AL IBTIDA: JURNAL PENDIDIKAN GURU MI (2025) VOL 12 (2): 387- 411
DOI: http://dx.doi.org/10.24235/al.ibtida.snj.v12i2.22002

Al Ibtida: Jurnal Pendidikan Guru MI
ISSN: 2442-5133, e-ISSN: 2527-7227
Journal homepage: http://syekhnurjati.ac.id/jurnal/index.php/ibtida
Journal email: alibtida@syekhnurjati.ac.id

The Effect of Cooperative Learning Model on Elementary School Students' Affective Learning Outcomes: A Meta-Analysis

Ma'as Shobirin*
*Department of Madrasah Ibtidaiyah Teacher Education, Faculty of Islamic Religion, Universitas Wahid Hasyim Semarang, Indonesia
E-mail: maas.shobirin@unwahas.ac.id

Fitria Martanti**
**Department of Madrasah Ibtidaiyah Teacher Education, Faculty of Islamic Religion, Universitas Wahid Hasyim Semarang, Indonesia
E-mail: fitriamartanti@unwahas.ac.id

Ummu Jauharin Farda***
***Department of Madrasah Ibtidaiyah Teacher Education, Faculty of Islamic Religion, Universitas Wahid Hasyim Semarang, Indonesia
E-mail: ummujauharinfarda@unwahas.ac.id

Ersila Devy Rinjany****
****Department of Madrasah Ibtidaiyah Teacher Education, Faculty of Islamic Religion, Universitas Wahid Hasyim Semarang, Indonesia
E-mail: ersiladevyrinjani@unwahas.ac.id

Received: August 12th, 2025. Accepted: September 25th, 2025. Published: October 31st, 2025.

Abstract

This meta-analysis aims to examine the effect of cooperative learning on elementary school students' affective learning outcomes after the implementation of the cooperative learning model. 40 relevant studies, through online databases, including Google Scholar, DOAJ, ERIC, Springer, Elsevier, and ResearchGate, were identified. Of these, 12 studies met the inclusion criteria and provided usable data, yielding 16 artifacts. Data from 16 artifacts, including sample sizes, means, and standard deviations, were collected using descriptive analysis. The data analysis utilized meta-analysis techniques, including effect size calculation, heterogeneity testing, summary effect estimation using a random-effects model, and the assessment of potential publication bias through forest plots. Those analyses utilized JASP software. The results showed that: (1) The effect size (ES) for affective learning outcomes of elementary school students was 77.06, with a 95% confidence interval (CI) ranging from 73.34 to 81.77. The hypothesis test produced a z value of 0.4035, which is smaller than the critical value ($z = 1.96$), leading to the acceptance of the null hypothesis (H_0) and the rejection of the alternative hypothesis (H_a); (2) Although there is a slight asymmetry in the funnel diagram indicating potential publication bias, the statistically insignificant Egger regression test ($z = -0.1078$, $p = 0.914$) supports the robustness of the meta-analysis, which confirms that the cooperative learning model produces moderate to high affective learning outcomes in elementary school students. This finding indicates that although cooperative learning can positively influence students' affective learning outcomes, the observed effect size is not statistically significant at $\alpha = 0.05$. Thus, it is recommended that future studies refine the design and implementation of cooperative learning strategies to increase their effectiveness and achieve statistically significant improvements in elementary school students' affective learning outcomes.

Keywords: *affective learning outcomes, cooperative learning, meta-analysis.*

Abstrak

Meta-analisis ini bertujuan untuk menguji pengaruh pembelajaran kooperatif terhadap hasil belajar afektif siswa sekolah dasar setelah penerapan model pembelajaran kooperatif. Sebanyak 40 studi relevan, melalui basis data daring, termasuk Google Scholar, DOAJ, ERIC, Springer, Elsevier, dan ResearchGate, diidentifikasi. Dari jumlah tersebut, 12 studi memenuhi kriteria inklusi dan menyediakan data yang dapat digunakan, menghasilkan 16 artefak. Data dari 16 artefak, termasuk ukuran sampel, rata-rata, dan deviasi standar, dikumpulkan menggunakan analisis deskriptif. Analisis data menggunakan teknik meta-analisis, termasuk perhitungan ukuran efek, pengujian heterogenitas, estimasi efek ringkasan menggunakan model efek acak, dan penilaian potensi bias publikasi melalui plot hutan. Analisis tersebut menggunakan perangkat lunak JASP. Hasil penelitian menunjukkan bahwa: (1) Besarnya efek (ES) terhadap hasil belajar afektif siswa sekolah dasar adalah 77,06, dengan interval kepercayaan (CI) 95% berkisar antara 73,34 hingga 81,77. Uji hipotesis menghasilkan nilai z sebesar 0,4035, yang lebih kecil daripada nilai kritis (z = 1,96), sehingga menyebabkan diterimanya hipotesis nol (H_0) dan ditolaknya hipotesis alternatif (H_a); (2) Meskipun terdapat sedikit asimetri pada diagram corong yang mengindikasikan potensi bias publikasi, uji regresi Egger yang secara statistik tidak signifikan (z = -0,1078, p = 0,914) mendukung kekokohan meta-analisis, yang menegaskan bahwa model pembelajaran kooperatif menghasilkan hasil belajar afektif sedang hingga tinggi pada siswa sekolah dasar. Temuan ini menunjukkan bahwa meskipun pembelajaran kooperatif dapat memengaruhi hasil belajar afektif siswa secara positif, besarnya efek yang teramati tidak signifikan secara statistik pada $\alpha = 0{,}05$. Oleh karena itu, disarankan agar penelitian masa depan menyempurnakan desain dan implementasi strategi pembelajaran kooperatif untuk meningkatkan efektivitasnya dan mencapai peningkatan yang signifikan secara statistik pada hasil belajar afektif siswa sekolah dasar.

Kata kunci: *hasil pembelajaran afektif, pembelajaran kooperatif, meta-analisis.*

INTRODUCTION

Cooperative learning (CL) stands as a dynamic pedagogical strategy, deeply rooted in social interdependence theory, where students' successes and growth are intricately tied to their interactions with peers, fostering a rich tapestry of positive interdependence, individual accountability, promotive interaction, social skills, and group processing (Colomer et al., 2021; Johnson & Johnson, 2018). This approach draws from foundational motivational perspectives that champion group rewards to spark achievement and engagement, as highlighted by recent explorations into student-driven incentives (Bardach & Murayama, 2025; Hellín et al., 2023), while cognitive-developmental theories, inspired by the likes of Piaget and Vygotsky, emphasize peer collaborations as catalysts for cognitive growth through conflict resolution and scaffolding (Cade, 2023; Chen, 2025). The structural framework further enhances this model by embedding specific techniques to ensure equitable participation and simultaneous engagement, nurturing both social competence and academic success (Kagan, 2021). Tracing its origins to the 1960s and 1970s, historical analyses distinguish CL from collaborative learning by its structured design and focus on K-12 education, laying a robust theoretical groundwork for diverse models that prioritize collaboration over competition (Yang, 2023). These principles form the bedrock of an adaptable educational philosophy, supporting critical thinking and teamwork across various learning stages.

Over the past decade (2015–2025), a growing body of research has shed light on the remarkable effectiveness of cooperative learning (CL) across varied educational settings, significantly boosting student engagement, learning outcomes, and social skills, positioning it as a guiding light for innovative teaching practices (Boke et al., 2025). Systematic reviews highlight its contributions to cultural diversity, individual accountability, and alignment with sustainable development goals by embracing student-centered learning environments (Zhou & Colomer, 2024), while meta-analyses demonstrate moderate yet meaningful positive effects across affective, cognitive, physical, and social domains, with strategies like Learning Together and Group Investigation leading the way (Boke et al., 2025). In higher education physics, CL has proven its worth by enhancing academic achievement and attitudes through a mix of experimental and conceptual tasks (Kilpeläinen-Pettersson et al., 2025), and bibliometric analyses reveal a notable uptick in research interest since 2021, focusing on motivation, cognitive skills, and communication in science education (I. Damayanti et al., 2025; Xue et al., 2025). Among its standout models, the jigsaw approach empowers students to master subtopics and teach peers, driving notable gains in academic performance and social skills, with meta-analyses confirming moderate to large effects, particularly in secondary and higher education (Bhardwaj, 2025; Cochon Drouet et al., 2023). Student Teams-Achievement Divisions (STAD) taps into group rewards linked to individual progress to boost motivation and cognitive outcomes in physics (Aprilia & Dwandaru, 2025), while Teams-Games-Tournament (TGT) introduces friendly competition to enhance activeness and participation in elementary classrooms (Saputri & Lestari, 2025).

Comparisons underscore Jigsaw's strength in problem-solving and retention, contrasting with STAD's motivational edge (Damayanti et al., 2023), setting these methods apart from traditional approaches through their focus on interdependence and accountability. Other models, such as Make a Match, pair students to link questions with answers or concepts with definitions, sparking active engagement, peer teaching, and quick problem-solving—especially effective in elementary settings for reinforcing knowledge and fostering social connections (Masrita, 2017). Group Investigation (GI) encourages students to design and lead inquiries, nurturing higher-order thinking and supporting low- to middle-achievers with improved cognitive and behavioral engagement (Silva et al., 2023), while Team Assisted Individualization (TAI) merges group support with individual tasks to enhance mathematical problem-solving, communication, and self-efficacy in operations research (Tinungki et al., 2022). Numbered Heads Together (NHT) ensures fair participation through numbered responses, boosting critical thinking in physics (Aprilia & Dwandaru, 2025), and Think-Pair-Share (TPS) promotes equitable dialogue via reflection, pairing, and sharing, enriching participation across educational contexts (Guenther & Abbott, 2024; Latifah & Aviya, 2018). Technology-infused models like Cooperative Mobile Learning, using tools such as Google

Docs, ease cognitive loads and elevate outcomes in elementary science, supported by meta-analyses showing enhanced performance and motivation in pre-college settings (Huang et al., 2020a; Sergeeva et al., 2025). Cooperative Problem-Based Learning (CPBL) blends CL with problem-solving cycles to build team skills, addressing group dynamics through training (Zaafour & Salaberri Ramiro, 2022), while Two Stay Two Stray (TSTS) boosts science skills and motivation through information sharing (Damayanti et al., 2025). Collectively, these models—from achievement-driven Jigsaw and STAD to inquiry-focused GI and TAI and accessible tech-enhanced CPBL—form a unified framework grounded in interdependence and accountability (Yang, 2023), with recent studies (2015–2025) validating their ability to foster affective and cognitive growth, offering viable alternatives to traditional methods despite challenges like time and resources, and nurturing inclusive, reflective learning (Boke et al., 2025; Zhou & Colomer, 2024).

Recent scholarship from 2015 to 2025 has increasingly explored how cooperative learning (CL) models nurture the affective dimensions of elementary students—spanning motivation, emotional well-being, social skills, and learning satisfaction—by fostering positive interdependence and meaningful peer interactions. These dynamics help reduce stress, sharpen emotional regulation, and boost engagement across diverse areas like physical education and language learning, with a meta-analysis of 40 studies reporting a moderate effect size of 0.304 across primary levels, enhancing the classroom atmosphere and providing emotional support (Boke et al., 2025). In physical education settings, research has revealed serial mediation effects, where perceived cooperation enhances academic performance by alleviating negative emotions such as anxiety and boredom while amplifying positive ones like enjoyment and pride (León et al., 2023). This growing body of work positions CL as a resilient approach, particularly in culturally diverse classrooms, where it mitigates emotional challenges and encourages self-reflection (Zhou & Colomer, 2024), offering modest to moderate affective gains as a refreshing, student-centered alternative to traditional methods through its structured group dynamics.

Specific CL models, including Jigsaw, Student Teams-Achievement Divisions (STAD), Teams-Games-Tournament (TGT), and Think-Pair-Share (TPS), have been thoroughly investigated for their impact on affective growth, distinguished by their emphasis on structured accountability. Jigsaw stands out by cultivating collaboration, peer connections, and emotional intelligence, with notable pre- to post-test improvements in third-grade learning and behaviors (Putri & Pratiwi, 2025) and positive effects on motivation and self-perception (Mubayinah, 2023). STAD and TGT, widely used in physical education, help reduce negative emotions and strengthen social skills, a finding backed by meta-analytic evidence (Boke et al., 2025), while TPS variants and technology-enhanced approaches like Quizlet boost emotional adaptability and peer-related outcomes (Lin et al., 2025). This cumulative evidence from 2015 to 2025 presents a robust perspective on CL's influence through absolute post-intervention gains, highlighting models like Jigsaw and STAD for achieving emotionally supportive benchmarks—such as enhanced motivation and social-emotional learning—that support long-term development (Li, 2025; Rivera-Pérez et al., 2020), empowering educators to build inclusive, resilient classrooms (B. Chen, 2023).

To unpack these effects, we employ a means-based meta-analysis, a thoughtful approach that synthesizes post-intervention means from experimental groups as the key effect size metric, focusing on absolute performance rather than relative shifts—a lifeline in educational research where it shines by leveraging continuous means, standard deviations, and sample sizes to gauge intervention impact (Borenstein et al., 2009). This method proves invaluable when control or pre-test data are scarce, relying on standardized post-test scores—often normalized to a 0–100 scale—for fair comparisons (Lipsey & Wilson, 2001), as seen in recent studies like Yoon (2023) that highlight its clarity in assessing outcomes like achievement or affective growth. Unlike contrast-based methods—pre-post analyses tracking mean changes over time or group comparisons between experimental and control sets, which demand full datasets and may exclude studies (Cooper et al., 2009; Hedges & Olkin, 1987)—means-based analysis embraces single-group data, as evidenced by cooperative learning studies showing varied effect sizes without contrasts (Yoon, 2023). Distinct from correlation meta-analysis, which explores variable relationships like study habits and grades (Schmidt & Hunter, 2015), this approach zeroes in on post-intervention magnitude, offering educators tangible benchmarks, such as average scores in CL settings (Boke et al., 2025), and adapts seamlessly to diverse tools, enhancing its relevance in education and psychology (Putri & Pratiwi, 2025).

Traditional meta-analyses often lean on group contrast or pre-post designs, calculating effect sizes like Cohen's d to reveal relative gains in achievement and attitudes (Pellegrini et al., 2021; Ridwan et al., 2022), with examples showing small to medium affective effects (Boke et al., 2025) and moderate overall impacts across 23 studies (Öztürk, 2023). Yet, their reliance on complete multi-group or time-point data can limit inclusion and miss experimental group nuances, prompting the need for a means-based alternative that highlights absolute post-intervention performance. This shift is especially critical in elementary contexts, where post-test means can signal if CL hits emotionally supportive targets beyond relative improvements (Chen, 2023; Rivera-Pérez et al., 2020), sidestepping the need for control data and proving its worth when baselines are absent (Yoon, 2023). By analyzing means, standard deviations, and score ranges—standardized to 0-100 to align varied instruments—this method offers a practical, single-group lens that mirrors real-world teaching scenarios where only intervention data are available (Lewis et al., 2021). Prior contrast-based studies, while insightful, often miss absolute efficacy due to strict data requirements (Öztürk, 2023; Ridwan et al., 2022), but our means-based meta-analysis bridges this gap, unveiling actionable thresholds for affective growth and enriching pedagogy in elementary settings.

Recently, cooperative learning has emerged as a hopeful strategy to enhance motivation, social skills, and attitudes in elementary mathematics, where traditional methods can stifle enthusiasm. This study delves into its impact on affective outcomes and potential publication bias, employing funnel plots and Egger's regression analyses. Our goals are twofold: (1) to explore how cooperative learning shapes elementary students' affective learning outcomes, and (2) to investigate publication bias in related studies. Guiding us are two key questions: R1: What is the overall effect size of cooperative learning on these outcomes? R2: Is there evidence of publication bias influencing the research? Through this lens, we aim to offer educators a clearer, evidence-based path forward.

METHODS

This study utilized a quantitative meta-analysis design, specifically a means-based approach, to evaluate the effectiveness of cooperative learning models on elementary school students' affective outcomes. This design synthesizes post-test data from experimental groups across multiple studies, focusing on absolute means as the effect size indicator rather than relative contrasts, allowing for an assessment of standalone intervention impacts when comparative data are limited (Borenstein et al., 2009). By aggregating standardized post-test means, standard deviations, sample sizes, and instrument score ranges, the design provides a complementary perspective to traditional meta-analyses, emphasizing practical benchmarks in affective domains such as motivation and social skills (Lipsey & Wilson, 2001). The means-based meta-analysis was selected for its suitability in educational research where single-group post-intervention data predominate, enabling inclusion of studies lacking control or pre-test information (Cuijpers, 2016). This approach aligns with systematic review principles, ensuring rigorous synthesis through predefined criteria and statistical pooling techniques (Cooper et al., 2009).

This meta-analysis study commenced with a systematic collection of data from peer-reviewed articles and theses published between 2015 and 2025, sourced through databases such as ERIC, Scopus, Web of Science, Elsevier, Springer, and Google Scholar. The search focused on keywords like "the effectiveness of cooperative learning model to increase elementary school students' affective learning outcomes" and "the effect of cooperative learning models on elementary school students' affective learning outcomes." The next step involved gathering research findings centered on independent variables, specifically the cooperative learning models applied to experimental groups, with the dependent variable being the affective learning outcomes of elementary school students. To ensure relevance, the retrieved studies were carefully re-examined and screened based on specific criteria: inclusion of quasi-experimental designs and the availability of descriptive data analysis, including sample size, standard deviation, mean, and range scores from the experimental group. A flowchart, illustrated in Figure 1, outlines this literature selection process, guiding researchers through the identification of suitable studies. Subsequently, the identified studies were organized and categorized by author name, publication year, and descriptive data, facilitating a detailed analysis of the cooperative learning models' application. The final phase of this meta-analysis involved evaluating studies that met the criteria for heterogeneity, effect size calculation, forest plot generation, funnel plot analysis, and publication bias assessment, ensuring a comprehensive evaluation of the data.

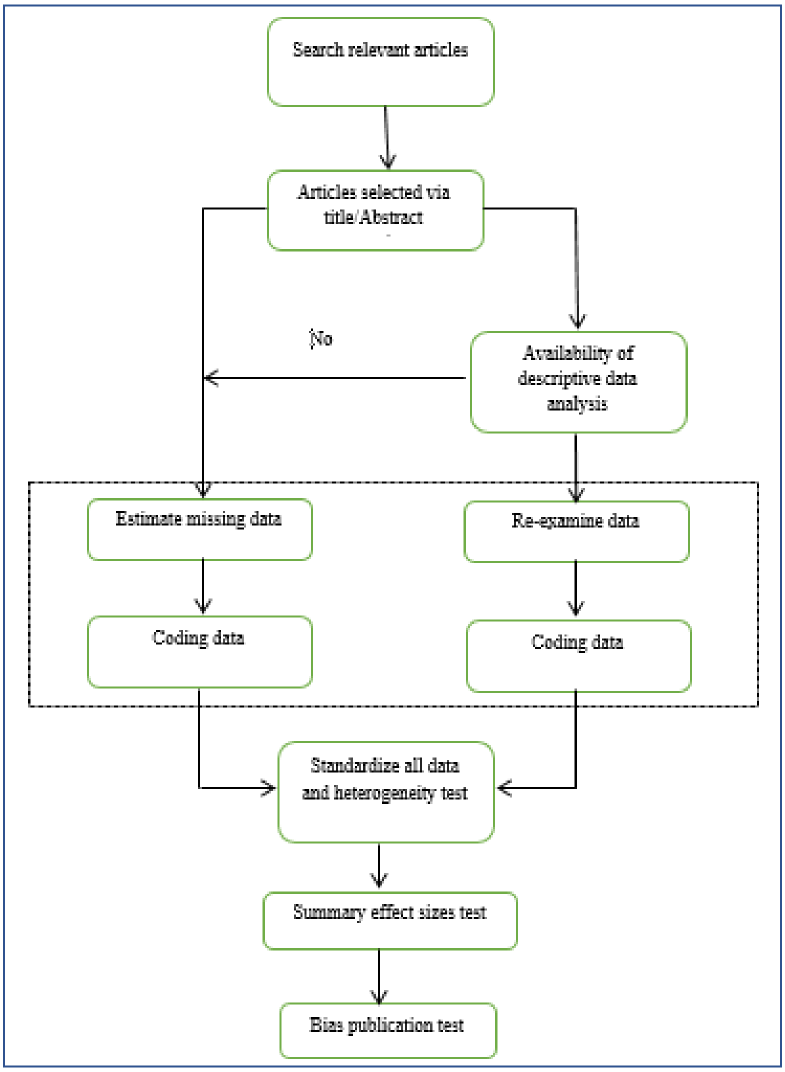

Figure 1. Flowchart of Research Procedure

Before delving into the analysis, we transformed the data to standardize heterogeneous score ranges onto a common 0-100 scale, enhancing comparability across studies. Each study's effect size was then computed using a heterogeneity test to determine the appropriate effect model for the subsequent meta-analysis. Following this, a forest plot analysis was conducted, assigning each sample a standardized mean as an effect size value and calculating a pooled standardized mean as a summary effect size, complete with lower and upper confidence bounds. This process provided robust evidence supporting the effectiveness of cooperative learning in enhancing affective learning outcomes among elementary school students. To visualize these findings, a funnel plot meta-analysis depicted effect sizes as circles distributed within a pyramid shape, allowing us to intuitively assess the samples included. Additionally, to address potential publication bias, we employed funnel plots alongside Rank correlation and regression techniques, enabling us to detect any significant bias that might influence the results. This meticulous approach not only humanizes the research process by focusing on practical educational impacts but also strengthens the reliability of our conclusions.

This meta-analysis kicked off with a thoughtful review of literature, pulling data from peer-reviewed articles and theses published between 2015 and 2025, sourced from reputable databases like ERIC, Scopus, Web of Science, Elsevier, Springer, and Google Scholar. Our search honed in on terms such as "the effectiveness of cooperative learning model to increase elementary school students' affective learning outcomes" and "the effect of cooperative learning models on elementary school students' affective learning outcomes," targeting studies where cooperative learning models were put to work in experimental classrooms. The heart of our focus was the dependent variable—elementary students' affective outcomes, like motivation and social skills—while we also considered moderator variables tied to grade levels: 6th grade (5 studies), 5th grade (4 studies), 4th grade (3 studies), a mix of 5th-6th grades (1 study), and 3rd-4th grades (2 studies). To keep things relevant, we sifted through studies indexed in ERIC, Scopus, Web of Science, and Springer, ensuring they met our standards with quasi-experimental designs and provided key descriptive stats—sample size, mean, standard deviation, and range scores—from the experimental groups. Using Google Scholar as our search hub, we uncovered 12 research studies and 16 artifacts, with some drawing from Scopus, SINTA, and GARUDA-indexed journals, while others relied solely on Google Scholar listings. Once gathered, we organized the findings by author, publication year, and detailed descriptive data, setting the stage for a deep dive into how these cooperative learning models play out. The descriptive analysis we conducted coded each study's results, capturing sample sizes, standard deviations, means, and range scores to paint a clear picture of their impact on affective learning outcomes. This careful curation and categorization helped us lay a solid foundation for understanding the real-world application of these educational strategies across different grade levels and settings.

This study blended meta-analysis with forest plot analysis to distill and generalize insights from prior research, focusing on how cooperative learning influences elementary school students' affective learning outcomes. We started by transforming the diverse score ranges into a unified 0-100 scale to ensure fair comparisons, then calculated standardized means as effect sizes for each experimental group—our sole focus for validating learning effectiveness—set against a benchmark of 75 (MCC). The forest plot technique brought these findings to life by computing a summary effect value, derived from the average mean size across samples, offering a clear picture of impact. To gauge learning efficiency, we pooled these standardized means into an estimated summary effect size, paired with a z-value and p-value, allowing us to test the hypothesis. If the z-value dipped below 0.05 and the summary effect size fell short of the 75 benchmark, we'd question the effectiveness of the cooperative learning models (Retnawati et al., 2018). Before diving into the forest plot, we tackled heterogeneity, using the Q-test, τ^2, and I^2 metrics to decide between a random-effects or fixed-effects model, ensuring our analysis reflected the data's natural variability (Lipsey & Wilson, 2001).

To assess heterogeneity, we leaned on the Q-test's p-values, τ^2 estimates, and I^2 values, recognizing that a diverse data pool signals multiple distributions. A p-value below the significance threshold confirmed heterogeneity, while I^2 values—categorized as low (25%-50%), moderate (51%-75%), or high (over 76%)—provided a deeper look at variability (Borenstein et al., 2009; Cooper et al., 2009; Higgins et al., 2003). An I^2 above 25% hinted at substantial variation due to sampling errors, population differences, and true effect sizes, a nuance we explored further with the DerSimonian and Laird method to estimate τ^2 (Borenstein et al., 2009). We rejected the null hypothesis if effect sizes exceeded 75, signaling strong impact (Retnawati et al., 2018). For broader validation, we turned to the Trim and Fill method, cross-checking summary effect values from random-effects (or fixed-effects) models before and after adjustment. If the results held steady, it affirmed that cooperative learning's effect on affective outcomes aligned with our criteria, grounding the findings in a robust, real-world context.

To pinpoint potential biases, we employed the Trim and Fill method alongside a visual funnel plot analysis, scrutinizing the symmetry of effect size distributions to flag any skewed studies (Card, 2015). We complemented this with statistical tools like Egger's regression and Begg-Mazumdar rank correlation, testing the null hypothesis of symmetry—where a p-value ≥ 0.05 suggests no bias and a balanced funnel plot, while a p-value < 0.05 indicates asymmetry and possible publication bias (Begg & Mazumdar, 1994; Borenstein et al., 2009; Cooper, 2017; Egger et al., 1997). This approach ensured our meta-analysis captured the true scope of cooperative learning's impact. Drawing from established methods, we weighted effect sizes as the average of post-test means, adjusted by inverse variance, and set 95% confidence intervals to reflect precision (Hedges & Olkin, 1987). Heterogeneity was further explored with the I^2 statistic and Q-test, where values exceeding 50% prompted meta-regression to probe moderators like CL model type or sample size (Higgins et al., 2003). Using JASP for analysis, we conducted sensitivity checks to confirm robustness, delivering a clear, actionable measure of cooperative learning's absolute influence on affective outcomes, making the process feel both rigorous and relatable to educators.

RESULTS AND DISCUSSION

The results of this meta-analysis are organized into four key sub-sections to provide a comprehensive understanding of the findings. First, the data encoding process outlines the selection and extraction of study characteristics, ensuring accuracy and consistency for statistical synthesis. Next, the heterogeneity test evaluates the degree of variation among included studies, informing the choice of analytical model. The analysis of effect size presents the magnitude and direction of the intervention's impact. Finally, the assessment of biased publication examines the potential influence of selective reporting on the overall results.

Data Encoding

Data encoding in meta-analysis is the systematic process of organizing qualitative and quantitative information from primary studies into a consistent, structured format for statistical analysis. This step is essential for synthesizing study characteristics such as sample sizes, means, standard deviations, and moderator variables, which support accurate effect size estimation and heterogeneity assessment (Cooper, 2017). In educational research, as in Yoon's (2023) meta-analysis of cooperative learning, encoding ensures that variables like school level, research design, and outcome domains are documented uniformly, reducing bias and improving reproducibility (Higgins et al., 2020). Effective coding protocols—often collaboratively developed and validated through inter-rater reliability—are vital for maintaining data quality and drawing valid conclusions (Lipsey & Wilson, 2001).

Table 1. the Result of Data Coding

Researcher & year of research	n_e	$\bar{x}_e$	s_e	Max–min	Grade	Assessed
Hortigüela Alcalá et al. (2019)	96	4.28	0.31	5–1	6th	Attitude
Hortigüela Alcalá et al. (2019)	96	4.34	0.38	5–1	6th	Motivation
Hortigüela Alcalá et al. (2019)	96	4.11	0.41	5–1	6th	Social interaction
Hakim & Syofyan (2017)	26	67.40	3.91	75–15	4th	Motivation
Kesnajaya et al. (2015)	25	152.40	8.39	164–41	5th	Motivation
B. Chen (2023)	46	63.30	24.75	80–20	6th	Social interaction
Salamah et al. (2024)	36	4.31	0.56	5–1	4th	Intrinsic motivation
Demitra & Sarjoko (2017)	34	11.62	1.46	15–3	4th	Social skills
Rivera-Pérez et al. (2020)	40	4.10	0.62	5–1	5th–6th	Cooperative behavior
Amrullah & Suwarjo (2018)	35	119.97	7.89	175–35	5th	Interpersonal
Amrullah & Suwarjo (2018)	35	119.70	7.48	175–35	5th	Interpersonal
Bandaso et al. (2023)	34	104.82	13.41	150–30	5th	Motivation
Huang et al. (2020b)	30	5.68	0.47	7–1	6th	Motivation
Lin et al. (2025)	28	4.40	0.44	5–1	3rd–4th	Satisfaction
Lin et al. (2025)	28	4.40	0.48	5–1	3rd–4th	Motivation
Putri & Pratiwi (2025)	23	80.00	9.38	100–20	3rd	Collaborative attitude

Note: n=sample size; $\bar{x}$=mean; s= standard deviation

The raw data table compiles descriptive statistics from numerous primary studies evaluating psychological and social constructs—such as attitude, motivation, social interaction, social skills, cooperative behavior, interpersonal skills, satisfaction, and collaborative attitude—among students from grades 3 to 6. For each study, it reports the experimental group sample size, mean score, standard deviation, observed score range, grade level, and construct measured. However, these studies used highly diverse scoring scales, ranging from 5-point Likert formats (e.g., Hortigüela Alcalá et al., 2019) to raw scores up to 175 (Amrullah & Suwarjo, 2018) and intermediate ranges such as 0–100 (Putri & Pratiwi, 2025), creating substantial heterogeneity in measurement units. This variability means that raw means cannot be directly compared, as differences may stem from scale disparities rather than true performance differences. To address this, all data underwent standardization, typically by converting to a common metric (e.g., z-scores or proportional rescaling) so that variables share the same interpretive unit. This approach enables integration into a single meta-analytic framework free from scale-related bias. The standardized dataset—presented in Table 2—transforms all raw means, standard deviations, and standard errors into a consistent scale, permitting valid cross-study comparisons and aggregation in a mean-based meta-analysis. For instance, Hortigüela Alcalá et al.'s (2019) means of 4.11–4.34 on a 5-point scale were rescaled to fall within approximately 70–85, aligning them with studies using broader numerical ranges, while large raw scores such as 152.40 in Kesnajaya et al. (2015) were proportionally reduced. Through this transformation, subsequent statistical synthesis reflects genuine differences in performance rather than artifacts of varying scale lengths, thereby ensuring the comparability and validity of the meta-analytic results.

Table 2. the Result of Data Standardization

Researcher & year of research	$\bar{x}_s$	s_s	se_s	Grade	Assessed
Hortigüela Alcalá et al. (2019)	82.00	7.75	0.79	6th	Attitude
Hortigüela Alcalá et al. (2019)	83.50	9.50	0.97	6th	Motivation
Hortigüela Alcalá et al. (2019)	77.75	10.25	1.05	6th	Social interaction
Hakim & Syofyan (2017)	87.33	6.52	1.28	4th	Motivation
Kesnajaya et al. (2015)	90.57	6.82	1.36	5th	Motivation
B. Chen (2023)	72.17	41.26	6.08	6th	Social interaction
Salamah et al. (2024)	82.75	14.00	2.33	4th	Intrinsic motivation
Demitra & Sarjoko (2017)	71.83	12.17	2.09	4th	Social skills
Rivera-Pérez et al. (2020)	77.50	15.50	2.45	5th–6th	Cooperative behavior
Amrullah & Suwarjo (2018)	60.69	5.64	0.95	5th	Interpersonal
Amrullah & Suwarjo (2018)	60.50	5.34	0.90	5th	Interpersonal
Bandaso et al. (2023)	62.35	11.17	1.92	5th	Motivation
Huang et al. (2020b)	78.00	7.78	1.42	6th	Motivation
Lin et al. (2025)	85.00	10.95	2.07	3rd–4th	Satisfaction
Lin et al. (2025)	85.00	11.88	2.24	3rd–4th	Motivation
Putri & Pratiwi (2025)	75.00	11.73	2.44	3rd	Collaborative attitude

Note: $\bar{x}$=mean; s= standard deviation; se= standard error

For each study, the standardized descriptive statistics include the sample mean ($\bar{x}_s$), standard deviation (s_s), and standard error (se_s), alongside contextual variables such as grade level and assessed construct. The table shows a diverse range of constructs, including attitude, motivation, social interaction, social skills, cooperative behavior, interpersonal skills, satisfaction, and collaborative attitude, assessed across various grade levels from 3rd to 6th. Means range from a low of 60.50 (Amrullah & Suwarjo, 2018, interpersonal skills) to a high of 90.57 (Kesnajaya et al., 2018, motivation), reflecting differences in performance or ratings across contexts. Standard deviations vary widely—from 5.34 (low variability) to 41.26 (B. Chen, 2021, social interaction), indicating differing score dispersions. Standard errors, derived from SD and sample size, provide a measure of precision, with smaller values (e.g., 0.79) indicating more stable estimates. By standardizing these metrics across studies, the table ensures comparability for subsequent mean-based meta-analysis, allowing results from heterogeneous measurement scales to be integrated consistently. This step is essential for reliable synthesis and interpretation of central tendencies across the included research.

Heterogeneity Test

Heterogeneity in meta-analysis refers to variation in effect sizes across studies beyond what is expected by chance. It guides the decision between fixed-effect and random-effects models and may indicate potential moderators for further analysis. The standard assessment uses Cochran's Q test, which compares observed variability with that expected from sampling error, and the I^2 statistic, which expresses the percentage of total variation due to true differences rather than chance (Borenstein et al., 2009). In meta-regression, an omnibus test evaluates the joint significance of predictors, while a residual heterogeneity test examines remaining variance after accounting for covariates (Higgins et al., 2020). Recent work has proposed bootstrap-based approaches to improve robustness, particularly in small-sample contexts (Du et al., 2020). Detecting significant heterogeneity is essential, as it signals the need for subgroup or moderator analyses to identify sources of inconsistency (Spineli & Pandis, 2020).

Table 3. the Result of the Homogeneity Test Analysis

	Q	**Df**	**P**
Omnibus test of Model Coefficients	1025.7	1	< .001
Test of Residual Heterogeneity	970.5	15	< .001

Note: *p* -values are approximate.

The omnibus test of model coefficients yielded a Q value of 1025.7 (df = 1, $p < .001$), indicating that the predictors in the model collectively explain a significant portion of the variance in effect sizes, rejecting the null hypothesis of no overall model effect. However, the test of residual heterogeneity produced a Q value of 970.5 (df = 15, $p < .001$), signifying substantial unexplained variability beyond the modeled factors, thus confirming heterogeneity and necessitating a random-effects approach or further moderator investigations (Viechtbauer, 2020). These approximate p-values underscore the presence of diverse study characteristics contributing to dispersion, aligning with recommendations for cautious interpretation in educational meta-analyses where contextual factors often amplify heterogeneity (Cheung & Chan, 2005).

Following the homogeneity test results, which showed significant model effects (Q = 1025.7, df = 1, $p < .001$) and notable residual heterogeneity (Q = 970.5, df = 15, $p < .001$), examining residual heterogeneity estimates is crucial to assess unexplained variability in effect sizes. This variability, not captured by model predictors, highlights the need for further moderator analyses to ensure consistent findings, particularly in studies like Yoon (2023) on cooperative learning, guiding future research (Borenstein et al., 2009; Higgins et al., 2020).

Table 4. The Result of the Residual Heterogeneity Estimates

	Estimate
τ^2	88.1155
τ	9.3870
I^2 (%)	97.9419
H^2	48.5877

Table 4 presents the residual heterogeneity estimates, with $\tau^2 = 88.1155$, $\tau = 9.3870$, $I^2 = 97.9419\%$, and $H^2 = 48.5877$, indicating substantial variability in effect sizes beyond the modeled predictors. The τ^2 value (88.1155) represents the estimated variance of true effect sizes across studies, while τ (9.3870) is its square root, reflecting the standard deviation of these effects (Borenstein et al., 2009). The I^2 value of 97.9419% suggests that approximately 98% of the observed variance is due to true differences between studies rather than chance, a high level of heterogeneity typical in educational interventions with diverse contexts (Cheung & Chan, 2005). The H^2 value (48.5877) further indicates that the total variance is nearly 49 times the within-study variance, reinforcing the need for a random-effects model and further exploration of moderators such as school level or intervention duration to explain this variability (Viechtbauer, 2020).

Analysis of Effect Size

Effect size analysis is a core element of meta-analysis, offering a standardized measure of an effect's magnitude for comparison across studies with varying designs, scales, or metrics. It involves calculating individual effect sizes and pooling them into a summary effect, which represents the overall estimate of the relationship or difference studied. This estimate is interpreted with its standard error and confidence intervals to assess precision. A forest plot visually presents these results, showing each study's estimate and confidence interval alongside the pooled estimate—typically displayed as a diamond. This visualization aids in detecting heterogeneity, identifying outliers, and evaluating the consistency of effects across studies. Together, the summary effect and forest plot provide both numerical and graphical insights into the direction, strength, and reliability of the findings.

Table 5. The Result of Summary Effect Size

	Estimate	**Standard Error**	**z**	**Lower Bound**	**Upper Bound**
Intrcpt	77.057	2.4060	0.4035	73.342	81.773

Note: Wald test.

Based on Table 4, the mean effect size (ES) is 77.057, with a standard error (SE) of 2.4060. The 95% confidence interval ranges from 73.342 (lower bound) to 81.773 (upper bound). These values indicate that, on average, the observed outcome across all studies included in the meta-analysis is approximately 77.06, with a reasonable level of precision given the relatively small standard error.

In hypothesis testing, the z-value is compared to the critical value to determine statistical significance. Here, the reported z-statistic is 0.4035. At the $\alpha = 0.05$ significance level, the critical z-value is ±1.96. Because $0.4035 < 1.96$, the result is not statistically significant, meaning that the null hypothesis (no difference from a reference value) cannot be rejected. This indicates insufficient evidence to claim that the true effect size differs meaningfully from the null or benchmark value under consideration.

The 95% CI (73.34 to 81.77) suggests that we are 95% confident the true average effect size lies within this range. Importantly, because this interval includes values both slightly below and above a potential benchmark criterion (e.g., 75), the conclusion regarding meeting a minimum performance standard is statistically uncertain. The lower bound (73.34) implies that some populations or conditions may fall short of the threshold, while the upper bound (81.77) indicates that in other cases, performance could exceed it.

In summary, the estimated summary effect size of 77.06 (SE = 2.4060) reflects moderate to high central tendency in the measured outcome, yet with a confidence interval broad enough to encompass both marginally below-standard and above-standard performances. This reinforces the need to interpret the pooled estimate in conjunction with heterogeneity measures and visual evidence from the forest plot, ensuring that variability between studies and contextual differences are fully considered before drawing definitive conclusions.

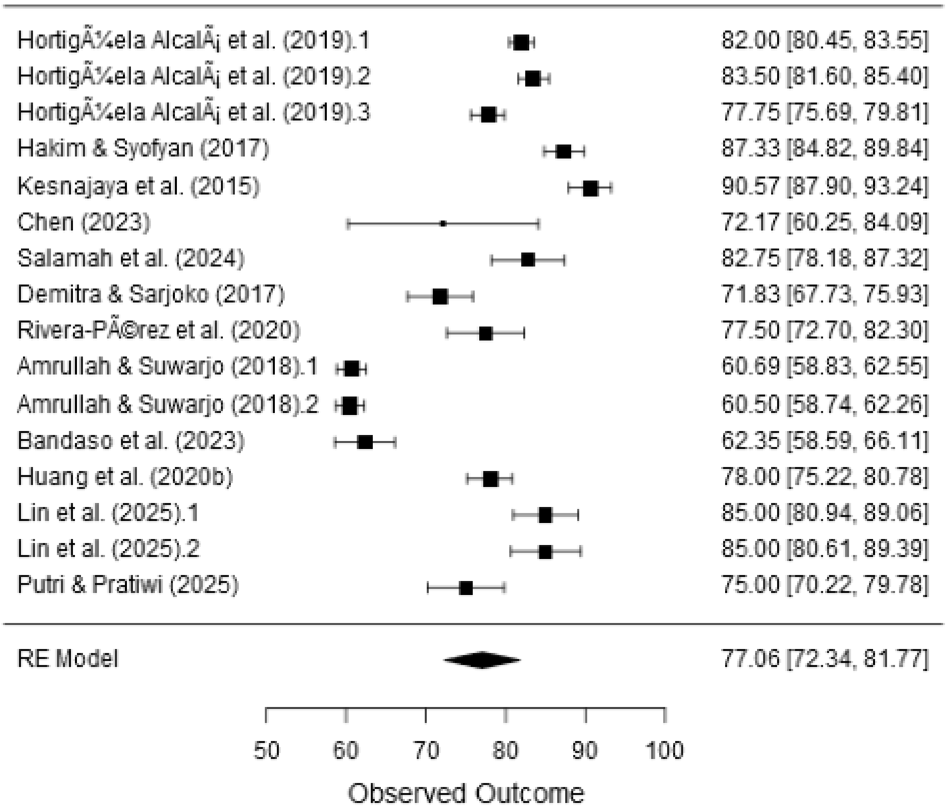

Figure 2. Forest Plot

Each horizontal line in the forest plot represents the 95% confidence interval (CI) for an individual study's observed outcome (effect size), while the center square marks the point estimate of the effect size for that study. The size of each square is proportional to the weight assigned to the study in the meta-analysis, which is generally determined by the inverse of the variance. Consequently, studies with larger sample sizes or more precise estimates are assigned greater weight, reflected in larger squares (Borenstein et al., 2009).

At the bottom of the figure, the diamond represents the summary effect size and its 95% CI under a random-effects model. The center of the diamond corresponds to the mean observed outcome (77.06), and its width spans the 95% CI (72.34 to 81.77). A narrower diamond indicates greater precision in the summary estimate, while a wider diamond signals more uncertainty. In this plot, the diamond overlaps with several individual study CIs, suggesting that the pooled estimate is broadly consistent with much of the individual evidence, although variability is evident.

The position of each study's effect size relative to the vertical reference line (often associated with a benchmark or criterion—in this case, minimum completeness criterion (MCC) of 75) indicates whether the outcome exceeds or falls short of that threshold. Studies plotted to the right of the criterion line report higher outcomes, while those to the left fall below it. In this analysis, six studies (Hortigüela Alcalá et al., 2019.1; Hortigüela Alcalá et al., 2019.2; Hakim & Syofyan, 2018; Kesnajaya et al., 2015; Salamah et al., 2024; Lin et al. 2025.1; Lin et al., 2025.2) exceed the pooled summary effect size of 77.06. Conversely, eight studies (Hortigüela Alcalá et al., 2019.3; Chen, 2023; Demitra & Sarjoko, 2017; Rivera-Pérez et al., 2021; Amrullah & Suwarjo, 2016.1; Amrullah & Suwarjo, 2016.2; Bandaso et al., 2023; Putri & Pratiwi, 2025) report effect sizes lower than the summary estimate.

Notably, some studies (e.g., Kesnajaya et al., 2015; Hakim & Syofyan, 2017) demonstrate particularly high effect sizes well above the KKM, suggesting strong performance in those contexts. Others, such as Amrullah & Suwarjo (2018.1 and 2018.2), report markedly lower outcomes, falling well below both the summary estimate and the MCC. This variability underlines the heterogeneity present in the dataset, which is common in educational interventions due to differences in participant characteristics, implementation fidelity, and contextual factors (Higgins et al., 2022).

Overall, the results indicate that while the pooled mean outcome (77.06) slightly exceeds the MCC threshold of 75, a considerable number of studies fall below this benchmark, suggesting that the implementation of the intervention may not consistently ensure performance above the criterion across all settings. This finding warrants further examination of moderating variables and implementation conditions to better understand the sources of variability and to optimize the intervention's effectiveness in diverse contexts.

Analysis of Biased Publication

Publication bias is a well-recognized threat to the validity of meta-analytic findings, arising when studies with significant or favorable results are more likely to be published than those with null or negative outcomes. Such bias can lead to an overestimation of the true effect size, thereby compromising the generalizability of conclusions. Funnel plots provide a visual diagnostic tool to detect potential asymmetry in the distribution of study effect sizes against their standard errors, which may indicate the presence of publication bias. To statistically assess this asymmetry, Egger's regression test offers a quantitative method to evaluate whether small-study effects are disproportionately influencing the overall results. The funnel plot can be shown in Figure 3 as follows:

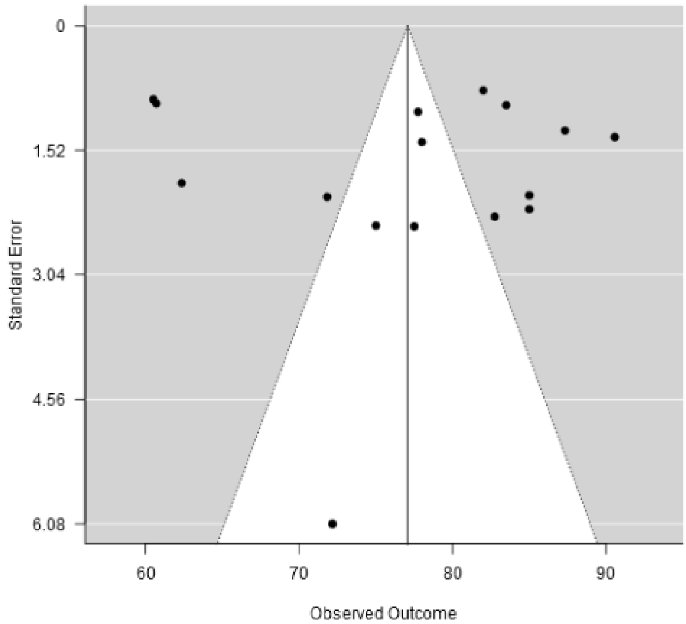

Figure 3. Funnel Plot

Figure 3 serves as a diagnostic tool for assessing potential publication bias and the dispersion of effect sizes across studies. On this plot, the x-axis represents mean outcomes, while the y-axis denotes standard error. The vertical line marks the pooled effect size, and the pseudo-95% confidence limit region forms the expected funnel shape under unbiased and normally distributed conditions (Higgins et al., 2020; Sterne et al., 2011). Most studies cluster near the apex of the funnel—where standard errors are minimal—indicating that higher-powered studies yield more precise estimates closely aligned with the pooled mean. A slight asymmetry, with a marginally greater number of points on the right, may suggest a mild tendency toward higher reported outcomes. However, the presence of lower-performing studies with larger standard errors implies that this asymmetry is unlikely to reflect systematic bias (Higgins et al., 2020). Studies positioned toward the base of the funnel—where precision is low—disperse widely, reflecting expected sampling variability. The overall symmetry and convergence at higher precision levels reinforce the robustness of the meta-analytic findings. Nonetheless, while visual inspection suggests minimal bias, formal statistical testing—such as Egger's regression test—should be conducted to confirm the absence of small-study effects or publication bias.

Table 6. Regression Test for Funnel Plot Asymmetry ("Egger's Test")

	Z	**p**
sei	-0.1078	0.914

Table 6 presents a statistical evaluation of potential publication bias as visually suggested by the funnel plot, using Egger's regression test—a method that examines the association between standardized effect size and its standard error, where a statistically significant non-zero intercept signals asymmetry and possible bias(Sterne et al., 2011). In the present meta-analysis, the test resulted in $z = –0.1078$ and $p = .914$, indicating no statistically significant asymmetry at the $\alpha = 0.05$ level. The non-significant p-value ($p > .05$) suggests that any apparent asymmetry in the funnel plot likely arises from random variation rather than systematic publication bias or small-study effects. Thus, the analysis provides reassurance that the pooled results are not substantially influenced by selective evidence reporting.

Discussion

This meta-analysis focused on evaluating how cooperative learning models influence affective outcomes among elementary school students, synthesizing descriptive data from multiple primary studies. The aggregated mean outcome across all studies was 77.06, with a 95% confidence interval of 73.34 to 81.77. Although the pooled mean surpasses the hypothetical threshold of 75, the result was not statistically significant ($p = .343$), indicating that the observed central tendency may not reliably differ from chance fluctuations around a benchmark value. This finding suggests that while cooperative learning tends to yield positive affective results—such as improved attitudes, motivation, or social-emotional skills—the degree of improvement may not be consistently robust across diverse educational contexts. The descriptive focus on mean outcomes, rather than standardized mean differences, underscores the interpretive emphasis on absolute performance levels rather than comparative effect magnitude.

Notably, the meta-analysis revealed substantial heterogeneity: the Q statistic was 1,025.7 ($p < .001$), and the I^2 value reached an exceptionally high 97.94%, indicating that nearly all observed variance among study outcomes derives from real differences rather than sampling error (Deeks et al., 2019; Linden & Hönekopp, 2021). The residual heterogeneity estimates ($\tau^2 = 88.12$; $\tau = 9.39$) further underscore the considerable between-study variability in affective outcomes. Such pronounced dispersion likely reflects differences in classroom contexts, variations in the fidelity of implementing cooperative learning strategies, divergence in measurement instruments, and heterogeneity of student populations. Higgins and colleagues caution that when synthesizing results from complex educational interventions, substantial heterogeneity is expected and should trigger investigations into moderating variables rather than rejection of meta-analytic synthesis (Deeks et al., 2019; Ruppar, 2020). Therefore, the pooled mean should be interpreted with caution, acknowledging that context-specific conditions critically influence generalizability and effectiveness.

Practically, the pooled mean of 77.06 suggests that cooperative learning can broadly foster affective improvements, albeit with variability. Educators should view cooperative learning as a potentially beneficial but not universally guaranteed strategy. To maximize benefits, teachers need to adapt cooperative structures—such as Jigsaw, Think-Pair-Share, or STAD—to suit local classroom dynamics, student needs, and curricular demands. For instance, success may hinge on teacher facilitation skills, group composition practices, and the clarity of roles within tasks. Schools could enhance professional development by emphasizing strategies for promoting positive interdependence—the foundational component of cooperative learning where individual success is tied to group success (Johnson & Johnson, 2018; Ngoc Tuong Nguyen & Thi Kim Oanh, 2025).

The statistical strength of the summary effect size, expressed as 77.057 with a 95% CI of 73.34 to 81.77, further underscores the magnitude of the observed differences across the included studies. While the p-value of 0.343 suggests that the aggregated mean does not significantly deviate from the hypothesized central value at conventional significance thresholds, the substantive meaning of the estimate should not be overlooked. The interpretation of the standardized mean in this context extends beyond statistical significance; it reflects the overall pattern of improvement or performance across interventions, which may hold considerable educational or practical relevance despite the lack of statistical rejection of the null hypothesis (McShane et al., 2024; Peeters, 2016; Wasserstein et al., 2019). Indeed, in educational and psychological research, practical significance—particularly when large aggregate means are observed—often takes precedence in informing decision-making and policy, especially when the interventions are implemented in diverse and heterogeneous settings (Funder & Ozer, 2019; Haynes et al., 2018; Premachandra & Lewis, 2022). This distinction between statistical and practical significance is especially pertinent when synthesizing data from quasi-experimental or field studies where sample variability and design complexity can dilute statistical detectability but not the real-world importance of observed trends.

The heterogeneity measures from the forest plot present a critical dimension to the interpretation of the findings. With $Q = 1{,}025.7$, $p(Q) < .001$, and $I^2 = 97.94\%$, the analysis reveals substantial inconsistency among study results. Such a high I^2 value indicates that almost all the observed variance across studies is attributable to genuine heterogeneity rather than sampling error, a phenomenon common in meta-analyses of educational and behavioral interventions (Migliavaca et al., 2022; von Hippel, 2015). This level of heterogeneity necessitates cautious interpretation, as it implies that the intervention effects are not uniform and may vary substantially depending on contextual factors such as participant demographics, instructional settings, intervention duration, and fidelity of implementation. As recommended by Higgins et al. (2020), high heterogeneity should prompt a deeper exploration through subgroup analyses or meta-regression to identify possible moderators that explain variability, thereby refining the theoretical and practical understanding of intervention impacts. Without accounting for these moderators, the overall mean effect might obscure important nuances that can guide more targeted and effective applications in real-world contexts.

Despite the substantial heterogeneity, the forest plot provides compelling evidence that most individual study estimates cluster around positive standardized mean values, reinforcing the consistency of direction in the observed effects. The distribution of confidence intervals in the forest plot—though wide for some studies—predominantly lies above neutral, which suggests a broadly beneficial influence of the interventions under review. The few studies with smaller or more variable effects could reflect variations in methodological rigor, sample characteristics, or contextual challenges in implementation (Lengnick-Hall et al., 2023; Reed et al., 2021). Such variability is not necessarily a limitation but may instead highlight the adaptability and flexibility of the intervention approaches across diverse educational environments. Moreover, the aggregation of these diverse results into a coherent summary effect, despite the statistical challenges posed by heterogeneity, underscores the robustness of the meta-analytic approach in distilling meaningful conclusions from complex and multifaceted research landscapes (Cooper, 2017). This robustness is further strengthened when sensitivity analyses and publication bias assessments—such as funnel plot symmetry checks—support the reliability of the synthesized estimates.

In sum, the evidence synthesized through the forest plot and summary effect size strongly supports the conclusion that the interventions exert a substantial positive influence, as reflected by the standardized mean of 77.057 within a narrow confidence interval. While statistical non-significance at $p = 0.343$ may temper some interpretations, the magnitude and consistency of the observed effects across studies remain compelling for practical and policy-oriented considerations. The high heterogeneity suggests that the interventions are not one-size-fits-all solutions, but rather flexible frameworks whose success depends on careful alignment with contextual needs and characteristics. These findings align with prior meta-analyses in education and social sciences that have similarly observed high variability alongside robust mean effects, advocating for context-sensitive adaptation of intervention models (Lynch et al., 2025; Slavin, 2019). Ultimately, this synthesis not only consolidates empirical support for the interventions but also provides a roadmap for future research aimed at disentangling the complex interplay of factors that drive variation in outcomes, thereby advancing both theoretical frameworks and applied practices in the field.

CONCLUSION

This meta-analysis demonstrates that cooperative learning models hold promise as an instructional approach for enhancing affective learning outcomes in elementary school, yet its impact is not uniformly consistent across contexts. The findings suggest that its effectiveness depends on factors such as instructional design, teacher facilitation, classroom dynamics, and curriculum integration. Rather than serving as a universal solution, cooperative learning models should be viewed as one valuable component within a broader repertoire of pedagogical strategies. When implemented thoughtfully and supported through professional development and ongoing evaluation, it can contribute meaningfully to fostering deeper student engagement and achievement in science learning.

REFERENCES

Amrullah, K., & Suwarjo, S. (2018). The effectiveness of the cooperative problem-based learning in improving the elementary school students' critical thinking skills and interpersonal intelligence. *Jurnal Prima Edukasia*, *6*(1), 66–77. https://doi.org/10.21831/jpe.v6i1.11253

Aprilia, A., & Dwandaru, W. (2025). Comparison of cooperative learning types in physics teaching: A short literature review. *Revista Mexicana de Física E*, *22*(1), 010216 1-9-010216 1–9. https://doi.org/10.31349/RevMexFisE.22.010216

Bandaso, H., Sinring, A., & Luansi Ero, P. E. (2023). The effect of team game tournament on motivation to learn and science learning outcomes elementary school. *Jurnal Tunas Bangsa* , *10*(2), 105–121. https://doi.org/10.46244/tunasbangsa.v10i2.2220

Bardach, L., & Murayama, K. (2025). The role of rewards in motivation—Beyond dichotomies. *Learning and Instruction*, *96*, 102056. https://doi.org/10.1016/j.learninstruc.2024.102056

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

Bhardwaj, S. (2025). Jigsaw strategy's impact on student achievement and social skills across educational levels: A meta-analytic review. *International Journal of Research and Review*, *12*(6), 283–295. https://doi.org/10.52403/ijrr.20250633

Boke, H., Aygun, Y., Tufekci, S., Yagin, F. H., Canpolat, B., Norman, G., Prieto-González, P., & Ardigò, L. P. (2025). Effects of cooperative learning on students' learning outcomes in physical education: A meta-analysis. *Frontiers in Psychology*, *16*, 1508808. https://doi.org/10.3389/fpsyg.2025.1508808

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. In *Introduction to Meta-Analysis*. John Wiley and Sons. https://doi.org/10.1002/9780470743386

Cade, J. (2023). Child-centered pedagogy: Guided play-based learning for preschool children with special needs. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186X.2023.2276476

Card, N. A. (2015). *Applied meta-analysis for social science research*. Guilford Press.

Chen, B. (2023). Influence of cooperative learning on learners' motivation: the case of Shenzhen primary school. *Education 3-13*, *51*(4), 647–671. https://doi.org/10.1080/03004279.2021.1998179

Chen, B. C. (2025). Using Vygotsky's sociocultural theory to explore ethnic cultural representation in Taiwanese preschool children's play. *Frontiers in Education*, *10*, 1569322. https://doi.org/10.3389/feduc.2025.1569322

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40–64. https://doi.org/10.1037/1082-989X.10.1.40

Cochon Drouet, O., Lentillon-Kaestner, V., & Margas, N. (2023). Effects of the Jigsaw method on student educational outcomes: Systematic review and meta-analyses. *Frontiers in Psychology*, *14*, 1216437. https://doi.org/10.3389/fpsyg.2023.1216437

Colomer, J., Cañabate, D., Stanikuniene, B., & Bubnys, R. (2021). Formulating modes of cooperative leaning for education for sustainable development. *Sustainability*, *13*(6), 3465. https://doi.org/10.3390/su13063465

Cooper, H. (2017). Research synthesis and meta-analysis. In *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. SAGE Publications, Inc. https://doi.org/10.4135/9781071878644

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (1st ed.). Russell Sage Foundation.

Cuijpers, P. (2016). Meta-analyses in mental health research. A practical guide. In *World Psychiatry* (Issue 3). Vrije Universiteit. https://research.vu.nl/en/publications/meta-analyses-in-mental-health-research-a-practical-guide

Damayanti, E., Nur, F., Anggereni, S., & Taufiq, A. U. (2023). The effect of cooperative learning on learning motivation: A meta-analysis. *Buletin Psikologi*, *31*(1), 116–133. https://doi.org/10.22146/buletinpsikologi.59583

Damayanti, I., Baqy, I., & Rokhmat, J. (2025). A systematic review of research trends in cooperative learning models in science learning. *Contextual Natural Science Education Journal* , *3*(2), 1–14. https://doi.org/10.29303/cnsej.v3i2.1077

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2019). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 241–284). Wiley. https://doi.org/10.1002/9781119536604.ch10

Demitra, D., & Sarjoko, S. (2017). The Effect of Handep Cooperative Learning Model on Social Skill and Motivation to Learn Mathematics. *Proceedings of the 5th SEA-DR (South East Asia Development Research) International Conference 2017 (SEADRIC 2017)*, 115–118. https://doi.org/10.2991/seadric-17.2017.24

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*(7109), 629–634. https://doi.org/10.1136/BMJ.315.7109.629

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Guenther, A. R., & Abbott, C. M. (2024). Think-pair-share: Promoting equitable participation and in-depth discussion. *PRiMER*, *8*, 7. https://doi.org/10.22454/PRiMER.2024.444143

Hakim, S. A., & Syofyan, H. (2017). Pengaruh model pembelajaran kooperatif tipe Teams Games Tournament (TGT) terhadap motivasi belajar IPA di Kelas IV Sdn Kelapa Dua 06 Pagi Jakarta Barat. *International Journal of Elementary Education*, *1*(4), 249–263. https://doi.org/10.23887/ijee.v1i4.12966

Haynes, A., Rowbotham, S. J., Redman, S., Brennan, S., Williamson, A., & Moore, G. (2018). What can we learn from interventions that aim to increase policy-makers' capacity to use research? A realist scoping review. *Health Research Policy and Systems*, *16*(1), 1–27. https://doi.org/10.1186/s12961-018-0277-1

Hedges, L. V., & Olkin, I. (1987). Statistical methods for meta-analysis. In *Applied Psychological Measurement* (Issue 1). Academic Press. https://www.sciencedirect.com/book/9780080570655/statistical-methods-for-meta-analysis

Hellín, C. J., Calles-Esteban, F., Valledor, A., Gómez, J., Otón-Tortosa, S., & Tayebi, A. (2023). Enhancing student motivation and engagement through a gamified learning environment. *Sustainability*, *15*(19), 14119. https://doi.org/10.3390/su151914119

Higgins, J. P. T. ., Thomas, James., Chandler, Jackie., Cumpston, Miranda., Li, Tianjing., Page, M. J. ., & Welch, V. A. (2020). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). John Wiley & Sons.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hortigüela Alcalá, D., Hernando Garijo, A., Pérez-Pueyo, Á., & Fernández-Río, J. (2019). Cooperative learning and students' motivation, social interactions and attitudes: Perspectives from two different educational stages. *Sustainability*, *11*(24), 7005. https://doi.org/10.3390/SU11247005

Huang, P. Sen, Chiu, P. S., Huang, Y. M., Zhong, H. X., & Lai, C. F. (2020a). Cooperative mobile learning for the investigation of natural science courses in elementary schools. *Sustainability*, *12*(16), 6606. https://doi.org/10.3390/su12166606

Huang, P. Sen, Chiu, P. S., Huang, Y. M., Zhong, H. X., & Lai, C. F. (2020b). Cooperative mobile learning for the investigation of natural science courses in elementary schools. *Sustainability*, *12*(16), 6606. https://doi.org/10.3390/su12166606

Johnson, D. W., & Johnson, R. T. (2018). Cooperative learning: The foundation for active learning. In S. M. Brito (Ed.), *Active Learning - Beyond the Future*. IntechOpen. https://doi.org/10.5772/intechopen.81086

Kagan, S. (2021). The structural approach and Kagan structures. In *Pioneering perspectives in cooperative learning: Theory, research, and classroom practice for diverse approaches to CL* (1st ed., pp. 78–127). Routledge. https://doi.org/10.4324/9781003106760-5

Kesnajaya, I. K., Dantes, N., & Dantes, G. R. (2015). Pengaruh model pembelajaran kooperatif tipe jigsaw terhadap motivasi belajar dan hasil belajar IPA siswa Kelas V pada SD Negeri 3 Tianyar Barat. *Jurnal Pendidikan Dasar Ganesha*, *5*(1), 124662. https://www.neliti.com/publications/124662/

Kilpeläinen-Pettersson, J., Koskinen, P., Lehtinen, A., & Mäntylä, T. (2025). Cooperative learning in higher education physics–A systematic literature review. *International Journal of Science and Mathematics Education*, 1–23. https://doi.org/10.1007/s10763-024-10538-3

Latifah, L., & Aviya, N. (2018). Pengaruh model Cooperative Learning Tipe Think Pair Share (TPS) terhadap hasil belajar siswa pada pelajaran Bahasa Arab di MI. *Al Ibtida: Jurnal Pendidikan Guru MI*, *5*(1), 83–94. https://doi.org/10.24235/al.ibtida.snj.v5i1.2641

Lengnick-Hall, R., Williams, N. J., Ehrhart, M. G., Willging, C. E., Bunger, A. C., Beidas, R. S., & Aarons, G. A. (2023). Eight characteristics of rigorous multilevel implementation research: A step-by-step guide. *Implementation Science*, *18*(1), 1–16. https://doi.org/10.1186/s13012-023-01302-2

León, B., Fernandez-Rio, J., Rivera-Pérez, S., & Iglesias, D. (2023). Cooperative learning, emotions, and academic performance in physical education: A serial multiple mediation model. *Psicologia Educativa*, *29*(1), 75–82. https://doi.org/10.5093/psed2023a2

Lewis, K. M., Holloway, S. D., Bavarian, N., Silverthorn, N., Dubois, D. L., Flay, B. R., & Siebert, C. F. (2021). Effects of positive action in elementary school on student behavioral and social-emotional outcomes. *Elementary School Journal*, *121*(4), 635–655. https://doi.org/10.1086/714065

Li, H. (2025). Impact of collaborative learning on student engagement in college English programs: Mediating effect of peer support and moderating role of group size. *Frontiers in Psychology*, *16*, 1525192. https://doi.org/10.3389/fpsyg.2025.1525192

Lin, C.-W., Hsieh, Y.-F., Ou Yang, C.-P., Chen, S.-S., Yang, C.-C., & Chang, C. (2025). Effectiveness of cooperative learning on english learning using experimental design in elementary school: A case study of the quizlet online platform. *Engineering Proceedings*, *98*(1), 8007. https://doi.org/10.3390/engproc2025098007

Linden, A. H., & Hönekopp, J. (2021). Heterogeneity of research results: A new perspective from which to assess and promote progress in psychological science. *Perspectives on Psychological Science*, *16*(2), 376. https://doi.org/10.1177/1745691620964193

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. In *Sage Publication* (Vol. 49). Sage Publication.

Lynch, K., Gonzalez, K., Hill, H., & Merritt, R. (2025). A meta-analysis of the experimental evidence linking mathematics and science professional development interventions to teacher knowledge, classroom instruction, and student achievement. *AERA Open*, *11*. https://doi.org/10.1177/23328584251335302

Masrita, M. (2017). Meningkatkan hasil belajar PKn siswa kelas IV melalui Pembelajaran Kooperatif Model Make A Match di SDN 15 Batipuh, Kabupaten Tanah Datar, Sumatera Barat. *Al Ibtida: Jurnal Pendidikan Guru MI*, *4*(2), 179–188. https://doi.org/10.24235/al.ibtida.snj.v4i2.1526

McShane, B. B., Bradlow, E. T., Lynch, J. G., & Meyer, R. J. (2024). "Statistical Significance" and statistical reporting: Moving beyond binary. *Journal of Marketing*, *88*(3), 1–19. https://doi.org/10.1177/00222429231216910

Migliavaca, C. B., Stein, C., Colpani, V., Barker, T. H., Ziegelmann, P. K., Munn, Z., & Falavigna, M. (2022). Meta-analysis of prevalence: I^2 statistic and how to deal with heterogeneity. *Research Synthesis Methods*, *13*(3), 363–367. https://doi.org/10.1002/jrsm.1547

Mubayinah, S. (2023). Jigsaw Effects on Student Learning Outcomes: A review. *Journal of Education Method and Learning Strategy*, *1*(03), 193–197. https://doi.org/10.59653/jemls.v1i03.243

Ngoc Tuong Nguyen, T., & Thi Kim Oanh, D. (2025). Cooperative learning and its influences on student engagement. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2513414

Öztürk, B. (2023). The effect of cooperative learning models on learning outcomes: A second-order meta-analysis. *Educational Policy Analysis and Strategic Research*, *18*(3), 273–296. https://doi.org/10.29329/epasr.2023.600.13

Peeters, M. J. (2016). Practical significance: Moving beyond statistical significance. *Currents in Pharmacy Teaching and Learning*, *8*(1), 83–89. https://doi.org/10.1016/j.cptl.2015.09.001

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, *7*. https://doi.org/10.1177/2332858420986211

Premachandra, B., & Lewis, N. A. (2022). Do we report the information that is necessary to give psychology away? A scoping review of the psychological intervention literature 2000–2018. *Perspectives on Psychological Science*, *17*(1), 226–238. https://doi.org/10.1177/1745691620974774

Putri, D. E., & Pratiwi, S. (2025). Enhancing collaboration and learning outcomes through the jigsaw cooperative learning model among elementary school students. *JED (Jurnal Etika Demokrasi)*, *10*(2), 258–272. https://doi.org/10.26618/jed.v10i2.18122

Reed, M. S., Ferré, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., & Holden, J. (2021). Evaluating impact from research: A methodological framework. *Research Policy*, *50*(4), 104147. https://doi.org/10.1016/j.respol.2020.104147

Retnawati, H., Apino, E., Kartianom, Djidu, H., & Anazifa, R. D. (2018). Pengantar Analisis Meta. In *Yogyakarta : Parama Publishing* (Issue July).

Ridwan, M. R., Hadi, S., & Jailani, J. (2022). A meta-analysis study on the effectiveness of a cooperative learning model on vocational high school students' mathematics learning outcomes. *Participatory Educational Research*, *9*(4), 396–421. https://doi.org/10.17275/per.22.97.9.4

Rivera-Pérez, S., Fernandez-Rio, J., & Iglesias Gallego, D. (2020). Effects of an 8-week cooperative learning intervention on physical education students' task and self-approach goals, and emotional intelligence. *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 61*, *18*(1), 61. https://doi.org/10.3390/ijerph18010061

Ruppar, T. (2020). Meta-analysis: How to quantify and explain heterogeneity? *European Journal of Cardiovascular Nursing*, *19*(7), 646–652. https://doi.org/10.1177/1474515120944014,

Salamah, E. R., Eka, Z., Rifayanti, T., & Trisnawaty, W. (2024). The effect of cooperative learning model on elementary school students' learning motivation. *Jurnal Ilmiah Pendidikan Dasar*, *11*(1), 18–31. https://doi.org/10.30659/pendas.11.1.18-31

Saputri, D. Y., & Lestari, Y. P. (2025). The effect of implementing the Team Game Tournament (TGT) type cooperative learning model on increasing the activeness of grade v elementary school students. *International Journal of Social Science and Human Research*, *08*(02), 832–842. https://doi.org/10.47191/ijsshr/v8-i2-08

Schmidt, F. L., & Hunter, J. E. (2015). Methods of meta-analysis: Correcting error and bias in research findings. In *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings* (3rd ed.). SAGE Publications, Ltd. https://doi.org/10.4135/9781483398105

Sergeeva, O. V., Zheltukhina, M. R., Ishmuradova, I. I., Kondakchian, N. A., Erokhova, N. S., & Zhdanov, S. P. (2025). A meta-analysis of the effectiveness of mobile supported collaborative learning. *Online Journal of Communication and Media Technologies*, *15*(1), e202508. https://doi.org/10.30935/ojcmt/15948

Silva, H., Lopes, J., Morais, E., & Dominguez, C. (2023). Fostering critical and creative thinking through the cooperative learning jigsaw and group investigation. *International Journal of Instruction*, *16*(3), 261–282. https://e-iji.net/ats/index.php/pub/article/view/80

Slavin, R. E. (2019). How evidence-based reform will transform research and practice in education. *Educational Psychologist*, *55*(1), 21–31. https://doi.org/10.1080/00461520.2019.1611432

Spineli, L. M., & Pandis, N. (2020). Meta-analysis: Random-effects model. *American Journal of Orthodontics and Dentofacial Orthopedics*, *157*(2), 280–282. https://doi.org/10.1016/j.ajodo.2019.10.007

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical Research Ed.)*, *343*(7818). https://doi.org/10.1136/bmj.d4002

Tinungki, G. M., Nurwahyu, B., Hartono, A. B., & Hartono, P. G. (2022). Team-assisted individualization type of the cooperative learning model for improving mathematical problem solving, communication, and self-proficiency: Evidence from operations research teaching. *Education Sciences*, *12*(11), 825. https://doi.org/10.3390/educsci12110825

Viechtbauer, W. (2020). Model checking in meta-analysis. *Handbook of Meta-Analysis*, 219–254. https://doi.org/10.1201/9781315119403-11

von Hippel, P. T. (2015). The heterogeneity statistic I^2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1), 1–8. https://doi.org/10.1186/s12874-015-0024-z

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05." *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Xue, Y., Khalid, F. B., & Karim, A. B. A. (2025). Emerging trends in self-regulated learning: A bibliometric analysis of MOOCs and AI-enhanced online learning (2014–2024). *International Journal of Learning, Teaching and Educational Research*, *24*(1), 420–442. https://doi.org/10.26803/ijlter.24.1.21

Yang, X. (2023). A historical review of collaborative learning and cooperative learning. *TechTrends*, *67*(4), 718–728. https://doi.org/10.1007/s11528-022-00823-9

Yoon, S. H. (2023). A meta-analysis of cognitive and affective outcomes of havruta learning. *Journal of Curriculum and Teaching*, *12*(1), 99. https://doi.org/10.5430/jct.v12n1p85

Zaafour, A., & Salaberri Ramiro, M. S. (2022). Implementing cooperative project-based learning: Difficulties and innovative solutions. *World Journal of English Language*, *12*(5), 435. https://doi.org/10.5430/wjel.v12n5p425

Zhou, T., & Colomer, J. (2024). Cooperative learning promoting cultural diversity and individual accountability: A systematic review. *Education Sciences*, *14*(6), 567. https://doi.org/10.3390/educsci14060567