random effects, it was found that it is not possible to conclude the effect sizes. This is due to interference from research characteristics. Therefore, the researchers investigated the influence of individual variables on research characteristics, as detailed in Table 2.

Table 2. Research Characteristics Influencing Students' Analytical Thinking

Research characteristics	k	$\overline{d}$	SE	95%CI	z-value	Q_a	Q_b	τ^2	I^2
Learning Management Models									
Collaborative Learning	16	1.654	0.550	[0.58,2.73]	6.272***	4.577	1184.007***	0.631	88.852%
Constructivism	19	1.350	0.207	[0.94,1.76]	0.059				
Learning Through Techniques	25	1.364	0.171	[1.03,1.70]	-0.133				
Activity Kit and Media	12	1.734	0.294	[1.16,2.31]	1.260				
Inquiry-Based Learning	39	1.738	0.306	[1.14,2.34]	0.889				
Problem-Based Learning	20	1.204	0.190	[0.83,1.58]	-0.391				
Total	131	1.404	0.095	[1.22,1.59]	14.782***				
Year of publication									
2002 - 2006	10	0.881	0.223	[0.44,1.32]	3.300***	5.044	1168.705***	0.622	88.717%
2007 – 2011	50	1.572	0.248	[1.09,2.06]	2.077*				
2012 – 2016	44	1.599	0.218	[1.17,2.03]	2.003*				
2017 – 2021	27	1.523	0.189	[1.15,1.89]	2.110*				
Total	131	1.396	0.145	[1.11,1.68]	9.601***				
Field of Research									
Learning Management, Elementary and Secondary Education	13	1.003	0.168	[0.68,1.33]	4.337***	8.347*	1145.258***	0.608	88.497%
Research, Measurement, Evaluation, and Educational Psychology	20	1.183	0.220	[0.75,1.61]	0.803				
Curriculum and Instruction	60	1.608	0.244	[1.13,2.09]	1.665				
Science and Technology Education Teaching	38	1.729	0.153	[1.43,2.03]	2.649**				
Total	131	1.380	0.163	[1.06,1.70]	8.466***				
Courses									
Mathematics	33	1.211	0.152	[0.91,1.51]	7.933***	3.444	1184.894***	0.632	88.854%
Thai Language	11	1.833	0.784	[0.30,3.37]	0.619				
Science	58	1.694	0.218	[1.27,2.12]	1.793				
Social Studies, Religion, and Culture	14	1.352	0.194	[0.97,1.73]	0.417				
Career and Technology, Health and Physical, Arts, Foreign Languages, Activities	15	1.438	0.231	[0.99,1.89]	0.890				
Total	131	1.383	0.095	[1.20,1.57]	14.592***				
Level									
Primary School	40	1.168	0.154	[0.87,1.47]	8.378***	5.868	1170.424***	0.623	88.763%
Junior High School	54	1.747	0.261	[1.24,2.26]	2.163*				
Senior High School	37	1.563	0.160	[1.25,1.88]	2.070*				
Total	131	1.440	0.135	[1.18,1.71]	10.651***				
Duration per Plan									
Hour	27	1.020	0.145	[0.74,1.30]	6.374***	10.546*	1155.874***	0.614	88.610%
2 hours	55	1.808	0.265	[1.29,2.33]	2.718**				
3 hours	39	1.364	0.141	[1.09,1.64]	1.222				
More than 3 hours	10	1.872	0.311	[1.26,2.48]	2.527*				
Total	131	1.423	0.157	[1.11,1.73]	9.046***				
Total Duration									
1 - 12 hours	30	1.376	0.154	[1.07,1.68]	8.819***	2.340	1201.216***	0.642	89.028%
13 - 16 hours	48	1.782	0.297	[1.20,2.36]	1.004				
17 - 20 hours	41	1.364	0.166	[1.04,1.69]	-0.288				
More than 20 hours	12	1.346	0.182	[0.99,1.70]	-0.179				
Total	131	1.403	0.091	[1.22,1.58]	15.370***				

Table 2. Continued

Research characteristics	k	$\bar{d}$	SE	95%CI	z-value	Q_a	Q_b	τ^2	I^2
Learning Management Process									
Introduction, Teach, Practice, and Summarize	28	1.550	0.177	[1.20,1.90]	9.351***	7.963	1159.873***	0.617	88.590%
Generate Interest, Teach, Practice, Assess, and Reward	16	1.654	0.550	[0.58,2.73]	-0.709				
Define the problem, Solution criteria, Solution research, Pick a solution, Create, Run, and inspect the solution, and Reflect on a solution	13	0.882	0.244	[0.40,1.36]	-2.152*				
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	17	1.849	0.297	[1.27,2.43]	0.866				
Encounter problems, Stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	26	1.415	0.172	[1.08,1.75]	-0.569				
Engage, Explore, Explain, Elaborate, and Evaluate.	20	1.701	0.546	[0.63,2.77]	-0.310				
Choose a topic, Search, Plan, Execute, Present, and Assessment	11	1.394	0.191	[1.02,1.77]	-0.488				
Total	131	1.427	0.096	[1.24,1.62]	14.798***				
Student Learning Process									
No Grouping and Discussion	14	1.690	0.321	[1.06,2.32]	7.267***	3.170	1184.048***	0.631	88.875%
Grouping and Discussion	18	1.655	0.606	[0.47,2.84]	-1.332				
Grouping, Discussion, and Presentation	67	1.369	0.107	[1.16,1.58]	-1.476				
Grouping, Discussion, Presentation, and Reinforcement	32	1.678	0.300	[1.09,2.27]	-0.617				
Total	131	1.435	0.095	[1.25,1.62]	15.112***				
Learning Media									
Song/Storytelling/Video. (Use 1 type)	10	1.339	0.205	[0.94,1.74]	4.908***	8.873*	1186.448***	0.633	88.892%
Worksheet/Quiz (Use 1 type)	17	1.841	0.306	[1.24,2.44]	1.166				
Blend 2 types of learning media	56	1.771	0.259	[1.26,2.28]	0.674				
Integrate learning from more than 2 types of media	48	1.146	0.108	[0.93,1.36]	-0.674				
Total	131	1.413	0.137	[1.15,1.68]	10.323***				

Table 2. Continued

Research characteristics	k	$\bar{d}$	SE	95%CI	z-value	Q_a	Q_b	τ^2	I^2
Measurement and Evaluation									
Evaluating behavior	24	1.178	0.161	[0.86,1.49]	6.663***	3.906	1192.583***	0.637	88.950%
Posttest, Checking assignments	56	1.449	0.127	[1.20,1.70]	1.220				
Posttest, Checking assignments, Evaluating work	25	2.044	0.545	[0.98,3.11]	1.974*				
Posttest, Checking assignments, Evaluating behavior	26	1.477	0.186	[1.11,1.84]	1.095				
Total	131	1.391	0.087	[1.22,1.56]	16.038***				
Research Design									
Randomized control group pretest-posttest design	84	1.591	0.141	[1.31,1.87]	15.959***	3.043	1187.265***	0.633	88.912%
Non-Randomized control group pretest-posttest design	36	1.221	0.147	[0.93,1.51]	-1.743				
Non-Equivalent control group pretest-posttest design	11	1.932	0.971	[0.03,3.84]	-0.395				
Total	131	1.419	0.115	[1.19,1.64]	12.375***				
Research Statistics									
ANOVA, MANOVA, ANCOVA, MANCOVA	50	1.213	0.126	[0.97,1.46]	9.676***	3.848	1196.905***	0.639	89.026%
t–test Independent Sample	81	1.707	0.188	[1.34,2.08]	1.962				
Total	131	1.417	0.165	[1.09,1.74]	8.608***				
Research Quality									
Moderate level	14	2.605	0.612	[1.41,3.80]	9.826***	16.073***	1146.360***	0.609	88.532%
Good level	84	1.455	0.152	[1.16,1.75]	-3.548***				
Excellent level	33	1.218	0.154	[0.92,1.52]	-3.932***				
Total	131	1.412	0.154	[1.11,1.71]	9.191***				

Note: k = Sample Size, $\bar{d}$ = Mean Effect Size, SE = Standard Error, CI = Confidence Interval, Q_a = Omnibus Test of Model Coefficients, Q_b = Test of Residual Heterogeneity, *** $p < .001$, ** $p < .01$, * $p < .05$

From Table 1, the effect sizes of the 131 research studies, with a mean effect size of 1.428, were analyzed. Research on learning management models significantly influences students' analytical thinking at a high level ($d > 0.80$). When considering Table 2, it becomes evident that all research characteristics exhibit τ^2 and I^2 values indicative of a high level of variability ($\tau^2 > 0$, $I^2 > 75$), underscoring substantial fluctuations in effect sizes across studies. This variability is attributed to various research characteristics. Seven statistically significant research characteristics influence students' analytical thinking at the level of .05, including year of publication, field of research, level, duration per plan, learning management process, measurement and evaluation, and research quality. It is evident that the effect sizes are still influenced by research characteristics. Therefore, it is not possible to conclude the development of students' analytical thinking. It is necessary to eliminate the influence of research characteristics first to obtain clearer conclusions. Hence, the researchers proceeded to analyze the results of adjusting the effect sizes using propensity score matching, as outlined in objective 2.

Propensity Score Matching

The researchers divided the data into two groups based on the mean effect size ($\bar{d}$ = 1.428): the low-effect size group ($\bar{d}$ < 1.428) and the high-effect size group ($\bar{d}$ > 1.428), as shown in Table 3. Subsequently, the data were analyzed using propensity score matching, and the propensity scores for both groups ranged from 0.06338 to 0.78071. The scores exhibited a similar distribution, allowing the researchers to define three continuous score ranges: Q_1 = 0.06338–0.30249, Q_2 = 0.30250–0.54160, and Q_3 = 0.54161–0.78071, as illustrated in Figure 3. This was done to examine the initial agreement in the two-way ANOVA. Considering the close similarity in the data size, propensity score matching was then applied to both sample groups for comparison and ensuring 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	86	131	0.833	0.414
	High	45		2.828	1.798
Q_1	Low	51	64	0.192	0.066
	High	13		0.226	0.039
Q_2	Low	25	43	0.397	0.046
	High	18		0.422	0.048
Q_3	Low	10	24	0.618	0.082
	High	14		0.612	0.064

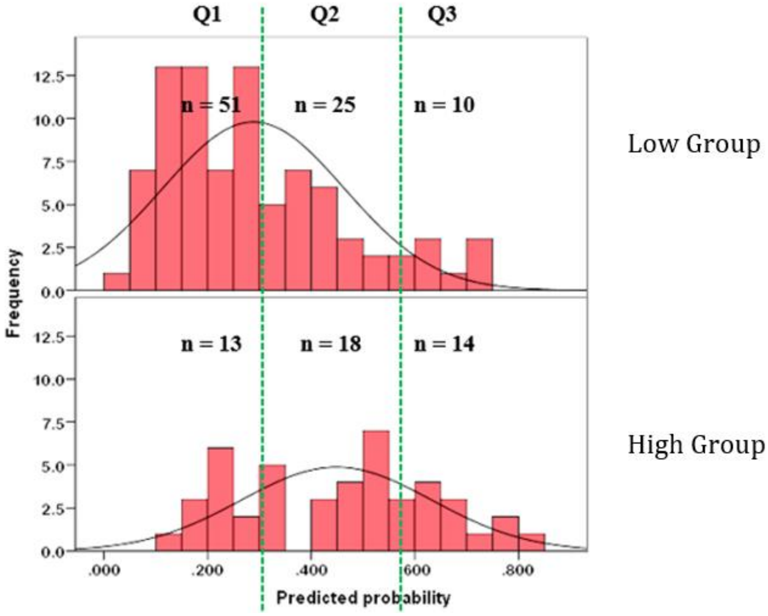


Figure 3. The Results of Comparing the Distribution of Propensity Scores Between the Low and High Effect Sizes Groups

To eliminate the variance of research characteristics influencing the effect sizes, an independent sample *t*-test and Two-way ANOVA will be used. This will provide the effect size after propensity score matching, as detailed in Table 4.

Table 4. Basic Statistics and Comparing of Effect Sizes Before and After Adjustment Using Propensity Score Matching Between Effect Size Groups and Propensity Score Groups

Research characteristics	Effect sizes				Before		After			
	G	*n*	$\bar{d}$	*SD*	*t*-value	*p*-value	*F*-value	*p*-value	*F**-value	*p*-value
Learning Management Models	Low	86	2.779	1.711	0.229	.819	0.096	.757	0.728	.485
	High	45	2.711	1.561						
Year of publication	Low	86	1.628	0.934	-0.817	.416	0.019	.892	0.835	.436
	High	45	1.756	0.802						
Field of research	Low	86	1.814	0.939	-2.269*	.025	1.256	.265	1.715	.184
	High	45	2.178	0.834						
Courses	Low	86	1.756	1.246	0.094	.925	1.788	.184	2.271	.107
	High	45	1.733	1.321						
Level	Low	86	0.872	0.779	-2.250*	.027	0.089	.766	1.447	.239
	High	45	1.178	0.716						
Duration per Plan	Low	86	2.279	1.411	0.059	.953	1.002	.319	0.660	.519
	High	45	2.267	0.986						
Total Duration	Low	86	17.233	0.806	1.586	.115	1.824	.179	0.352	.704
	High	45	15.467	0.768						
Learning Management Process	Low	86	7.233	5.944	1.066	.289	0.400	.528	1.408	.248
	High	45	6.133	5.421						

Table 4. Continued

Research characteristics	Effect sizes				Before		After			
	G	***n***	$\bar{d}$	***SD***	***t*-value**	***p*-value**	***F*-value**	***p*-value**	***F*-value***	***p*-value**
Student Learning Process	Low	86	2.140	1.190	0.026	.979	0.081	.777	0.787	.457
	High	45	2.133	1.342						
Learning Media	Low	86	2.233	0.890	2.737**	.007	0.321	.572	1.090	.339
	High	45	1.800	0.842						
Measurement and Evaluation	Low	86	1.395	1.032	-0.148	.883	0.045	.832	0.572	.566
	High	45	1.422	0.965						
Research Design	Low	86	0.523	0.681	2.132*	.035	0.246	.621	0.699	.499
	High	45	0.289	0.549						
Research Statistics	Low	86	1.849	1.467	-0.066	.948	0.004	.950	0.579	.562
	High	45	1.867	1.471						
Research Quality	Low	86	2.683	0.257	1.578	.118	3.604	.060	0.618	.541
	High	45	2.604	0.280						

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

From Table 4, it was found that four research characteristic variables significantly influence the effect size groups with a statistical significance level of .05 after adjusting for effect sizes. It can be observed that the *F*-statistic of the model with interactions from the two-way ANOVA is low and not statistically significant. This suggests that research characteristics no longer significantly influence the effect size, as shown in Figure 4, the funnel plot. After adjusting the effect sizes (Figure 4b), most of the research studies were conducted using large samples. They approached a zero effect size, as observed in the upper part and within the triangular funnel. This allows for clear conclusions regarding the influence of research characteristics on effect size.

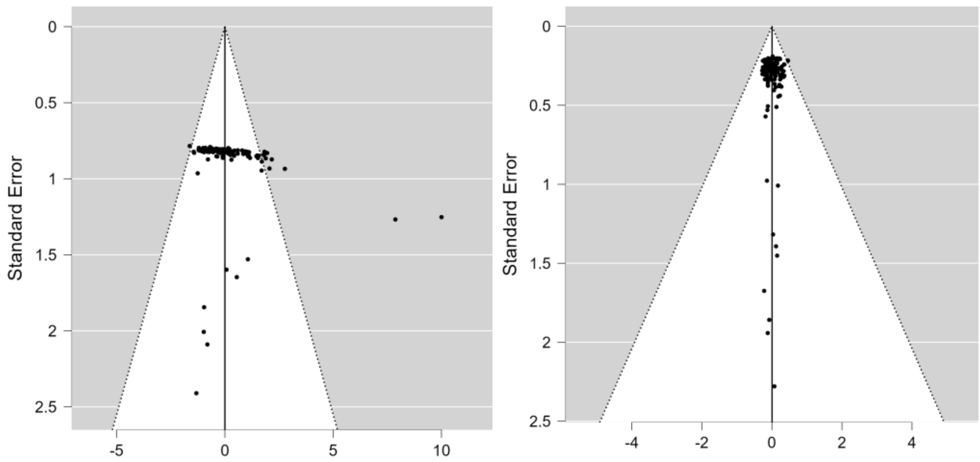


Figure 4a. Before *Figure 4b. After*

Figure 4. Mean Effect Size According to the Learning Management Models That Develop Students' Analytical Thinking Before and After Propensity Score Matching

The Results of Comparing Learning Management Models That Develop Students' Analytical Thinking

After propensity score matching, when testing the type of influence, it was found that the omnibus test of model coefficient yielded a Q value of 2.923 (p-value > .05), indicating that the mean effect size of all studies did not differ significantly from zero. As for the test of residual heterogeneity, a Q value of 45.443 was obtained (p-value > .05), suggesting that the studies did not cause the mean effect size to differ from zero, and the effect sizes for each study did not differ significantly. Therefore, fixed effects model should be employed.

After conducting the fixed effects analysis, it was found that when estimating coefficients using collaborative learning as the baseline for comparison, there were no significant differences among the various learning management models. However, learning through techniques had the greatest influence on students' analytical thinking, followed by activity kits and media, and problem-based learning, in that order, as detailed in Table 5.

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

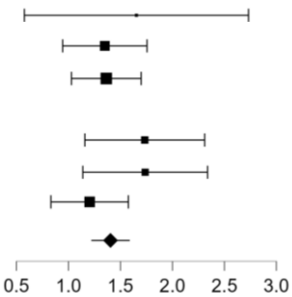
Learning Management Models	*k*	$\bar{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b
Collaborative Learning	16	0.296	0.036	[0.23,0.37]	3.908***	2.923	45.443
Constructivism	19	0.317	0.036	[0.25,0.39]	0.616		
Learning Through Techniques	25	0.415	0.036	[0.35,0.49]	1.597		
Activity Kit and Media	12	0.376	0.046	[0.29,0.47]	0.750		
Inquiry-Based Learning	39	0.318	0.028	[0.26,0.37]	0.672		
Problem-Based Learning	20	0.346	0.038	[0.27,0.42]	0.954		
Total	131	0.340	0.015	[0.31,0.37]	23.436***		

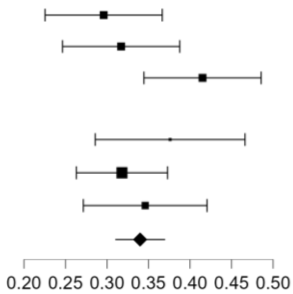
Note: *** $p < .001$, **$p < .01$, * $p < .05$, no values for τ^2 and I^2 due to fixed effects

Before propensity score matching (Table 2), although no statistically significant learning management models influencing students' analytical thinking were found, it was observed that inquiry-based learning had the greatest influence on students' analytical thinking. After propensity score matching, learning through techniques had the greatest influence on students' analytical thinking, as detailed in Table 6.

Table 6. Comparative Results of Learning Management Models Before and After Propensity Score Matching

Learning Management Models	Effect sizes and 95% confidence interval: Before Propensity Score Matching	Effect sizes and 95% confidence interval: After Propensity Score Matching
Collaborative Learning	1.65[0.58,2.73]	0.30[0.23,0.37]
Constructivism	1.35[0.94,1.76]	0.32[0.25,0.39]
Learning Through Techniques	1.36[1.03,1.70]	0.41[0.35,0.49]
Activity Kit and Media	1.73[1.16,2.31]	0.38[0.29,0.47]
Inquiry-Based Learning	1.74[1.14,2.34]	0.32[0.26,0.37]
Problem-Based Learning	1.20[0.83,1.58]	0.35[0.27,0.42]
Total	1.40[1.22,1.59]	0.34[0.31,0.37]





When examining the type of influence in the learning management process, the omnibus test of the model coefficient yielded a Q value of 6.502 (p-value > .05), indicating that the mean effect size of all studies did not differ significantly from zero. As for the test of residual heterogeneity, a Q value of 41.864 was obtained (p-value > .05), suggesting that the studies did not cause the mean effect size to differ from zero, and the effect sizes for each study did not differ significantly. Therefore, fixed effects model should be employed.

After conducting the fixed effects analysis and estimating coefficients using the introduction, teach, practice, and summarize as the baseline for comparison, it was found that engaging, exploring, explaining, elaborating, and evaluating significantly influenced students' analytical thinking at a statistically significant level of .05, as detailed in Table 7.

Table 7. Comparative Results of the Effect Sizes Based on the Learning Management Process

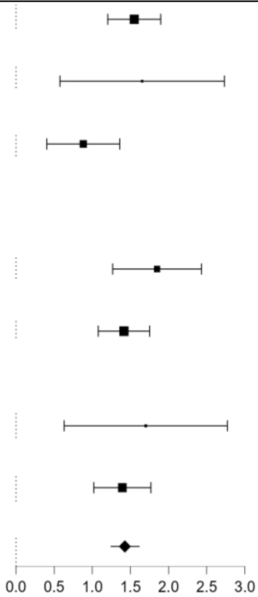
Learning Management Process	*k*	$\bar{d}$	*SE*	95%CI	*z*-value	Q_a	Q_b
Introduction, Teach, Practice, and Summarize	28	0.419	0.036	[0.35,0.49]	7.693***	6.502	41.864
Generate Interest, Teach, Practice, Assess, and Reward	16	0.296	0.036	[0.23,0.37]	-1.783		
Define the Problem, Solution Criteria, Solution Research, Pick a solution, Create, Run, and inspect the solution, and Reflect on the Solution	13	0.325	0.043	[0.24,0.41]	-0.999		
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	17	0.361	0.048	[0.27,0.46]	-0.368		
Encounter problems, stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	26	0.344	0.030	[0.29,0.40]	-1.102		
Engage, Explore, Explain, Elaborate, and Evaluate	20	0.258	0.033	[0.19,0.32]	-2.163*		
Choose a topic, Search, Plan, Execute, Present, and Assessment	11	0.373	0.037	[0.30,0.45]	-0.537		
Total	131	0.336	0.014	[0.31,0.36]	24.443***		

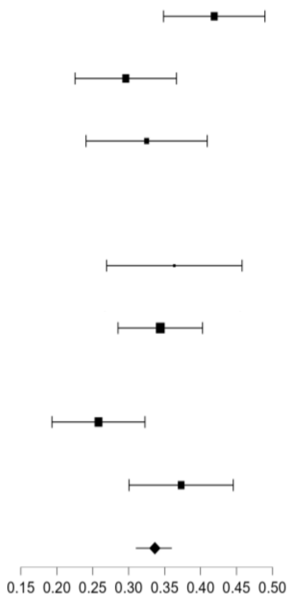
Note: *** $p < .001$, **$p < .01$, * $p < .05$, no values for τ^2 and I^2 due to fixed effects

Before propensity score matching (Table 2), it was found that the learning management process involving elicit, engage, explore, explain, elaborate, evaluate, and extend had the greatest influence on students' analytical thinking. After propensity score matching, it was observed that the introduction, teach, practice, and summarize significantly influenced students' analytical thinking at .05 significance level, as detailed in Table 8.

Table 8. Comparative Results of Learning Management Process Before and After Propensity Score Matching

Learning Management Process	**Effect sizes and 95% confidence interval**	
	Before Propensity Score Matching	**After Propensity Score Matching**
Introduction, Teach, Practice, and Summarize	1.55[1.20,1.90]	0.42[0.35,0.49]
Generate Interest, Teach, Practice, Assess, and Reward	1.65[0.58,2.73]	0.30[0.23,0.37]
Define the problem, Solution criteria, Solution research, Pick a solution, Create, Run, and inspect the solution, and Reflect on a solution	0.88[0.40,1.36]	0.32[0.24,0.41]
Elicit, Engage, Explore, Explain, Elaborate, Evaluate, and Extend	1.85[1.27,2.43]	0.36[0.27,0.46]
Encounter problems, stimulate intellectual conflict, Analyze, Elaborate, and Evaluate	1.42[1.08,1.75]	0.34[0.29,0.40]
Engage, Explore, Explain, Elaborate, and Evaluate.	1.70[0.63,2.77]	0.26[0.19,0.32]
Choose a topic, Search, Plan, Execute, Present, and Assessment	1.39[1.02,1.77]	0.37[0.30,0.45]
Total	1.43[1.24,1.62]	0.34[0.31,0.36]





Discussion

The findings of this research indicate that inquiry-based learning had the most significant impact on students' analytical thinking, as evidenced by diverse studies involving students, teachers, administrators, and various contexts. However, there is a limitation in excluding quality research due to the inability to calculate effect sizes, limited access to unpublished research, and experimental studies lacking experimental and control groups. When considering the funnel plot analysis in this meta-analysis, reveals asymmetry and deviation of effect sizes from the overall mean, indicating publication bias. This bias stems from institutions and fields of education prioritizing research differently. Some institutions and fields still lack rigor in research design, sampling, controlling confounding variables in experimental research, and verifying statistical assumptions, leading to the small-study effect, where small sample groups tend to exhibit higher effect sizes, thus causing differences in research outcomes (Card, 2012; Harbord et al., 2009; Vevea et al., 2019), as well as low research quality with high effect sizes, similar to the studies by Dowdy et al. (2020), Ferguson and Brannick (2012). This is attributed to budget constraints, institutional differences in rigor, negative publication bias, acceptance of positive research outcomes, and statistically significant results, all contributing to Publication Bias (Ahmed et al., 2012; Lin & Chu, 2018; Vevea et al., 2019). This research underscores the critical importance of sample selection in meta-analysis. If publication bias is detected, the conclusions cannot be fully utilized and become a limitation that should be cautiously considered, as they cannot be generalized to the population (Egger et al., 1997; Esterhuizen & Thabane, 2016; Lee, 2019; Lin & Chu, 2018).

Following meta-analysis, it was revealed that the learning management process significantly impacts students' analytical thinking. Upon careful consideration, it is evident that the characteristics of the learning management process create engaging experiences for students, foster understanding, impart knowledge through teaching, understanding through teaching, practice, expand thinking through mutual exchange, and evaluate outcomes. Consequently, students can engage in critical thinking, analyze sub-components, and establish profound connections, enabling them to assess data, identify problem-solving strategies, and make decisions (Rodrangsee et al., 2022; Spaska et al., 2021; Suyatman et al., 2021). However, it was discovered that various research characteristics, including the year of publication, field of research, level, duration per plan, learning management process, measurement and evaluation, and research quality, positively influence the development of students' analytical thinking. This is evidenced in studies exploring the impacts of teaching management from meta-analyses conducted by Itsarangkul Na Ayutthaya and Damrongpanit (2022a), Niu et al. (2013), and Xu et al. (2023). Additionally, researchers identified intriguing issues regarding research characteristics, such as certain variables demonstrating high effect sizes but small sample sizes, including Thai courses, nonequivalent control group pretest-posttest design, and t-test independent sample. Although these components tend to positively influence analytical thinking, their statistical significance is not firmly established due to the limited number of studies examined. Nevertheless, the conclusions drawn from the meta-analysis are still influenced by research characteristics. To address this concern, propensity score matching should be utilized to balance the data, thereby reducing confounding variables and enhancing the clarity and reliability of the conclusions (Austin, 2009; Morgan, 2018).

After propensity score matching, researchers identified three crucial issues. Firstly, when exploring research characteristics (14 variables) that impact effect sizes in groups categorized based on mean effect size as a criterion ($\bar{d}$ = 1.428), it was found that high effect size groups (n = 45) had higher propensity scores than low effect size groups (n = 86). Comparing the distribution of propensity scores by overlapping score ranges revealed a balanced and similar distribution in both groups, allowing for effective propensity score matching to mitigate inequality between the two groups (Badhiwala et al., 2021; Benedetto et al., 2018). Secondly, upon considering the influence of research characteristics following propensity score matching, it was found that propensity score matching completely eradicated the influence of research characteristics. (Rosenbaum & Rubin, 2023). Thirdly, before conducting propensity score matching, a comparison of learning management models developed for students' analytical thinking revealed that the mean effect size of the studies did not differ significantly from zero (Q_a = 4.577), and the residuals from the estimation were not zero (Q_b = 1184.007***). The effect sizes of each study varied significantly (τ^2 = 0.631, I^2 = 88.852), which stemmed from the influence of different research characteristics. Therefore, they cannot be used as research conclusions. After propensity score matching, it was found that the mean effect size did not differ significantly from zero (Q_a = 2.923), and the residuals from the estimate were close to zero (Q_b = 45.443). The effect sizes of each research study were not significantly influenced by research characteristics statistically. Therefore, the adjusted effect sizes can be used in the meta-analysis, as demonstrated in the study by Itsarangkul Na Ayutthaya and Damrongpanit (2022b). Therefore, the conclusion is that learning through techniques such as KWL, KWL-plus, Six Thinking Hats, 4MAT, and Mind Mapping had the greatest influence on students' analytical thinking, with the learning management process encompassing lesson introduction, teaching, hands-on practice, and summarization. These processes enable students to build understanding, stimulate thinking, practice step-by-step thinking, and draw logical conclusions, following Bloom's theory (Bloom, 1956; Krathwohl, 2002), as supported by the research of Sitthipon (2012) and Spaska et al. (2021). However, research indicates that there is no difference in learning management models. The conclusions from the studies of Ramadani et al. (2021), Rodrangsee et al. (2022), and Suyatman et al. (2021) suggest that inquiry-based learning can also develop students' analytical thinking. When considering the learning management process, crucial processes for students' analytical thinking include creating opportunities for students to face problems, pose questions, make predictions, analyze relationships, make connections, verify, and summarize findings from various situations is a method that significantly enhances students' analytical thinking and can be effectively applied in problem-solving in daily life and various situations (Amer, 2005; Elder & Paul, 2007; Rasheva-Yordanova et al., 2018; Robbins, 2011; Sartika, 2018).

Conclusion

In this study, it was found that (a) research related to learning management models significantly influences students' analytical thinking at a high level. Seven research characteristics, including year of publication, field of publication, level, duration per plan, learning management process, measurement and evaluation, and research quality, have a statistically significant influence on students' analytical thinking at the .05 significance level, making it impossible to draw conclusions without using propensity score matching to eliminate the influence of these research characteristics, and (b) after adjusting the effect sizes using propensity score matching, it was found that learning techniques such as KWL, KWL-plus, Six Thinking Hats, 4MAT, and Mind Mapping had the greatest influence on students' analytical thinking. Moreover, the learning management process that provides students with opportunities for fostering understanding, systematic practical training, expand thinking through collaborative exchanges, and assessments will lead to the development of students' analytical thinking.

Recommendations

The researchers have three recommendations for this research. Firstly, to enhance the effectiveness of learning management models, teachers can apply learning thought techniques such as KWL, KWL-plus, Six Thinking Hats, 4MAT, and Mind Mapping. These techniques enable students to foster understanding, systematic practical training, expanding thinking through collaborative exchanges, and assessments, particularly in Thai courses, to develop students' analytical thinking according to the context. Secondly, in controlling confounding variables in research design, conclusions related to confounding variables influence students' analytical thinking to design research and reduce research outcome variability. Thirdly, propensity score matching for meta-analysis should be employed to help reduce publication bias and type I errors in research outcomes, making the research findings more comprehensive.

Recommendations for future research consist of two aspects. Firstly, addressing unclear conclusions: if there is a sufficient sample size in the future, clear conclusions can be drawn. Secondly, meta-analysis should be employed to synthesize research findings from various sources for more precise research outcomes. This should be accompanied by propensity score matching to match sample groups in research studies, reducing the variability of confounding variables, and thus enhancing the completeness of research results.

Limitations

This research encounters two limitations. Firstly, despite the high effect sizes, the small sample size leads to statistically insignificant and unclear conclusions. Secondly, the process of sample selection may not provide access to quality research or unpublished studies.

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

Adelana, O. P., Ishola, A. M., & Adeeko, O. (2021). Development and validation of instructional package for teaching and learning of genetics in secondary schools. *Asian Journal of Assessment in Teaching and Learning*, *11*(2), 32-41. https://doi.org/10.37134/ajatel.vol11.2.4.2021

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

Ahmed, I., Sutton, A. J., & Riley, R. D. (2012). Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: A database survey. *BMJ*, *344*, Article d7762. https://doi.org/10.1136/bmj.d7762

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

Akkerman, S., Admiraal, W., Brekelmans, M., & Oost, H. (2008). Auditing quality of research in social sciences. *Quality and Quantity*, *42*, 257-274. https://doi.org/10.1007/s11135-006-9044-4

Amer, A. (2005). *Analytical thinking*. Center for Advancement of Postgraduate Studies and Research in Engineering Sciences, Faculty of Engineering- Cairo University. https://bit.ly/3TuvErD

Ariyasinsomboon, W. (2001). *การสังเคราะห์งานวิจัยในสาขาจิตวิทยาการศึกษา : การวิเคราะห์อภิมาน* [Synthesis of research in the field of educational psychology: A meta-analysis] [Doctoral dissertation, Chulalongkorn University]. Chulalongkorn University Intellectual Repository. https://bit.ly/3W2zB8v

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

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals.* Longmans.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (Eds.). (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons Ltd.

Burch, G. F., Giambatista, R., Batchelor, J. H., Burch, J. J., Hoover, J. D., & Heller, N. A. (2019). A meta-analysis of the relationship between experiential learning and learning outcomes. *Decision Sciences Journal of Innovative Education*, *17*(3), 239-273. https://doi.org/10.1111/dsji.12188

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

Clark, K. R. (2018). Learning theories: Constructivism. *Radiologic Technology*, *90*(2), 180-182. https://bit.ly/3SafIsq

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

Cornell, D. (2024, January 3). *17 Analytical thinking examples*. Helpful Professor. https://bit.ly/3S3lImV

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

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

Elder, L., & Paul, R. (2007). *The thinker's guide to analytic thinking*. The Foundation for Critical Thinking. https://bit.ly/4bOFcE1

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

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

Few, S. (2015, November). *A course of study in analytical thinking*. Perceptual Edge. https://bit.ly/3yz8SGT

Gillies, R. M. (2014). Cooperative learning: Developments in research. *International Journal of Educational Psychology*, *3*(2), 125-140. http://dx.doi.org/10.4471/ijep.2014.08

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

González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of education 4.0 in 21st century skills frameworks: Systematic review. *Sustainability*, *14*(3), Article 1493. https://doi.org/10.3390/su14031493

Harbord, R. M., Harris, R. J., & Sterne, J. A. (2009). Updated tests for small-study effects in meta-analyses. *The Stata Journal*, *9*(2), 197-210. https://doi.org/10.1177/1536867X0900900202

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, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2017). Publication bias in strategic management research. *Journal of Management*, *43*(2), 400-425. https://doi.org/10.1177/0149206314535438

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

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

Kao, C.-Y. (2014). Exploring the relationships between analogical, analytical, and creative thinking. *Thinking Skills and Creativity*, *13*, 80-88. https://doi.org/10.1016/j.tsc.2014.03.006

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

Khairani, Z., Nasution, D., & Bukit, N. (2021). Analysis of science process skills using learning cycle 7e. *Journal of Physics: Conference Series*, *1811*, Article 012085. https://doi.org/10.1088/1742-6596/1811/1/012085

Kmet, L. M., Cook, L. S., & Lee, R. C. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. Alberta Heritage Foundation for Medical Research. https://bit.ly/3JuWIAU

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

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, 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

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, *74*(3), 785-794. https://doi.org/10.1111/biom.12817

Moallem, M., Hung, W., & Dabbagh, N. (Eds.). (2019). *The Wiley handbook of problem-based learning*. John Wiley & Sons. 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

Morrison, G. R., Ross, S. J., Morrison, J. R., & Kalman, H. K. (2019). *Designing effective instruction*. John Wiley & Sons.

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

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*(1), 4-21. https://doi.org/10.1111/2041-210X.13724

Nicol, C. B., Gakuba, 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

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5), 1189-1208. https://bit.ly/3RrxWpb

Ramadani, A. S., Supardi, Z. A. I., Tukiran, & Hariyono, E. (2021). Profile of analytical thinking skills through inquiry-based learning in science subjects. *Studies in Learning and Teaching*, *2*(3), 45-60. https://doi.org/10.46627/silet.v2i3.83

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

Rasheva-Yordanova, K., Iliev, E., & Nikolova, B. (2018). Analytical thinking as a key competence for overcoming the data science divide. In L.G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *EDULEARN18 Proceedings: 10th International Conference on Education and New Learning Technologies* (pp. 7892-7898). IATED Academy. https://doi.org/10.21125/edulearn.2018.1833

Robbins, J. K. (2011). Problem solving, reasoning, and analytical thinking in a classroom environment. *The Behavior Analyst Today*, *12*(1), 41-48. https://doi.org/10.1037/h0100710

Rodrangsee, B., Tuntiwongwanich, S., Pimdee, P., & Moto, S. (2022). Development of an online active learning model using the theory of multiple intelligence to encourage Thai undergraduate student analytical thinking skills. *Journal of Higher Education Theory and Practice*, *22*(12), 63-75. https://doi.org/10.33423/jhetp.v22i12.5463

Rosenbaum, P. R., & Rubin, D. B. (2023). Propensity scores in the design of observational studies for causal effects. *Biometrika*, *110*(1), 1-13. https://doi.org/10.1093/biomet/asac054

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127*(8), 757-763. https://doi.org/10.7326/0003-4819-127-8_Part_2-199710151-00064

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

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*, 77-112. https://doi.org/10.1007/s10648-019-09498-w

Sartika, S. B. (2018). Teaching models to increase students' analytical thinking skills. In T. W. Abadi, N. E. T. I. Santosa, M. T. Multazam, F. Megawati (Eds.), *Proceedings of the 1st International Conference on Intellectuals' Global Responsibility* (pp. 216-218). Atlantis Press. https://doi.org/10.2991/icigr-17.2018.52

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

Sitthipon, A.-I. (2012). Development of teachers' learning management emphasizing on analytical thinking in Thailand. *Procedia-Social and Behavioral Sciences*, *46*, 3339-3344. https://doi.org/10.1016/j.sbspro.2012.06.063

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

Smith, P. L., & Ragan, T. J. (2005). A framework for instructional strategy design. In P. L. Smith, & T. J. Ragan (Eds.), *Instructional design* (3rd ed., pp. 127-150). Wiley & Sons. https://bit.ly/3U79n28

Spaska, A. M., Savishchenko, V. M., Komar, O. A., Hritchenko, T. Y., & Maidanyk, O. V. (2021). Enhancing analytical thinking in tertiary students using debates. *European Journal of Educational Research, 10*(2), 879-889. https://doi.org/10.12973/eu-jer.10.2.879

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

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

Suyatman, Saputro, S., Sunarno, W., & Sukarmin. (2021). The implementation of research-based learning model in the basic science concepts course in improving analytical thinking skills. *European Journal of Educational Research*, *10*(3), 1051-1062. https://doi.org/10.12973/eu-jer.10.3.1051

Tan, O.-S. (2021). *Problem-based learning innovation: Using problems to power learning in the 21st century*. Gale Cengage Learning. https://bit.ly/3SYB9yw

Thaneerananon, T., Triampo, W., & Nokkaew, A. (2016). Development of a test to evaluate students' analytical thinking based on fact versus opinion differentiation. *International Journal of Instruction, 9*(2), 123-138. https://doi.org/10.12973/iji.2016.929a

Thoemmes, F. (2012). *Propensity score matching in SPSS*. arXiv. https://doi.org/10.48550/arXiv.1201.6385

Triantafyllou, S. A. (2022). Constructivist learning environments. In *Proceedings of the 5th International Conference on Advanced Research in Teaching and Education* (pp. 1-6). Diamond Scientific Publishing. https://doi.org/10.33422/5th.icate.2022.04.10

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., & Sutton, A. (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

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need, and what we can do about it*. Basic Books.

World Economic Forum. (2023). *Future jobs report 2023*. https://bit.ly/3Hqg8Gn

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