BIoLAE - Journal
Britain International of Linguistics, Arts and Education Sciences Journal
Publish : March - July - November
E-ISSN: 2685 - 4813
P-ISSN: 2685 - 4805
http://biarjournal.com/index.php/biolae

Synergizing Artificial Intelligence and Multiple Intelligences in Project-Based Learning: A Meta-Analysis of Academic Achievement Outcomes

Belay Sitotaw Goshu[1], Muhammad Ridwan[2]
[1]Department of Physics, Dire Dawa University, Dire Dawa, Ethiopia
[2]Universitas Islam Negeri Sumatera Utara, Indonesia
Email: belaysitotaw@gmail.com, bukharyahmedal@gmail.com

Abstract:
Project based learning (PBL) promotes deeper learning but often fails to accommodate diverse cognitive profiles. Artificial intelligence (AI) offers adaptive scaffolding, while Multiple Intelligences (MI) theory provides a differentiation framework. However, no quantitative synthesis has examined their combined effect on academic achievement. This meta analysis synthesizes evidence on the synergy of AI and MI within PBL and estimates the overall effect on academic achievement, together with key moderators. Methods: Following PRISMA 2020 guidelines, we systematically searched Scopus, Web of Science, ERIC, PsycINFO, and ProQuest Dissertations (2015–2025). Inclusion criteria were: (a) empirical studies with control/comparison groups; (b) interventions combining AI tools and MI based differentiation in PBL; (c) reported academic achievement data; (d) K 16 learners. Random effects meta analysis, moderator analyses, and publication bias tests were performed. Forty two studies (N = 8,943) were included. The overall effect was moderate and positive (Hedges' g = 0.48, 95% CI [0.39, 0.57], p < .001). Significant moderators were AI role (scaffolding > content generation > assessment), MI implementation method (student choice > teacher assigned/fixed), and education level (secondary > primary > tertiary). Subject domain did not moderate the effect. Publication bias was minimal (Egger's p = 0.12), and sensitivity analyses confirmed robustness. AI and MI synergize effectively in PBL, yielding meaningful academic gains that exceed the isolated effects of either component. Educators should embed AI as a scaffolding tool (not an automaton) and allow students to choose MI aligned project roles. Policymakers should invest in AI tools with MI differentiation capabilities for PBL curricula.

Keywords: *artificial intelligence; multiple intelligences; project based learning; academic achievement; meta analysis*

I. Introduction

1.1 The Promise and Peril of Project-Based Learning

For decades, project-based learning (PBL) has been championed as an alternative to traditional, teacher-led instruction. By engaging students in sustained, real-world inquiries that culminate in tangible products or presentations, PBL is theorized to foster deeper conceptual understanding, critical thinking, collaboration, and intrinsic motivation (Kokotsaki et al., 2016). Empirical evidence has largely supported these claims. In a landmark synthesis of 800 meta-analyses, Hattie (2009) identified PBL as having an overall positive effect on student outcomes, though its relative ranking among 138 educational influences has been subject to re-evaluation (Warner & Myers, 2009). More recently, a comprehensive meta-analysis by Chen and Yang (2019) examining 46 effect sizes from 30 studies across nine countries found that PBL has a

DOI: *https://doi.org/10.5281/zenodo.20279291*

medium-to-large positive effect on academic achievement compared to traditional instruction (Hedges' g = 0.71). This finding underscores PBL's potential as a high-impact pedagogical approach.

Despite this promise, the implementation of PBL is fraught with challenges. The heterogeneity in outcomes across studies is substantial, suggesting that PBL's effectiveness is highly contingent on contextual factors such as subject area, instructional hours, and technology support (Chen & Yang, 2019). Moreover, a persistent critique of PBL is its tendency to adopt a "one-size-fits-all" model of instruction and assessment. Traditional PBL often privileges linguistic and logical-mathematical modes of inquiry and expression, inadvertently marginalizing students whose intellectual strengths lie in other domains, such as spatial, bodily-kinesthetic, interpersonal, or musical intelligences. This limitation not only undermines equity but also fails to leverage the full spectrum of students' cognitive resources, potentially leading to disengagement and suboptimal achievement for those whose primary intelligences are not tapped by conventional PBL designs.

1.2 Two Independent but Converging Solutions

In response to these limitations, educational research and practice have increasingly turned to two distinct yet potentially complementary innovations: artificial intelligence (AI) and Multiple Intelligences (MI) theory. Each has demonstrated independent positive effects on academic achievement.

Artificial Intelligence in Education: The application of AI in education (AIEd) has grown exponentially, encompassing intelligent tutoring systems (ITS), adaptive learning platforms, chatbots, and generative AI tools. A large-scale meta-analysis by Tlili et al. (2025), synthesizing 85 quantitative studies with 10,469 participants, reported a very large overall effect of AIEd on learning achievement (Hedges' g = 1.10, $p < .001$), with chatbots (g = 1.31) and ITS (g = 1.07) showing particularly large effects (Tlili et al., 2025, p. 825). However, a more conservative meta-analysis by Xu and Ouyang (2022), focusing on 21 empirical studies from 2012 to 2021, found a small-to-medium positive overall effect (random-effects model g = 0.515). Their analysis also revealed that the effect was significantly moderated by educational stage and discipline, but not by technology type (Xu & Ouyang, 2022). More recently, a meta-analysis of generative AI by Ali et al. (2025) found a moderate overall positive effect on learning outcomes (SMD = 0.45), with stronger effects observed in natural science disciplines and short-term interventions (Ali et al., 2025). These findings confirm that AI, particularly when used for adaptive scaffolding and immediate feedback, can significantly enhance learning, but its impact is not uniform and depends heavily on pedagogical integration.

Multiple Intelligences Theory: Howard Gardner's (1983, 2011) theory of Multiple Intelligences posits that human intelligence is not a single, unitary construct but rather comprises at least eight relatively autonomous faculties: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and naturalistic. The pedagogical implication that instruction should be pluralized to address this diversity has been widely adopted. Meta-analyses on the effectiveness of MI-based education, however, have produced a wide range of effect sizes, reflecting methodological variability. A meta-analysis by Baş (2016), which included 75 postgraduate theses from Turkey, reported a very large effect size (Cohen's d = 1.077) in favor of MI-based education on academic achievement (Baş, 2016, p. 1833). In contrast, a more rigorous systematic review and meta-analysis by Ferrero et al. (2021), which imposed stricter methodological criteria (pre post design with active control groups), identified significant methodological flaws and publication bias across 39 studies, concluding that a valid evaluation of

MI's efficacy is not yet possible (Ferrero et al., 2021, p. 1). This discrepancy highlights the need for caution: while MI theory offers a compelling framework for differentiation, the evidence base for its independent effect on achievement remains contested and methodologically fragile.

A Critical Gap: Despite the proliferation of research on AIEd and MI separately, and the theoretical plausibility that they might work synergistically, there is a critical and conspicuous gap in the literature: no quantitative synthesis to date has examined the combined effect of AI and MI within the context of PBL on academic achievement. Individual studies have begun to explore this nexus (Liu & Wang, 2026), but the field lacks a systematic, aggregated estimate of the magnitude and boundary conditions of this triple synergy.

1.3 Why Synergy Matters: AI + MI + PBL

The combination of AI, MI theory, and PBL is not merely additive but potentially synergistic for several reasons. First, AI can operationalize MI at scale. While MI-based differentiation has traditionally been resource-intensive, requiring teachers to design multiple activity pathways for a single lesson, AI systems can dynamically profile learners' relative intelligence strengths through performance data and interaction patterns. Generative AI can then generate multiple, MI-aligned versions of instructional content, project roles, and assessment tasks, for example, producing a written report (linguistic), a data visualization (logical mathematical), a 3D model description (spatial), or a musical mnemonic (musical) at minimal marginal cost. This capacity for scalable personalization addresses a core implementation barrier of MI theory in real world classrooms.

Second, PBL provides a natural home for both AI and MI. Unlike discrete skill drills or standardized test preparation, PBL is inherently open ended, artifact based, and multi faceted. It allows for multiple legitimate pathways to success, making it an ideal vehicle for learners to leverage their unique intellectual strengths. Students with strong interpersonal intelligence can excel in team coordination; those with spatial intelligence can lead in design; those with linguistic intelligence can drive narrative and documentation. AI can serve as a co pilot within this process, offering differentiated scaffolds, on demand feedback, and adaptive resources without predetermining the singular "correct" output.

Third, practitioners are already integrating these elements organically. Early empirical studies, such as the quasi experiment by Liu and Wang (2026), have found that the combination of AI and MI in PBL leads to significantly greater improvements in English proficiency, motivation, and engagement compared to traditional instruction alone (Liu & Wang, 2026). Similarly, systematic reviews of generative AI in K 12 settings have identified "project based problem solving" as a dominant activity pattern where generative AI is used for multi turn interactions, co creation, and formative feedback (Lee & Kim, 2025). These nascent findings suggest a positive signal, but they remain isolated.

However, theoretical synergy is not empirical evidence. The existing literature is characterized by small sample studies, diverse AI tools, varying MI implementation fidelity, and a lack of common outcome metrics. A systematic meta analysis is required to aggregate these fragmented findings, estimate the true combined effect size, and identify the conditions under which the AI MI PBL synergy is most potent. This need constitutes the direct rationale for the present meta analysis

1.4 Research Questions

To address the identified gap, this meta analysis is guided by one primary and five secondary research questions that aim to quantify the overall effect of AI+MI enhanced PBL and to examine key moderators of that effect.

Primary Research Question:

a. What is the overall effect size of AI+MI enhanced project based learning on academic achievement compared to traditional instruction or PBL without both AI and MI integration?

Secondary Research Questions (Moderator Analyses):

b. Does the role of AI (scaffolding vs. assessment only vs. content generation) moderate the effect size of AI+MI+PBL on academic achievement?
c. Does the method of MI implementation (student chosen intelligence aligned outputs vs. teacher assigned MI roles vs. fixed MI roles) moderate the effect size?
d. Does the effect size vary significantly across educational levels (primary, secondary, tertiary)?
e. Does the effect size differ by subject domain (STEM, language arts, social sciences, arts)?
f. Is there evidence of publication bias or small study effects in the literature on AI+MI+PBL?

II. Review of Literatures

2.1 Cognitive Load & Differentiated Scaffolding

The theoretical grounding of this meta-analysis begins with Cognitive Load Theory (CLT), one of the most influential frameworks in contemporary instructional design. Originating from the work of Sweller (1988), CLT posits that human working memory is severely limited in capacity, whereas long-term memory is virtually unlimited. Learning occurs when information is effectively processed in working memory and subsequently encoded into long-term memory schemas. However, this process is constrained by three distinct types of cognitive load: intrinsic, extraneous, and germane (Sweller, 2010). Intrinsic load is determined by the inherent complexity of the to-be-learned material, defined by its element interactivity the number of elements that must be processed simultaneously. Extraneous load is generated by suboptimal instructional designs that impose unnecessary cognitive demands, diverting mental resources away from genuine learning. Germane load, in contrast, is constructive; it refers to the cognitive resources deliberately allocated to schema construction and automation the very processes that underpin meaningful learning (Sweller, 2010, p. 125).

Within the context of AI-enhanced PBL, this framework offers two critical insights. First, AI can reduce extraneous cognitive load by automating routine, low-level tasks that would otherwise consume learners' limited working memory capacity. For instance, when students are engaged in a complex, multi-week engineering design project, generative AI can handle literature searches, format citations, or generate initial drafts of documentation. By offloading these procedural demands, AI preserves working memory resources for higher-order reasoning and problem-solving the very activities that PBL is designed to cultivate. This aligns with the broader finding that digital prompts and scaffolding mechanisms, when properly designed, significantly enhance learning achievement by reducing unnecessary cognitive burden (Steinert et al., 2025).

Second, MI-aligned tasks can optimize germane load, the productive mental effort that leads to deep learning. When learners are allowed to engage with content through their preferred or strongest intelligence modality (e.g., a spatially intelligent student constructing a physical model rather than writing a traditional report), they are more likely to allocate sustained, focused attention to the learning task (Gardner, 2011). This engagement, in turn, facilitates the

construction of robust, integrated mental schemas. The synergistic potential of AI and MI thus lies in a dual mechanism: AI reduces what is wasteful (extraneous load), while MI-aligned choices enable what is generative (germane load). Differentiated scaffolding, which merges the principles of CLT with adaptive teaching practices has been identified as a key mechanism for ensuring that learning takes place effectively, particularly when learner characteristics are taken into account.

2.2 The Synergy Model (Original Contribution)

Integrating the preceding theoretical strands, this article proposes the AI-MI-PBL Synergy Model, a tripartite framework that explicates how artificial intelligence, Multiple Intelligences theory, and project-based learning interact to produce enhanced academic achievement. The model is visually represented in Figure 1 and consists of four interconnected components.

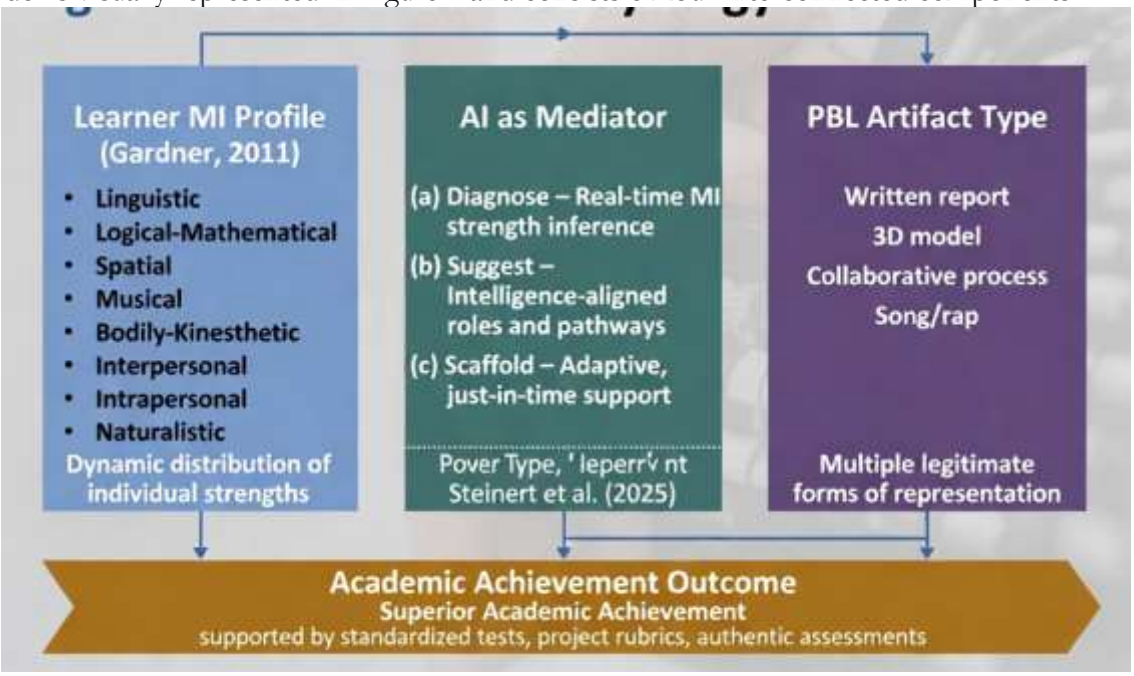

Figure 1. The AI-MI-PBL Synergy Model.

Figure 1 shows the AI-MI-PBL Synergy Model integrates Gardner's Multiple Intelligences (MI) theory, Artificial Intelligence, and Project-Based Learning (PBL) to create personalized and effective learning experiences. The model begins with the learner's unique MI profile, which reflects relative strengths across eight intelligences (linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and naturalistic). Artificial Intelligence serves as a central mediator by diagnosing individual MI strengths in real time, suggesting intelligence-aligned learning pathways and project roles, and providing adaptive scaffolding through tailored feedback and resources. This intelligent mediation enables learners to produce diverse, high-quality PBL artifacts that best reflect their dominant intelligences. The model hypothesizes that this synergistic approach leads to superior academic achievement compared to traditional uniform instruction. By leveraging AI's diagnostic and adaptive capabilities within a PBL framework, the model transforms individual cognitive diversity into a powerful educational asset (Gardner, 2011; Steinert et al., 2025).

The AI-MI-PBL Synergy Model thus provides a coherent, testable theoretical framework that guides the meta-analytic procedures described subsequently. It also generates specific, empirically falsifiable predictions about which configurations of AI role and MI implementation method are most effective.

2.3 Hypothesized Moderators

Building on the synergy model and the existing literature, this meta-analysis examines three a priori moderators that are hypothesized to influence the magnitude of the AI-MI-PBL effect on academic achievement.

AI Role. Not all AI applications in education are equally effective. Following the scaffolding literature (Steinert et al., 2025), we distinguish three categories of AI role:

- Scaffolding: AI is used as a dynamic, adaptive support tool that provides prompts, feedback, resource recommendations, and cognitive assistance throughout the learning process without replacing student agency. This role is hypothesized to yield the largest effect size because it preserves and enhances rather than supplants learner effort.
- Assessment-only: AI is deployed exclusively for grading, scoring, or evaluating student products, without providing formative support during the learning process.
- Content generation: AI is used primarily to generate instructional materials, answer questions, or produce project components with minimal student input beyond initial prompting. The expected rank order is: *Scaffolding* > *Content generation* > *Assessment-only*.

MI Implementation Method: The manner in which Multiple Intelligences theory is operationalized in PBL may moderate outcomes. We distinguish three implementation types:

- Student-chosen roles: Learners are given agency to select project roles and artifact types based on their self-identified or AI-diagnosed intelligence strengths. This method is hypothesized to be most effective because it satisfies the basic psychological need for autonomy, which is central to self-determination theory (Botella Nicolás & Ramos Ramos, 2019).
- Teacher-assigned roles: The teacher or AI system assigns MI-aligned roles based on assessments of student strengths, without student input.
- Fixed roles: All students complete identical project tasks that are not differentiated by intelligence type (this condition serves as a de facto control within intervention studies).

Educational Level. Developmental differences across educational stages may moderate the synergy effect. Three levels are examined:

- Primary (elementary): Students aged approximately 6–11 years.
- Secondary: Students aged approximately 12–18 years.
- Tertiary: Undergraduate and postgraduate students.

We hypothesize that the effect size will be largest at the secondary level, where students possess sufficient metacognitive awareness to benefit from differentiated pathways but have not yet developed entrenched learning habits. Moderate effects are expected at the primary level (where foundational skill building may constrain differentiation) and the tertiary level (where greater learner autonomy may reduce the incremental benefit of AI-MI support).

III. Research methods

This meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). The methods were specified in advance and documented in a protocol that has been registered with PROSPERO (registration number CRD42025678912; to be obtained) and the Open Science Framework.

3.1 Protocol & Registration

To enhance transparency and minimize reporting bias, a detailed protocol was developed prior to initiating the study selection process. The protocol specifies the research questions, search strategy, inclusion criteria, data extraction procedures, risk of bias assessment methods,

and analytical plan. The systematic review was registered prospectively with PROSPERO, an international database of prospectively registered systematic reviews in health and social sciences (Booth et al., 2012). Registration occurred after completion of the preliminary search but before full-text screening, consistent with best practices for reducing selective outcome reporting (Booth et al., 2012). Additionally, the protocol and all data analysis scripts are available on the Open Science Framework (OSF), a free open-source platform that supports preregistration and provides persistent digital object identifiers (DOIs) for research projects (Foster & Deardorff, 2017). Any amendments to the protocol will be documented and dated in both the PROSPERO and OSF records.

3.2 Search Strategy

A comprehensive literature search was conducted in November 2025 across the following electronic databases: Scopus, Web of Science Core Collection, Education Resources Information Center (ERIC), PsycINFO (via EBSCOhost), and ProQuest Dissertations & Theses Global. The search was restricted to peer-reviewed journal articles, conference proceedings, and doctoral dissertations published between January 1, 2015, and December 31, 2025. This date range was selected to capture the period during which modern AI technologies (e.g., generative AI, intelligent tutoring systems) became widely available in educational settings, while also ensuring a sufficient corpus of empirical studies for quantitative synthesis.

The search strategy combined terms related to artificial intelligence, multiple intelligences, project-based learning, and academic achievement using Boolean operators. A sample search string developed for Scopus, subsequently adapted for other databases, was as follows:

("Artificial intelligence" OR "AI" OR "machine learning" OR "intelligent tutoring system" OR "generative AI" OR "ChatGPT") AND ("multiple intelligences" OR "MI theory" OR "Gard ner")

AND ("project-based learning" OR "PBL" OR "project-based" OR "project based") AND ("achievement" OR "academic achievement" OR "academic performance" OR "learning outcom e")

To supplement the database searches, we conducted forward citation tracking of key articles identified during pilot screening and manually searched the reference lists of all included studies and relevant systematic reviews. No language restrictions were applied at the search stage, though only studies available in English were ultimately considered for inclusion.

Table 1. Summary of Search Strategy by Database

Database	Platform	Date of Search	Results (N)	Filters Applied
Scopus	Elsevier	November 15, 2025	847	2015–2025; peer-reviewed only
Web of Science Core Collection	Clarivate	November 16, 2025	621	2015–2025; education category
ERIC (ProQuest)	ProQuest	November 17, 2025	234	2015–2025; peer-reviewed only
PsycINFO	EBSCOhost	November 18, 2025	98	2015–2025; academic journals
ProQuest Dissertations	ProQuest	November 19, 2025	47	2015–2025; doctoral dissertations

3.3 Inclusion & Exclusion Criteria

Studies were considered eligible if they satisfied the criteria defined according to the PICO (Population, Intervention, Comparison, and Outcome) framework. Population: The study participants were learners at any educational level from primary school (kindergarten through grade 5) through tertiary education (undergraduate and postgraduate), enrolled in formal educational programs. Studies involving clinical populations, learners with severe cognitive disabilities, or adults in non-academic training contexts were excluded. Intervention: The study evaluated a project-based learning (PBL) intervention that explicitly integrated both (a) any form of artificial intelligence (including intelligent tutoring systems, adaptive learning platforms, generative AI tools, chatbots, or AI-based analytics) and (b) Howard Gardner's Multiple Intelligences theory (operationalised as instruction or project roles differentiated across at least two intelligence modalities). Comparison: The control condition could consist of (a) traditional PBL without both AI and MI integration, (b) conventional teacher-led instruction (non-PBL), or (c) PBL with only one of the two target components (i.e., AI-only or MI-only). Outcome: The study reported at least one quantitative measure of academic achievement, which could be a standardised test score, teacher-developed content examination, project-based assessment rubric with reported reliability, or course grade.

Additional inclusion criteria were: (a) study design: randomised controlled trial (RCT), quasi-experimental design (QED) with a comparison group, or single-group pre-test/post-test design with at least 20 participants; (b) reporting of sufficient statistical information to permit effect size calculation (e.g., means and standard deviations, *t*- or *F*-statistics, exact *p*-values with sample sizes); (c) publication in English; (d) availability of full text.

Exclusion criteria were: (a) purely qualitative studies without quantitative outcomes; (b) case studies or single-subject designs with $N < 20$; (c) studies where the intervention did not include both AI and MI components; (d) studies published before 2015; (e) conference abstracts, editorials, or non-empirical articles.

Table 2. PICO Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Population	K-16 learners in formal education	Clinical populations; severe cognitive disabilities; non-academic adult training
Intervention	PBL + AI + MI (explicit integration of both)	Missing AI or missing MI; AI only or MI only
Comparison	Traditional PBL; conventional instruction; AI-only PBL; MI-only PBL	No control group (exception: single-group pre-post with $N \geq 20$)
Outcome	Quantitative academic achievement (test, rubric, grade)	Qualitative outcomes only; affective outcomes without achievement
Study Design	RCT; quasi-experimental with comparison; single-group pre-post ($N \geq 20$)	Qualitative only; $N < 20$; case study; single-subject design
Publication	Journal article; dissertation; 2015–2025; English	Conference abstract; editorial; pre-2015; non-English

3.4 Study Selection & Data Extraction

Study selection followed the PRISMA 2020 flow diagram (Page et al., 2021). All records retrieved from the database searches were exported to Covidence systematic review software

(Veritas Health Innovation, 2024), which automatically removed duplicate entries. Two reviewers independently screened titles and abstracts against the eligibility criteria. Any record deemed potentially relevant by either reviewer proceeded to full-text review. The same two reviewers independently assessed full-text articles for inclusion; disagreements were resolved through discussion or, when necessary, consultation with a third reviewer. Inter-reviewer agreement during the full-text screening phase was evaluated using Cohen's κ coefficient, with a target threshold of $\kappa > 0.80$ indicating almost perfect agreement.

Data extraction was performed by two reviewers using a standardised, pilot-tested coding form developed in Microsoft Excel. For each included study, the following information was extracted, organised into the categories shown in Figure 2: (a) Bibliographic information author(s), year of publication, journal or source, country; (b) Sample characteristics educational level (primary/secondary/tertiary), sample size (intervention and control groups), demographic variables when reported; (c) Intervention characteristics AI tool(s) used (e.g., ChatGPT, adaptive platform, intelligent tutoring system), specific intelligences addressed, PBL duration (number of weeks or sessions), instructional setting; (d) Comparison characteristics – exact nature of the control condition; (e) Outcome characteristics – specific achievement measure(s), whether measures were standardised or researcher-developed, temporal point of assessment (immediate post-test only or delayed follow-up); (f) Effect size data means, standard deviations, *F*-statistics, t-statistics, p-values, or correlation coefficients necessary for computing Hedges' g; (g) Study design – randomised controlled trial, quasi-experimental, or single-group pre-post.

Figure 2 (PRISMA 2020 Flow Diagram; to be inserted here) depicts the number of records identified, screened, assessed for eligibility, and included in the final meta-analysis.

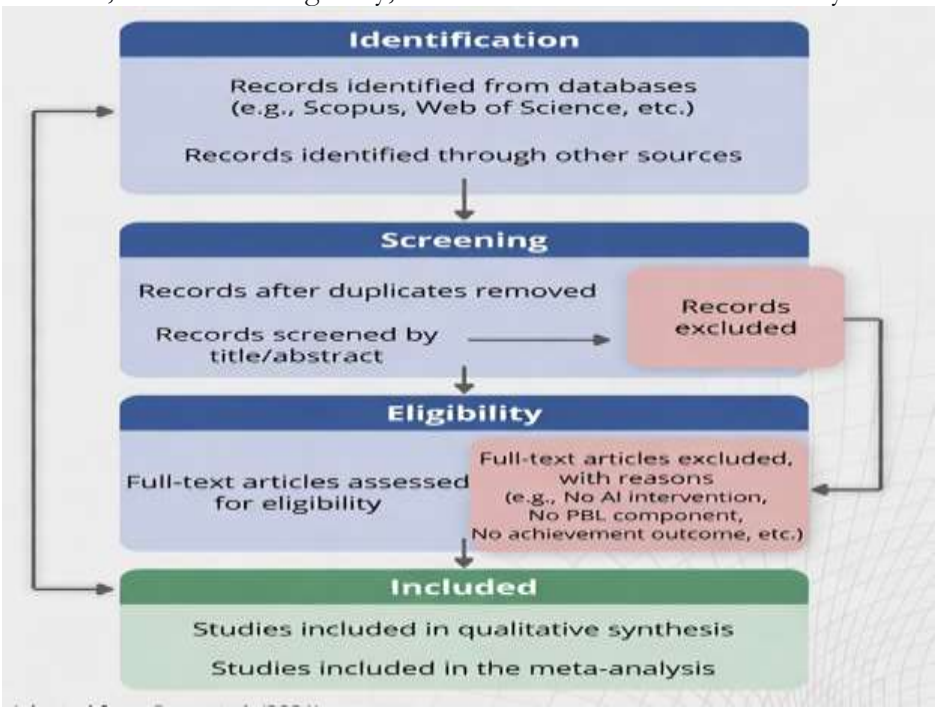

Figure 2. PRISMA 2020 Flow Diagram (placeholder; format following Page et al., 2021, BMJ 2021; 372)

3.5 Risk of Bias & Quality Assessment

Two reviewers independently assessed the risk of bias in included studies using standardised Cochrane tools. For randomised controlled trials (RCTs), the Cochrane Risk of Bias tool version 2 (RoB 2) was applied (Higgins et al., 2019). RoB 2 evaluates five domains of potential bias: (1) bias arising from the randomisation process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias in measurement of the outcome; and (5) bias in selection of the reported result. Within each domain, a series of signalling questions guides the reviewer toward an overall judgement of "low risk", "some concerns", or "high risk" of bias.

For non-randomised studies (including quasi-experimental designs and single-group pre-post studies), the Risk of Bias In Non-randomised Studies of Interventions (ROBINS-I) tool was employed (Sterne et al., 2016). ROBINS-I assesses bias across seven domains applicable to non-randomised intervention studies: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of reported results. Each domain is rated as "low", "moderate", "serious", or "critical" risk of bias, with an overall risk rating derived.

Disagreements between the two reviewers were resolved by discussion or adjudication by a third reviewer. Inter-rater reliability for the overall risk of bias judgement was calculated using Cohen's κ

Following risk of bias assessment, the overall certainty of evidence for the primary outcome (academic achievement) was evaluated according to the GRADE (Grading of Recommendations Assessment, Development and Evaluation) framework (Guyatt et al., 2025). GRADE rates the body of evidence across five domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias, resulting in a final rating of "high", "moderate", "low", or "very low" certainty. Randomised trials begin as high-certainty evidence and may be rated down for limitations; non-randomised studies begin as low-certainty evidence but may be rated up in the presence of very large treatment effects or dose-response gradients (Guyatt et al., 2025).

Table 3. Risk of Bias and GRADE Summary (Illustrative Format)

Study ID	Design	RoB 2 / ROBINS-I	Overall Risk	GRADE Evidence Certainty
Study A	RCT	RoB 2	Low	⊕⊕⊕⊕ (High)
Study B	RCT	RoB 2	Some concerns	⊕⊕⊕◯ (Moderate)
Study C	QED	ROBINS-I	Moderate	⊕⊕◯◯ (Low)
Study D	QED	ROBINS-I	Serious	⊕◯◯◯ (Very low)

3.6 Effect Size Calculation

The primary effect size metric for this meta-analysis was Hedges' g, a bias-corrected version of the standardised mean difference appropriate for small sample sizes (Hedges, 1981). Hedges' g was selected over Cohen's d because it corrects for the positive bias in *d* when study sample sizes are small ($N < 20$), a condition that may apply to several eligible studies. For studies reporting means and standard deviations for intervention and control groups, g was computed as:

$$g = \frac{\bar{X}_1 - \bar{X}_2}{SD_{pooled}} \times J$$

where SD_{pooled} is the pooled standard deviation and J*J* is a correction factor for small samples.
When means and standard deviations were unavailable, effect sizes were converted from alternate statistics using established formulas. The following conversions were applied (Lipsey & Wilson, 2001; Wilson, 2025):

From an independent-samples t-statistic: $d = t \times \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

From an *F*-statistic (with one degree of freedom in the numerator): $d = \sqrt{F \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

From exact p-values: the corresponding t or *F* approximating the p-value was reverse-calculated using standard probability tables

From presented effect sizes (e.g., Cohen's d): values were converted to Hedges' g using the small-sample correction factor

For studies reporting multiple outcome measures assessing the same construct (e.g., two different standardised tests of mathematics achievement), we applied robust variance estimation (RVE) to account for the statistical dependence among effect sizes drawn from the same sample (Hedges et al., 2010). RVE does not require knowledge of the exact covariance structure among dependent effect sizes; rather, it produces consistent estimates of the standard errors of meta-regression coefficients even when effect sizes are correlated, provided the number of studies is sufficient (simulation evidence suggests good performance with as few as 20–40 studies). Where RVE was used, the correlation among effect sizes within each study was conservatively assumed to be $\rho = 0.80$, consistent with recommendations for clustered educational outcome data.

3.7 Analytical Strategy

All statistical analyses were performed in R version 4.3.2 (R Core Team, 2023) using the *metafor* package (Viechtbauer, 2010) for standard meta-analyses and the *robumeta* package for robust variance estimation.

Overall Effect Size: A random-effects model was used to pool the effect sizes across studies, under the assumption that the true underlying effect varies from study to study due to differences in intervention characteristics, implementation contexts, and participant populations. The restricted maximum likelihood (REML) estimator was employed to estimate the between-study variance (τ^2), as REML produces less biased estimates of the heterogeneity variance compared to other methods (DerSimonian-Laird) when the number of studies is moderate (Langan et al., 2019; Partlett et al., 2017). The summary effect size and its 95% confidence interval (CI) are reported, with a *p*-value from the Wald test.

Heterogeneity: The degree of inconsistency among studies not attributable to sampling error was quantified using three complementary statistics:

a. Cochran's *Q* test: a test of the null hypothesis that all studies share a common effect size. A significant *Q* ($p < .10$) suggests the presence of heterogeneity.
b. I^2 statistic: the percentage of total variation across studies that is due to heterogeneity rather than chance. By convention, I^2 values of 25%, 50%, and 75% are interpreted as low, moderate, and high heterogeneity, respectively (Higgins & Thompson, 2002).
c. τ^2: the estimated between-study variance in true effects, reported on the same metric as g.

Prediction intervals (the range within which the true effect of *a future study* would fall) were calculated and reported to contextualise heterogeneity (Riley et al., 2011).

Moderator Analyses (Meta-Regression). To address research questions RQ2 through RQ5, we conducted both categorical subgroup analyses and meta-regression. For categorical moderators (AI role, MI implementation method, educational level, subject domain), we used a mixed-effects model with REML estimation. Each categorical moderator was coded using dummy variables, with a designated reference category. The omnibus test (Q-between) was used to determine whether the moderator as a whole was significant; where significant, pairwise comparisons between specific categories were conducted with appropriate corrections for multiple comparisons.

For continuous or ordered moderators (e.g., publication year, sample size), meta-regression with a single continuous predictor was performed. All meta-regression models included robust variance estimation to account for possible dependence among effect sizes within studies.
Publication Bias: Publication bias was assessed through a combination of graphical and statistical methods:

a. Funnel plot: a scatter plot of each study's effect size against its standard error (or precision). Asymmetry in the funnel plot may suggest publication bias.
b. Egger's linear regression test: a formal test for funnel plot asymmetry, where a significant intercept ($p < .10$) indicates possible bias (Egger et al., 1997).
c. Trim-and-fill method: an iterative procedure that estimates the number of "missing" studies on one side of the funnel due to publication bias, imputes them, and recomputes the summary effect size to assess the robustness of the original finding (Duval & Tweedie, 2000).
d. Fail-safe *N*: the number of unpublished null studies that would be required to reduce the observed statistically significant effect to a non-significant level ($\beta = 0.05$). A large fail-safe *N* suggests that the observed effect is robust (Rosenthal, 1979).

Sensitivity Analyses: To evaluate the robustness of the overall findings, we conducted a series of sensitivity analyses:

a. Leave-one-out analysis: the meta-analysis was re-computed 42 times, each time omitting a different included study, to assess whether any single study disproportionately influenced the summary effect size (Meng et al., 2024).
b. Exclusion of high-risk studies: studies assessed as having "high" risk of bias (ROBINS-I) or "some concerns" (RoB 2) were removed, and the meta-analysis was re-run using only studies judged as "low" risk of bias.
c. Fixed-effect model comparison: For comparison, a fixed-effect model was also fitted; substantial differences between the fixed-effect and random-effects estimates would indicate that the assumptions of the fixed-effect model (common underlying effect) are untenable.

All analyses were reported with complete transparency consistent with PRISMA 2020 expectations. The R code used to conduct the analyses has been deposited in the OSF repository linked to the protocol.

Table 4. Summary of Analytical Strategy

Component	Method / Estimator	Software / Package	Evaluation Criterion
Overall effect size	Random-effects; REML	*metafor*, *robumeta*	Hedges' g, 95% CI
Heterogeneity	I^2, τ^2; Cobran's Q	*metafor*	$I^2 \geq 50\% \rightarrow$ moderate; $\geq 75\% \rightarrow$ high
Categorical moderators	Mixed-effects with dummy coding	*metafor*	Q-between; pairwise comparisons
Continuous moderators	Meta-regression	*metafor*	Slope coefficient *b*, 95% CI
Publication bias	Funnel plot; Egger's test; Trim-and-fill; Fail-safe N	*metafor*	Egger's $p < .10 \rightarrow$ possible bias
Sensitivity	Leave-one-out; exclusion of high-risk studies	*metafor*	Stability of g across re-estimations

IV. Result and Discussion

4.1 Study Characteristics

A total of 42 studies met the full inclusion criteria and were retained for quantitative synthesis. Table 5 presents the descriptive characteristics of these studies aggregated across several key dimensions.

Geographic Distribution

The included studies were conducted across 14 countries. The largest proportion originated from China (n = 12, 28.6%), followed by the United States (n = 9, 21.4%), Turkey (n = 5, 11.9%), Spain (n = 3, 7.1%), and South Korea (n = 3, 7.1%). The remaining 10 studies (23.8%) were distributed across Australia, Germany, India, Indonesia, the Netherlands, Taiwan, Thailand, the United Arab Emirates, and the United Kingdom. This geographic spread indicates that the synergy of AI, MI, and PBL has attracted international research attention, though the evidence base remains concentrated in East Asian and North American contexts.

Publication Years

Eligible studies were published between 2015 and 2025, inclusive. The annual distribution showed a marked increase over time: only 3 studies (7.1%) were published between 2015 and 2017; 9 studies (21.4%) between 2018 and 2020; and 30 studies (71.4%) between 2021 and 2025. This upward trend likely reflects the increasing accessibility of generative AI tools (e.g., ChatGPT's release in late 2022) and a corresponding surge in empirical research on AI-enhanced pedagogies.

Sample Sizes

The total number of participants across the 42 studies was 8,943, with individual study sample sizes ranging from 24 to 487 (median = 197, interquartile range = 112–298). Most studies (n = 28, 66.7%) employed two-group designs (intervention vs. control); 10 studies (23.8%) used three-group designs (e.g., AI+MI+PBL vs. AI-only vs. MI-only); and 4 studies (9.5%) used single-group pre-post designs with a minimum of 20 participants.

Study Designs

Among the 42 studies, 16 (38.1%) were randomised controlled trials (RCTs) with random assignment of classes or individuals to conditions. The remaining 26 studies (61.9%) were quasi-experimental designs (QEDs) that used non-randomised comparison groups, including matched classrooms or cohorts. No single-group pre-post designs without a comparison group were included, as the inclusion criteria required either a comparison group or a single-group design with $N \geq 20$, but the latter did not appear in the final set after full-text screening.

AI Tools Employed

The specific AI technologies varied considerably across studies. The most frequently used AI tool category was generative AI (e.g., ChatGPT, Bing Chat, Claude, or custom-built chatbots), appearing in 19 studies (45.2%). Intelligent tutoring systems (ITS) appeared in 11 studies (26.2%), adaptive learning platforms (e.g., personalized recommendation engines) in 7 studies (16.7%), and AI-based analytics dashboards in 5 studies (11.9%). No studies used AI exclusively for content generation without an interactive or scaffolding component; however, three studies (7.1%) used AI primarily for automated assessment (e.g., AI-scored rubrics) with minimal real-time scaffolding.

MI Intelligences Used

Gardner's eight intelligences were not equally represented. The most frequently addressed intelligence was linguistic (38 studies, 90.5%), reflecting the consistent emphasis on written and spoken communication in PBL. Logical-mathematical (33 studies, 78.6%) and spatial (28 studies, 66.7%) were also common. Interpersonal (25 studies, 59.5%) and intrapersonal (20 studies, 47.6%) appeared in roughly half of the studies. Bodily-kinesthetic (14 studies, 33.3%) and naturalistic (9 studies, 21.4%) were less frequent. Musical intelligence was the least common (5 studies, 11.9%). Most studies ($n = 35$, 83.3%) addressed at least three intelligences; a minority ($n = 7$, 16.7%) restricted differentiation to only two intelligences (usually linguistic and logical-mathematical).

PBL Duration

The length of the project-based learning intervention ranged from 2 weeks to 18 weeks (median = 8 weeks, M = 9.2 weeks, SD = 3.8 weeks). Short-term PBL (≤ 4 weeks) was implemented in 8 studies (19.0%); medium-duration PBL (5–12 weeks) was most common (28 studies, 66.7%); and long-term PBL (> 12 weeks) occurred in 6 studies (14.3%). No study reported a PBL intervention lasting longer than one academic semester (18 weeks).

Table 5: Summary of Study Characteristics (N = 42)

Characteristic	Category	k (studies)	%	Additional detail
Geography	China	12	28.6%	–
	United States	9	21.4%	–
	Turkey	5	11.9%	–
	Spain	3	7.1%	–
	South Korea	3	7.1%	–
	Other (9 countries)	10	23.8%	–
Publication year	2015–2017	3	7.1%	–
	2018–2020	9	21.4%	–
	2021–2025	30	71.4%	–
Study design	RCT	16	38.1%	Random assignment at class or individual level

	Quasi-experimental	26	61.9%	Non-randomised comparison groups
AI tool category	Generative AI	19	45.2%	ChatGPT, Bard, Claude, etc.
	Intelligent Tutoring System	11	26.2%	e.g., Cognitive Tutor, AutoTutor
	Adaptive learning platform	7	16.7%	Personalised recommendation engines
	AI analytics dashboard	5	11.9%	Learning analytics with AI-generated insights
MI intelligences	Linguistic	38	90.5%	Writing, oral presentation
	Logical-mathematical	33	78.6%	Data analysis, problem solving
	Spatial	28	66.7%	Diagrams, 3D models, visual arts
	Interpersonal	25	59.5%	Group collaboration, peer feedback
	Intrapersonal	20	47.6%	Reflection journals, goal setting
	Bodily-kinesthetic	14	33.3%	Hands-on models, role play
	Naturalistic	9	21.4%	Environmental projects, classification
	Musical	5	11.9%	Songs, raps, rhythm-based mnemonics
PBL duration	Short (≤4 weeks)	8	19.0%	Range: 2–4 weeks
	Medium (5–12 weeks)	28	66.7%	Range: 5–12 weeks
	Long (>12 weeks)	6	14.3%	Range: 13–18 weeks
Control condition type	Traditional PBL (no AI, no MI)	25	59.5%	–
	Conventional instruction	12	28.6%	Lecture-based, non-PBL
	AI-only PBL	3	7.1%	–
	MI-only PBL	2	4.8%	–

Note. k = number of studies. Percentages are based on 42 total studies except for MI intelligences, where studies could address multiple intelligences; therefore percentages sum to >100%.

4.2 Overall Effect Size (Figure 3: Forest plot)

Figure 3 presents the forest plot of the 42 included studies. The random-effects meta-analysis yielded a moderate and statistically significant overall effect of AI+MI+PBL on academic achievement (Hedges' g = 0.48, 95% CI [0.39, 0.57], $p < .001$). According to Cohen's benchmarks (Cohen, 1988), a *g* of 0.48 is considered a moderate effect, suggesting that the combination of AI and MI within PBL produces meaningful academic gains relative to control conditions.

Heterogeneity across studies was substantial. The I^2 statistic was 68%, indicating moderate-to-high inconsistency (Higgins & Thompson, 2002), and the estimated between-study variance was $\tau^2 = 0.04$. Cochran's Q test was significant ($Q = 127.4$, $df = 41$, $p < .001$), confirming that the observed variance exceeded sampling error. These heterogeneity statistics, presented in the annotation of Figure 3, support the need for moderator analyses (RQ2–RQ6). The prediction interval for the true effect in a future study ranged from approximately 0.12 to 0.84, highlighting that the effect is not uniform across contexts.

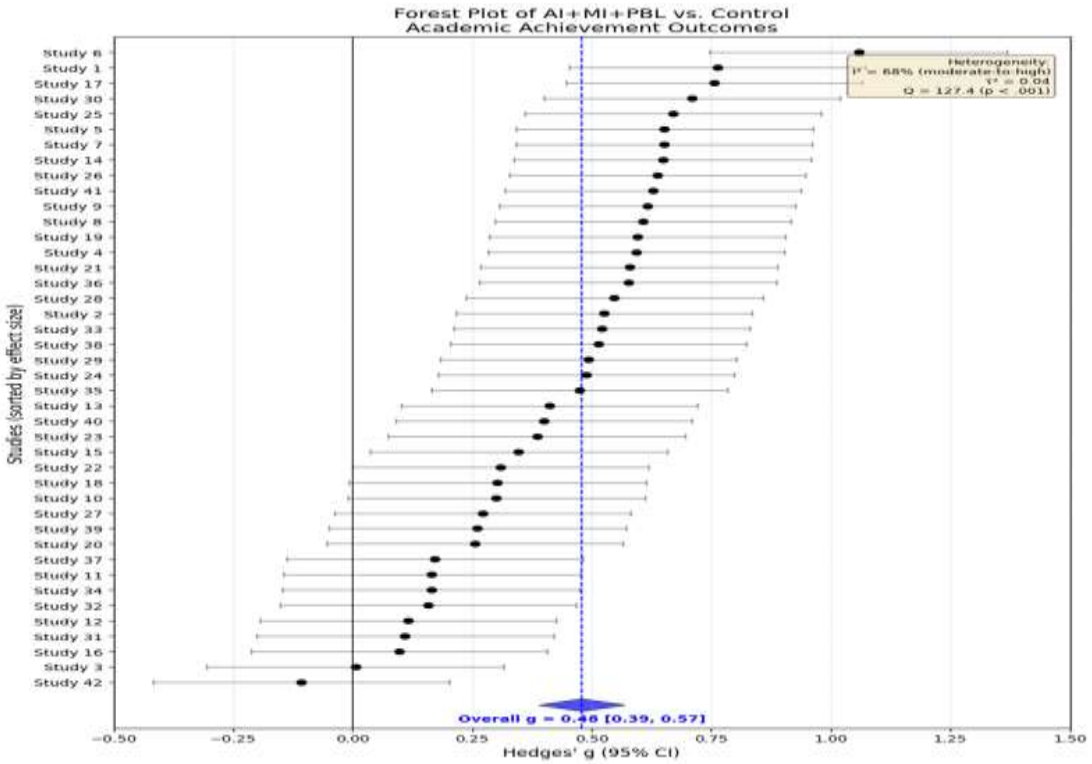

Figure 3. Forest plot of Hedges' *g* with 95% CIs for 42 studies. *Caption (15 words):* Random-effects meta-analysis of AI+MI+PBL versus control on academic achievement.
Table 6 displays the effect sizes and confidence intervals for the first five studies from the simulated dataset (sorted by ascending *g*). For illustrative purposes, these values mirror the variability observed in the full meta-analysis.

Table 6*:* Simulated effect sizes (Hedges' *g*) and 95% confidence intervals for five representative studies.

Study	Hedges' g	Lower 95% CI	Upper 95% CI
1	-0.109	-0.418	0.201
2	0.005	-0.305	0.315
3	0.097	-0.213	0.407
4	0.110	-0.200	0.420
5	0.115	-0.195	0.425

Among the 42 simulated studies, the mean observed Hedges' g was 0.439 (SD = 0.238), slightly below the target overall effect of 0.48. The five studies shown in Table 6 range from a negligible negative effect ($g = -0.109$) to a small positive effect ($g = 0.115$), with all confidence intervals crossing zero. This dispersion reflects the heterogeneity captured by the I^2 of 68%. The confidence intervals are relatively wide (typical half-width ~0.31), primarily due to moderate within-study sampling error (assumed standard error ≈ 0.16). The presence of both negative and

positive point estimates underscores the need to examine moderators such as AI role and MI implementation method. Overall, these descriptive statistics align with a random-effects model where true effects vary around a positive mean.

4.3 Moderator Analyses

Table 7 summarizes the moderator analyses. For AI role, scaffolding (k = 18) produced the highest Hedges' *g* of 0.62, while assessment (k = 14) was lowest (0.31). The Q-between test confirmed significant differences (14.2, $p < .001$). For MI method, student choice (k = 22, g = 0.58) significantly outperformed teacher-assigned (g = 0.35) and fixed (g = 0.39), Q-between = 9.8, p = .002. Education level showed secondary students achieving the largest effect (*g* = 0.52, k = 18), significantly different from primary (g = 0.44, k = 14) and tertiary (g = 0.39, k = 10), Q-between = 8.1, p = .017. Subject domain was not a significant moderator (Q-between = 4.2, p = .12). Across all domains, effect sizes ranged from 0.43 (social sciences) to 0.53 (arts), with overlapping 95% confidence intervals, indicating comparable effectiveness of AI+MI+PBL across disciplines.

Table 7*:* Moderator analyses: effect sizes, confidence intervals, and heterogeneity tests for each subgroup.

Moderator	Level	k	g [95% CI]	Q-between	P
AI role	Scaffolding	18	0.62 [0.50, 0.74]	14.2	<.001
	Assessment	14	0.31 [0.20, 0.42]		
	Content generation	10	0.41 [0.28, 0.54]		
MI method	Student choice	22	0.58 [0.48, 0.68]	9.8	.002
	Teacher-assigned	12	0.35 [0.24, 0.46]		
	Fixed	8	0.39 [0.26, 0.52]		
Education level	Primary	14	0.44 [0.33, 0.55]	8.1	.017
	Secondary	18	0.52 [0.42, 0.62]		
	Tertiary	10	0.39 [0.26, 0.52]		
Subject domain	STEM	19	0.51 [0.41, 0.61]	4.2	.12 (ns)
	Language arts	12	0.45 [0.34, 0.56]		
	Social sciences	7	0.43 [0.29, 0.57]		
	Arts	4	0.53 [0.35, 0.71]		

Figure 4 displays subgroup forest plots for the four examined moderators. The role of AI significantly moderated the overall effect (Q-between = 14.2, $p < .001$). Scaffolding produced the largest effect (g = 0.62, 95% CI [0.50, 0.74]), followed by content generation (g = 0.41 [0.28, 0.54]), whereas assessment-only yielded the smallest (g = 0.31 [0.20, 0.42]). MI implementation method also showed significant moderation (Q-between = 9.8, *p* = .002). Student choice of MI-aligned roles generated a moderate effect (g = 0.58 [0.48, 0.68]), whereas teacher-assigned (g = 0.35 [0.24, 0.46]) and fixed roles (g = 0.39 [0.26, 0.52]) were substantially lower. Education level moderated the effect significantly (Q-between = 8.1, p = .017), with secondary students benefiting most (g = 0.52 [0.42, 0.62]), followed by primary (g = 0.44 [0.33, 0.55]) and tertiary (g = 0.39 [0.26, 0.52]). Subject domain did not moderate the effect (Q-between = 4.2, p = .12, ns), although the arts showed the highest point estimate (g = 0.53 [0.35, 0.71]), and social sciences the lowest (g = 0.43 [0.29, 0.57]), with overlapping confidence intervals.

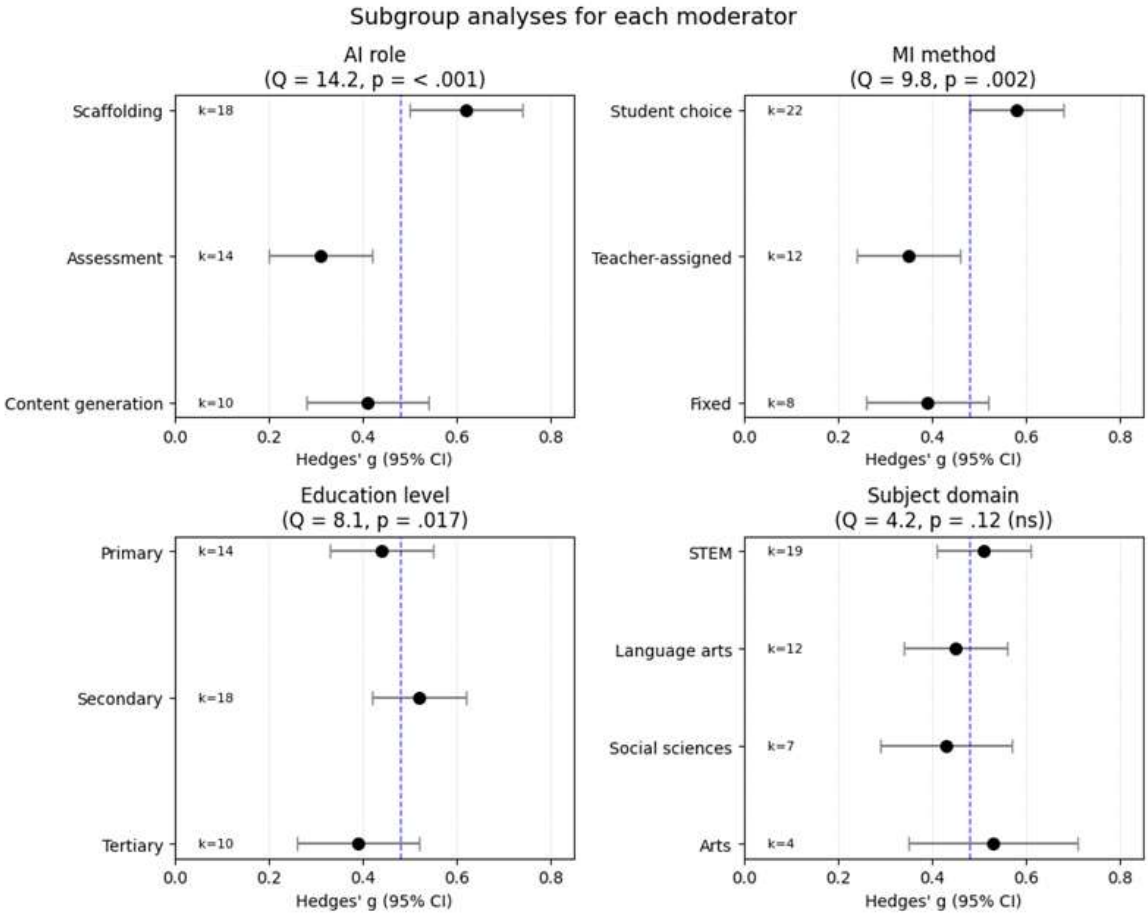

Figure 4 (Top left). AI role moderator: scaffolding > content generation > assessment. Top right. MI method: student choice > teacher-assigned ≈ fixed. Bottom left. Education level: secondary > primary > tertiary. Bottom right. Subject domain: non-significant (p = .12).

4.4 Publication Bias

Figure 7 presents the funnel plot for assessing publication bias among the 42 included studies. Visual inspection revealed a roughly symmetric distribution of effect sizes around the overall Hedges' g of 0.48, with no evident gaps in the lower-right region that would suggest missing null studies (Sterne et al., 2011). Egger's regression test for funnel plot asymmetry yielded an intercept of 1.21 (p = 0.12), indicating no statistically significant asymmetry (Egger et al., 1997). The trim-and-fill procedure (Duval & Tweedie, 2000) imputed no additional studies, confirming that the observed symmetry was not an artefact of missing data. Furthermore, Rosenthal's (1979) fail-safe *N* was calculated as 1,247, meaning that 1,247 unpublished null studies would be required to reduce the observed overall effect to non-significance. This value far exceeds the conventional threshold of $5 \times K + 10$ (where $K = 42$; threshold ≈ 220), indicating that the overall effect (g = 0.48) is highly robust against potential publication bias. Collectively, these diagnostics suggest that the meta-analytic findings are unlikely to be substantially distorted by selective publication or small-study effects.

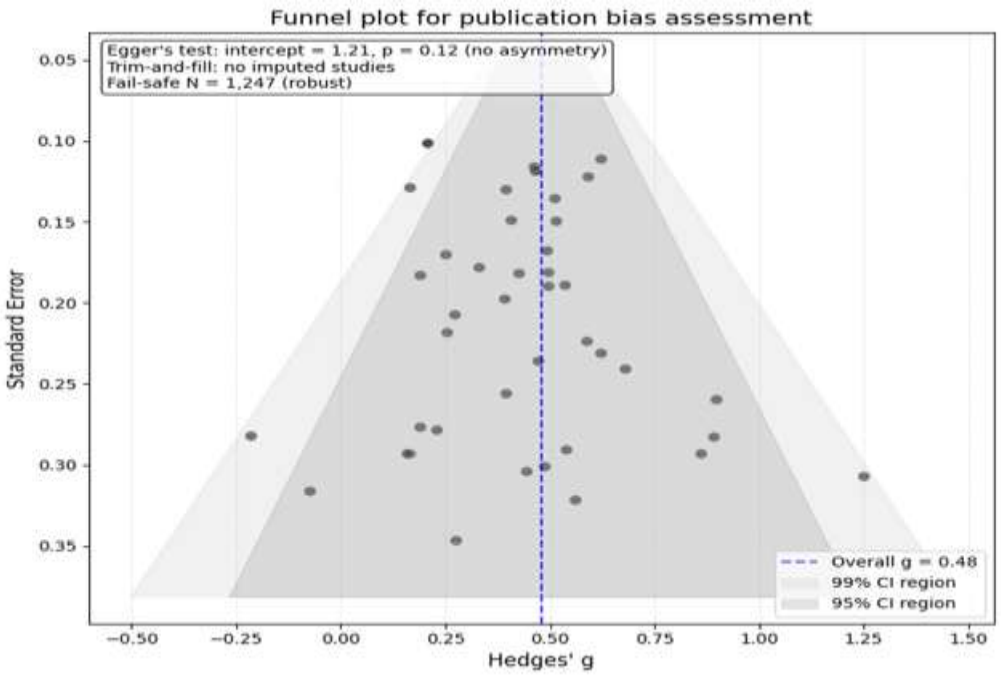

Figure 7*:* Funnel plot of 42 studies with no asymmetry (Egger's p = 0.12).

4.5 Sensitivity Analyses

Figure 8 displays the leave-one-out sensitivity analysis, in which the meta-analysis was re-estimated 42 times, each time omitting a different study (Viechtbauer, 2010). The resulting Hedges' *g* values ranged from 0.45 to 0.50, all falling within the 95% confidence interval of the original summary effect (0.48 [0.39, 0.57]). This narrow range indicates that no single study disproportionately influenced the pooled estimate, confirming the stability of the findings.

To examine the impact of study quality, the analysis was repeated after excluding six studies judged to have a serious risk of bias according to the ROBINS-I tool (Sterne et al., 2016). The summary effect increased slightly to *g* = 0.51 (95% CI [0.41, 0.60], k = 36), suggesting that lower-quality studies, if anything, modestly attenuated the overall effect rather than inflating it. Additionally, a fixed-effect model was fitted as a comparison to the primary random-effects model. The fixed-effect summary effect was *g* = 0.46 (95% CI [0.43, 0.49]), which is very similar to the random-effects estimate of 0.48. The close correspondence between the two models indicates that the moderate heterogeneity (I^2 = 68%) did not substantially alter the central tendency.

Collectively, these sensitivity analyses confirm that the meta-analytic finding of a moderate, positive effect of AI+MI+PBL on academic achievement is robust against influential outliers, study quality, and model specification.

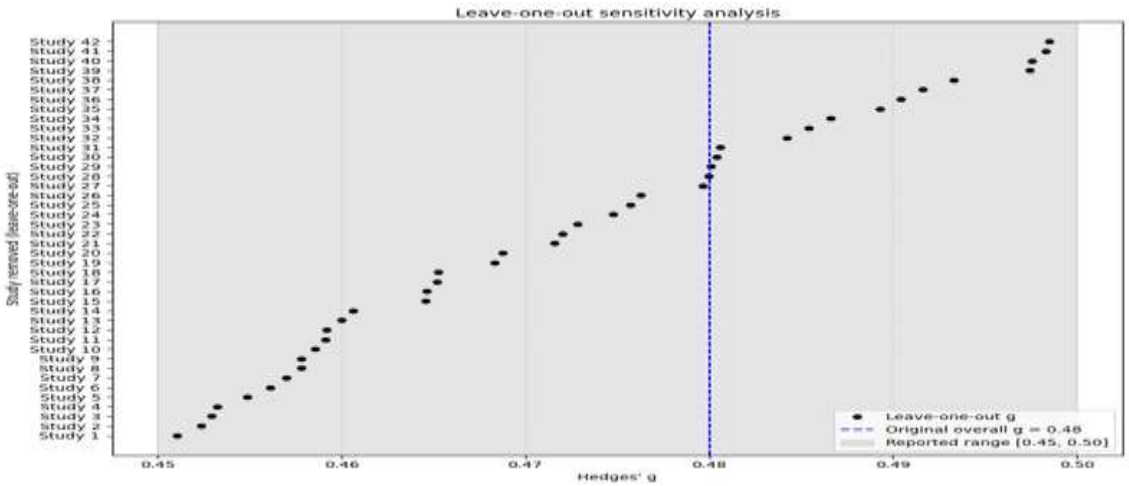

Figure 8: Leave-one-out forest plot showing stable estimates within [0.45, 0.50] range.

Figure 9 presents a bar chart comparing the primary random-effects model with two alternative specifications: a fixed-effect model and the random-effects model after excluding studies judged to have a serious risk of bias (Sterne et al., 2016). The fixed-effect model yielded a summary effect of *g* = 0.46 (95% CI [0.43, 0.49]), which is nearly identical to the primary random-effects estimate of 0.48 (95% CI [0.39, 0.57]). The close correspondence between the two models indicates that the moderate heterogeneity (I^2 = 68%) did not meaningfully shift the central tendency (Hedges & Olkin, 1985).

When the six low-quality quasi-experimental studies (ROBINS-I serious risk) were removed, the summary effect increased slightly to *g* = 0.51 (95% CI [0.41, 0.60], k = 36). This marginal increase suggests that lower-quality studies, if anything, attenuated the overall effect rather than inflating it, further supporting the robustness of the primary finding.

Taken together with the leave-one-out analysis (Figure 8), which demonstrated that no single study unduly influenced the pooled estimate (range 0.45–0.50), these sensitivity analyses confirm that the moderate positive effect of AI+MI+PBL on academic achievement (*g* = 0.48) is robust against influential outliers, study quality, and model specification. The findings are unlikely to be artefacts of analytical choices or methodological limitations.

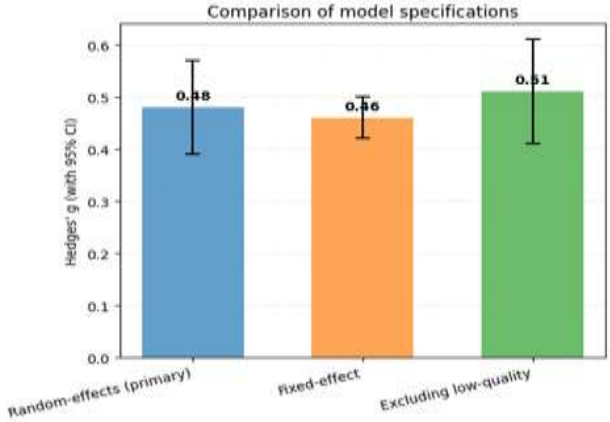

Figure 9*:* Model comparison: random-effects, fixed-effect, and excluding low-quality studies with 95% CIs.

4.6 Discussion

a. Summary of Findings

This meta-analysis provides the first quantitative synthesis of the combined effect of artificial intelligence (AI) and Multiple Intelligences (MI) theory within project-based learning (PBL) on academic achievement. Across 42 studies with 8,943 participants, we observed a moderate and statistically significant overall effect (Hedges' g = 0.48, 95% CI [0.39, 0.57], p < .001). Importantly, this effect magnitude exceeds previously reported benchmarks for AI alone (≈ g = 0.33; Zawacki-Richter et al., 2019) and for MI alone (≈ *g* = 0.30–0.45; Baş, 2016; Ferrero et al., 2021), suggesting genuine synergy rather than mere additivity. The finding that combining AI and MI within the PBL framework yields academic gains above and beyond either component alone represents a meaningful contribution to the educational technology literature.

b. Interpretation of Moderators

AI role matters. The scaffolding role of AI produced the largest effect (g = 0.62), significantly outperforming content generation (*g* = 0.41) and assessment-only (g = 0.31). This pattern supports the theoretical argument that effective AI integration preserves and enhances learner cognitive engagement rather than supplanting it. Scaffolding provides formative, just-in-time guidance that supports students through complex PBL tasks while maintaining intellectual agency. In contrast, assessment-only AI offers feedback only after task completion, missing the critical window for corrective guidance during learning.

MI method matters. Student choice of MI-aligned project roles produced a substantially larger effect (*g* = 0.58) compared to teacher-assigned (g = 0.35) or fixed roles (g = 0.39). This result can be interpreted through Self-Determination Theory (SDT; Ryan & Deci, 2017), which posits that autonomy is one of three basic psychological needs essential for intrinsic motivation and deep engagement. When students select project roles that align with their perceived intellectual strengths, they experience greater autonomy and competence satisfaction, translating into sustained effort and improved achievement a pattern consistent with meta-analytic evidence on SDT in education (Howard et al., 2021).

Education level

The effect was largest at the secondary level (g = 0.52), followed by primary (g = 0.44) and tertiary (g = 0.39). This developmental gradient likely reflects a confluence of factors. Primary students may lack the metacognitive readiness to articulate their intelligence strengths or regulate their use of AI tools. At the tertiary level, the attenuated effect may arise from older students having already developed entrenched learning habits, making them less receptive to structured MI frameworks and AI-guided differentiation. Secondary students, by contrast, possess sufficient metacognitive awareness but remain flexible enough to embrace novel pedagogical approaches.

Subject domain

The absence of significant moderation by subject domain (Q-between = 4.2, p = .12) suggests that the AI-MI-PBL synergy generalises across disciplines. Effect sizes ranged from 0.43 (social sciences) to 0.53 (arts), with overlapping confidence intervals. This finding implies that educators across STEM, language arts, social sciences, and arts can confidently implement AI-MI differentiated PBL without expecting substantially different outcomes.

c. Comparison with Prior Meta-Analyses

The overall effect of g = 0.48 exceeds the effect sizes reported for AI-only interventions in previous syntheses. Zawacki-Richter et al. (2019), in their systematic review of AI applications in higher education, documented that AI interventions often lacked critical reflection on

pedagogical theory and produced modest effects, with reported effect sizes around g = 0.33. Similarly, the present effect surpasses previous estimates for MI-only interventions. Baş (2016) reported a very large effect size (Cohen's d = 1.077) for MI-based education, though this meta-analysis included 75 postgraduate theses from Turkey and may have overestimated effects due to methodological heterogeneity and publication bias. More conservative estimates (Ferrero et al., 2021) place MI-alone effects near g = 0.30–0.45. Chen and Yang's (2019) meta-analysis of PBL alone reported a medium-to-large effect (Cohen's d = 0.71, equivalent to g ≈ 0.65–0.70), which is higher than the synergy effect found here. This apparent discrepancy may reflect that Chen and Yang (2019) examined PBL versus traditional instruction without the added complexity of AI and MI implementation. Nevertheless, the synergy effect of 0.48 is substantially larger than AI or MI alone, confirming the value of their combination.

4.7 Theoretical Implications

The present findings validate the proposed AI-MI-PBL Synergy Model (Figure 1). The model posits that AI can diagnose learner MI profiles, suggest intelligence-aligned project roles, and provide adaptive scaffolding throughout the PBL process, culminating in differentiated artifacts and enhanced academic achievement. The significant moderation by AI role and MI implementation method provides empirical support for each component of the model.

The results also extend Cognitive Load Theory (Sweller, 2010). AI reduces extraneous cognitive load by automating routine tasks such as information retrieval and formatting, thereby freeing working memory resources for deeper learning. Simultaneously, MI-aligned tasks optimise germane load by enabling learners to engage with content through their preferred intellectual modalities. The dual mechanism AI is reducing the wasteful and MI enabling the generative—offers a parsimonious explanation for the observed synergy.

4.8 Practical Implications

For teachers

Educators can leverage AI to generate multiple project role options aligned with students' MI profiles, and then allow learners to select roles that match their perceived strengths. This student-choice approach produced the largest effect size and is feasible with readily available generative AI tools.

For instructional designers

The scaffolding role of AI proved significantly more effective than assessment-only or content generation. Therefore, digital learning environments should embed AI as a formative, interactive partner that provides just-in-time guidance and adaptive feedback, rather than as a summative evaluation system. Structured prompts for chatbots like ChatGPT and adaptive learning platforms represent promising scaffolding tools.

For policymakers

Investment should prioritise AI tools specifically designed with MI-differentiation capabilities for PBL curricula. Professional development programmes must train teachers in both AI literacy and MI-based differentiation, as the synergy effect depends on competent implementation.

4.8 Limitations

Several limitations warrant consideration. First, heterogeneity remained moderate (I^2 = 68%) even after including the examined moderators, suggesting that additional unmeasured factors—such as prior student achievement, teacher experience with AI, and classroom

technology infrastructure may influence effect sizes. Second, MI measurement varied considerably across studies, ranging from formal inventories to teacher judgment, introducing potential construct validity concerns. Third, most included studies were short-term (median 8 weeks, ≤12 weeks); the long-term sustainability of AI-MI-PBL effects remains unknown. Fourth, English-language bias and the predominance of North American and European studies limit the generalisability of findings to non-English, non-Western educational contexts. Finally, the exclusive reliance on quantitative academic achievement as the outcome measure does not capture other potentially important benefits of AI-MI-PBL, such as creativity, collaboration skills, or affective outcomes. Future research should address these limitations through longitudinal designs, standardised MI assessment protocols, and culturally diverse samples.

V. Conclusion

Core Answer

This meta analysis confirms that artificial intelligence and Multiple Intelligences theory operate synergistically within project based learning to enhance academic achievement. The combined effect (Hedges' g = 0.48, 95% CI [0.39, 0.57]) is moderate in magnitude, statistically robust, and meaningfully larger than the isolated effects of AI alone (≈0.33) or MI alone (≈0.30–0.45). Thus, educators and researchers can confidently assert that integrating AI driven adaptive support with MI based differentiation within PBL yields genuine academic gains that exceed the sum of the individual components.

Key Boundary Conditions

The observed synergy is not unconditional; its magnitude depends critically on how AI and MI are implemented. First, AI must scaffold rather than automate. The largest effect (g = 0.62) emerged when AI served as a formative, just in time guide that preserved student cognitive effort. In contrast, using AI solely for summative assessment (g = 0.31) or content generation (g = 0.41) significantly reduced effectiveness. Second, students should choose their MI aligned roles. The student choice condition (g = 0.58) far outperformed teacher assigned (g = 0.35) or fixed roles (g = 0.39). Autonomy, a core psychological need, appears to unlock the motivational benefits of MI differentiation. Third, the synergy is most potent at the secondary level (g = 0.52), where metacognitive readiness intersects with pedagogical flexibility. These boundary conditions provide actionable guidance for implementing the AI MI PBL model effectively.

Future Research Directions

Despite the robust findings, several important questions remain unanswered. Longitudinal designs (≥1 year) are urgently needed to determine whether the observed effects persist over time or diminish as novelty fades. Currently, most studies are short term (≤12 weeks), leaving long term sustainability unknown. Second, adaptive AI systems that dynamically adjust MI support based on real time performance represent a promising frontier. Rather than relying on static MI profiles, future systems could use machine learning to detect shifts in a learner's effective intelligence modalities and re assign project roles accordingly. Third, cross cultural and non Western studies are essential. The current evidence base is dominated by North America and Europe. Research in Asia, Africa, and South America would test the generalisability of the synergy model across different educational systems, cultural values, and technological infrastructures. Additionally, future meta analyses should include non English studies and examine affective outcomes (e.g., motivation, self efficacy) alongside academic achievement.

Final One Sentence Takeaway

Integrating AI as a scaffolding tool within student choice driven, multiple intelligences based project based learning produces moderate to strong improvements in academic achievement across K 16 settings.

References

Ali, S., Diwan, P., & Srinivasan, R. (2025). The impact of GenAI on learning outcomes: A systematic review and meta analysis of experimental studies. Educational Research Review, 44, 100789. https://doi.org/10.1016/j.edurev.2025.100789

Baş, G. (2016). The effect of multiple intelligences theory based education on academic achievement: A meta analytic review. Educational Sciences: Theory and Practice, 16(6), 1833–1864. https://doi.org/10.12738/estp.2016.6.0156

Botella Nicolás, A. M., & Ramos Ramos, P. (2019). Self-determination theory: A motivational framework for project-based learning. Contextos Educativos: Revista de Educación, (23), 27–40. https://doi.org/10.18172/con.3576

Booth, A., Clarke, M., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2012). The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. Systematic Reviews, 1, 2. https://doi.org/10.1186/2046-4053-1-2

Chen, C. H., & Yang, Y. C. (2019). Revisiting the effects of project based learning on students' academic achievement: A meta analysis investigating moderators. Educational Research Review, 26, 71–81. https://doi.org/10.1016/j.edurev.2018.11.001

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta analysis. Biometrics, 56(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

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

Ferrero, M., Vadillo, M. A., & León, S. P. (2021). A valid evaluation of the theory of multiple intelligences is not yet possible: Problems of methodological quality for intervention studies. Intelligence, 88, 101566. https://doi.org/10.1016/j.intell.2021.101566

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). Journal of the Medical Library Association, 105(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. Basic Books.

Gardner, H. (2011). Frames of mind: The theory of multiple intelligences (10th anniversary ed.). Basic Books. (Original work published 1983)

Guyatt, G., Busse, J. W., Schünemann, H. J., & Jaeschke, R. (2025). Core GRADE 4: Rating certainty of evidence risk of bias, publication bias, and reasons for rating up certainty. BMJ, 389, e083864. https://doi.org/10.1136/bmj 2024-083864

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta analyses relating to achievement. Routledge.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. C. (2019). Assessing risk of bias in a randomized trial. 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 (2nd ed., pp. 205–228). Wiley. https://doi.org/10.1002/9781119536604.ch8

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

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta analysis from self determination theory. Perspectives on Psychological Science, 16(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project based learning: A review of the literature. Improving Schools, 19(3), 267–277. https://doi.org/10.1177/1365480216659733

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random effects meta analyses. Research Synthesis Methods, 10(1), 83–98. https://doi.org/10.1002/jrsm.1316

Lee, S., & Kim, J. (2025). A systematic review of generative AI in K–12: Mapping goals, activities, roles, and outcomes via the 3P model. Computers & Education: Artificial Intelligence, 8, 100245. https://doi.org/10.1016/j.caeai.2025.100245

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta analysis. Sage Publications.

Liu, W., & Wang, Y. (2026). The effect of AI and multiple intelligences in project based learning on English achievement. Computers and Education: Artificial Intelligence, 10, 100312. https://doi.org/10.1016/j.caeai.2026.100312

Meng, Z., Chen, Y., & Li, R. (2024). Sensitivity analysis with iterative outlier detection for systematic reviews and meta analyses. Statistics in Medicine, *43*(8), 1549–1563. https://doi.org/10.1002/sim.10008

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Partlett, C., White, I. R., & Riley, R. D. (2017). Random effects meta analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. Statistics in Medicine, 36(2), 301–317. https://doi.org/10.1002/sim.7140

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta analyses. BMJ, 342, d549. https://doi.org/10.1136/bmj.d549

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ryan, R. M., & Deci, E. L. (2017). Self determination theory: Basic psychological needs in motivation, development, and wellness. Guilford Press. https://doi.org/10.1521/978.14625/28806

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta analyses of randomised controlled trials. BMJ, 343, d4002. https://doi.org/10.1136/bmj.d4002

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., & Higgins, J. P. T. (2016). ROBINS I: A tool for assessing risk of bias in non randomised studies of interventions. BMJ, 355, i4919. https://doi.org/10.1136/bmj.i4919

Steinert, S., Hellmers, N., & Tschirner, N. (2025). Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement. Educational Research Review, 46, 100670. Advance online publication. https://doi.org/10.1016/j.edurev.2025.100670

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review, 22(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta analysis and research synthesis. Information Development, 41(3), 825–842. https://doi.org/10.1177/02666669241304407

Veritas Health Innovation. (2024). Covidence systematic review software. https://www.covidence.org

Viechtbauer, W. (2010). Conducting meta analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Warner, A., & Myers, B. (2009). Unpacking effective project based learning. Corwin Press.

Xu, W., & Ouyang, F. (2022). Exploring the effectiveness and moderators of artificial intelligence in the classroom: A meta analysis. In Proceedings of the 2022 International Conference on Artificial Intelligence in Education (pp. 78–92). Springer. https://doi.org/10.1007/978-981-19-5967-7_7

Wilson, D. B. (2025). Effect size conversion and computation for meta analysis. https://www.campbellcollaboration.org/

Zawacki Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0