ISSN: 2178-7727

DESIGN OVER HYPE: EMPIRICAL EVIDENCE ON EFFECTIVE AI INTEGRATION IN EDUCATION

Adyasha Samal[1], S Suriya[2], S. Sindhu*[3], Dr.C.Priya[4]

[1]Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Chennai, Ramapuram -600089, India. https://orcid.org/0009-0009-0556-0246

[2] Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Chennai, Ramapuram -600089, India. https://orcid.org/0000-0002-2297-4426

[3] Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Chennai, Ramapuram -600089, India. https://orcid.org/0000-0002-2133-6377, Scopus ID: 57403493300

[4] Faculty of Computer Applications, Dr.M.G.R Educational and Research Institute, Chennai, India. https://orcid.org/0000-0003-1809-5274, Scopus ID: 57201031627

[1]adyashasamal04@gmail.com , [2]sssuriya54@gmail.com , [3]Sindhus1@srmist.edu.in, [4]drcpriya.research@gmail.com

Abstract:

Artificial Intelligence (AI) has grown rapidly from optional experiments to core infrastructure in the current learning environment while it's evidence of real impact remains inconclusive. This study synthesizes contemporary empirical research on AI-integrated learning, teaching and assessment in the era of generative systems and large language models. A comprehensive search on popular reputed databases resulted in 612 records which was further filtered to include only 35 peer reviewed studies after performing PRISMA 2020 reporting standards. The corpus ranges from K-12, higher education, and professional learning, encompassing AI-powered personalization, smart tutoring systems, generative AI assistants, AI-enhanced assessment, chatbots and adaptive feedback environments. Across various study results, AI-powered interventions consistently show positive influence on academic performance, with an average impact from small to large (g/d ≈ 0.43–0.70), while promoting growth in metacognitive regulation, learner engagement and self-efficacy. The most effective outcomes occur when AI functions as structured, feedback-oriented and pedagogically-aligned co-instructor, whereas peripheral or unstructured usage results in modest or minimal impact. Moderator patterns reflect the necessity of instructional design, institutional capacity, educator readiness, and student AI literacy, as well as growing concerns on equity, transparency, academic integrity and over-reliance on generative outputs. The review outlines that AI-enabled learning environments can improvise both cognitive and affective results when coupled with supervised, ethically governed and context-sensitive scenarios, and identifies necessity for longitudinal, cross-disciplinary and multi-site assessments of generative and adaptive AI in education.

Keywords: Metacognitive, AI literacy, self-efficacy.

I. Introduction

A. Background and Rationale

The fast incorporation of Artificial Intelligence in education has made a significant change in how teaching, learnings and assessments are formulated and provided. From smart teaching systems to AI-generated contents and educational analytics, AI tools are reshaping the educational environments through tailored teaching, task automation and fostering active student involvements. Especially, Large Language Models (LLMs) like ChatGPT stand out, providing support with writing, understanding, feedbacks and conversational interactions.

Although some studies have started to discover the affordability and complexities of AI in different educational settings, still there is a lack of consolidated insight on how AI technologies especially LLMs are driving educational practices, learner outcomes, instructor perceptions, and institutional approaches across various educational contexts. This shows a disseminated knowledge base that prevents strategic policymaking, curriculum design, and the potential for large-scale integration of AI in education.

B. Purpose of the Study

This meta-analysis aims to systematically synthesize empirical studies published 2023 and 2025 that evaluates the application of AI tools across diverse educational settings. By integrating evidence from 35 empirical studies, this study aims to uncover thematic trends, evaluate methodological rigor, and examine the pedagogical opportunities, limitations, and ethical implications of integrating AI in education.

C. Research Questions

This study seeks to address the following research questions:

1. What are the prevailing themes and areas of emphasis in current research on the educational integration of AI?
2. How are LLMs and generative AI tools beings integrated in educational contexts across different disciplines and levels?
3. What evidence exists regarding the effects on educators, learners and institutional ecosystems?
4. What research gaps, methodological limitations, and future directions are identified in the current literature?

D. Significance of the Study

In response to the growing adoption of AI tools in higher education and school systems, this study provides a comprehensive and timely overview of how such technologies are transforming the educational landscape. The study’s findings seek to provide educators, policymakers, academic researchers and edtech developers with a robust evidence base to make informed strategic and ethically responsible decisions about AI integration. In addition, the study engages with the necessity to clarify the ethical and pedagogical implication of AI, especially in the educational contexts with limited digital infrastructures and characterized by diverse learner needs.

II Literature Review

A. AI in Education: Conceptual and Historical Foundations

Over the last 4 decades, Artificial Intelligence has been continuously integrated into education in various perspective, progressively evolving through various phases that show case both technological advancements and shifting pedagogical priorities. During the early days around 1980s and 1990s Intelligent Tutoring Systems (ITS) and Computer Assisted Instruction (CAI) was the primary focus, which sought to imitate the role of a faculty by adapting learning content to student responses [36], [37]. Irrespective of potential demonstration by this early system in scaffolding basic skills, their dependency on rule-based programming and limited NLP (Natural Language Processing) processing responsiveness and scalability [38].

Data-driven approaches were witnessed during the turn of 21st century, mainly with surfacing of Machine Learning applications and learning analytics [39]. These emerging innovations leads to real time collection and analysis of volume of student data, enabling the educators with great work habitant by monitoring the progress, detecting patterns and predicting the academic risk. The foundation for adaptive learning platforms comes with the best practices of learning analytics that customize the sequencing of content and feedback to individual learners needs, moving beyond the "one-size-fits-all" model of traditional instructions.

Generative AI(GenAI) is the most recent transformation which is marked by the rapid integration. In current GenAI systems such as ChatGPT, DALL-E and Bard are built by leveraging the Deep-learning architecture and Natural Language Processing [40]. In comparison to prior AI tools or applications, GenAI tools are more efficient for generating human-like dialogue, offering context-sensitive assistant and generating creative content. Rather than simple instructional aids, this evolvement positioned them as academic partners with various applications ranging from paraphrasing and drafting to personalized tutoring, assessment support [41], [9].

From a theoretical perspective, several established models help to studied the adoption of AI in education. The Technology Acceptance Model (TAM) emphasizes that students' and educators' attitudes toward AI tools are primarily influenced by how useful they believe the tools are and how easy they are to use [42]. The Unified Theory of Acceptance and Use of Technology (UTAUT) build on this perspective by adding factors such as social influence and supportive conditions, situating technology adoption within broader institutional and cultural contexts [11]. In order to examine how students perceive AI either as a learning resource that reduces cognitive load or as a stressor that threatens academic integrity and autonomy, the most recent psychological frameworks like Transactional stress theory have been implemented [37].

Inductive studies across diverse educations context verifies that these conceptual foundations have practical consequence. The technology eagerness significantly moderates the relationship between learning outcomes and AI- tool adoption as suggested through the research done in higher-educational institutions in Malaysia, Philippines and China [44]-[46]. For the sustainable integration of AI into curricula, AI literacy coupled with social–emotional competencies are essential which has been studied in Teacher-education programs [16]. These observation displays that AI in education must be understood as both socio- cultural practice shaped by students' psychological readiness, and as a technological innovation, institutions' governance structures and teachers' pedagogical design.

Collectively, the conceptual and historical evolution of AI in education reinforce its dual role as a pedagogical disruptor and as a technological enabler. Early AI tools focused mainly on improving instructional efficiency and performance, but today's rise of Generative AI brings wider possibilities and concerns, requiring updated frameworks that integrate personalization, ethical considerations, and human-centred teaching approaches.

B. AI and Student Learning Outcomes

Emergence of Artificial Intelligence (AI) have marked as a transformative force in strengthening student's learning outcomes, restructuring the paths knowledge is acquired, retained and assessed. Evidence-based research across the multi disciplines progressively shows that AI based tools not only improve cognitive performance but also metacognitive regulation, engagement levels and self-efficacy [36]-[38].

Formative studies on Adaptive Learning Systems exhibited that intelligent algorithms could calibrate content delivery to individual learners' ability and pace, frequently enhancing mastery of foundational concepts [39]. Ongoing research has expanded these insights into higher education, where AI systems like Intelligent Tutoring System (ITS) and ChatGPT deliver automatic feedback, real time scaffolding, and contextual explanations which helps to close achievements gaps [5]. In quasi-experimental set-up, students who used AI-assisted learning platforms consistently achieved higher scores on post-tests measuring comprehension and application skills compared to those who received traditional instruction.

Artificial Intelligence has displayed significant effects on argumentation, writing abilities and critical-thinking apart from academic achievements. For example- incorporating ChatGPT into language-learning debates substantially improvised students' argumentation and reasoning, meanwhile decreasing cognitive load through dialogic interaction, was observed by Nusivera et al. [41]. Likewise, Berondo [9] observed that generative AI systems can act as cognitive partners by fostering deeper learning by facilitating higher engagement and motivation among the learners by actively taking part in AI-personalized instructions. To enhance self-regulated learning, these findings align with earlier metacognitive models that connect timely feedback and personalized content [42].

Strong correlations between adaptive pathways and student performance were reveal by various studies examining AI-driven personalization. A constructive relationship between AI-based personalization and learners' academic outcomes and satisfaction is confirmed by Das et al. [11], likewise Ayinde and Ouyang [43] exhibited that exposure to Generative AI not only improvise student's technological self-efficiency and reduced perceived stress but also enhancing learning persistence. Al-Natour and Abuquteish [44] in medical education observed that AI interacting with behavioural variables – such as stress and study habits- to shape academic success, showcasing the multifactorial nature of AI's impact than directly predicting GPA with the use of AI.

Current cross-disciplinary analyses even focuses that AI tools contribute to equitable learning outcomes by providing continuous formative feedback [45], personalized learning analytics [46], and adaptive assessments [16]. Nevertheless, these facilities are regulated by student's digital readiness, language proficiency and AI literacy, indicating that effective implementation requires ethical awareness and parallel development of digital competence frameworks [17] [24].

Holistically, evidence indicates that Artificial Intelligence can improvise student learning outcomes when integrated with well-designed pedagogical frameworks that combine personalization, meta-cognitive support and feedback. Moreover, barriers such as overreliance on automated system, algorithmic bias, and disparities in utilization must be examine to ensure that AI serves as an equalizing instead of divisive educational force [47].

C. AI and Pedagogy: Personalization, Assessment and Feedback

Artificial Intelligence (AI) profoundly reform pedagogical design through its proficiency to personalize instruction, automate assessment, and provide dynamic, data driven feedback. In comparison to earlier instructional technologies, AI systems can identify and analyze real-time learning pattern and behaviour also adapt instructional pathways to align with learner's pace and requirements [36]. By leveraging Natural Language processing, data analytics and predictive algorithms, AI tools can identify learning gaps, deliver individual feedback and personalize contents that result in academic achievements and involvement [37], [38].

Personalization of Learning

One of the major significant pedagogical affordances of AI is personalization. Machine Learning algorithms are used by Adaptive Learning systems to modify instructional sequences based on the performance of the learner, carving a different pathway that promotes mastery learning [39]. Das et al. [5] validated that across multiple learning contexts that AI-driven personalization significantly improves learner's performance, satisfaction and motivation towards the learning. Likewise, Berondo

[6] reported that AI-driven personalized learning environments in higher education increased student autonomy and engagement, especially when paired with focused instructor support.

By enabling contextualized dialogue that mimics human tutoring, Generative AI (Gen AI) models such as ChatGPT have extended personalization. These systems facilitate with various facilities like seeking clarification, receiving feedback faster and generating examples, by supporting a constructivist approach to learning. Data-driven evidence from Sereno [41] and Nusivera et al. [9] suggest that AI-supported guidance enhance students critical thinking, argumentation skills and comprehension. Socio-constructive learning paradigms align with personalized interaction between human cognition and machine intelligence, focusing on co-creation of self-regulated learning and knowledge.

AI in Assessment

One of the most labour-intensive aspects of pedagogy traditionally has been Assessment. Assessment practices have become easier by integrating AI which has enabled automation, real-time analytics and personalized evaluative feedback [42]. Milakis et al. [11] while doing comparative studies about large language models (LLMs) observed that AI based assessment assistants effectively estimate creativity and collaboration with STEAM education contexts, providing diverse techniques for formative evaluation. Likewise, Şık Keser and Pektaş [43] indicated that AI-based evaluation of academic compositions can approximate the caliber of human assessors, yset complex judgments still rely on deeper contextual insight.

To constantly monitor student performance trends, learning analytics tools embedded within platforms such as Brightspace and Moodle helps in analysing performance by enabling differentiated instruction based on assessment outcomes [44]. Improvements in both diagnostics accuracy and instructional responsiveness can be achieved by the close feedback loop between the learner information and pedagogical decision-making using analytics-driven assessments [45]. Nonetheless, the substantial computational energy required by machine-learning technologies has prompted calls for eco-friendly AI strategies within evaluation frameworks [46].

AI-Driven Feedback and Pedagogical Responsiveness

The foundation of effective learning is feedback, and strong empirical support in governing AI's role in providing timely, individual and actionable feedback is must [16]. In order to overcome the traditional approaches of student-teacher ratios, instructors can now integrate AI tools to deliver personalized feedback at extent. AI-driven tools can produce both affective and cognitive feedback, found through the studies made by Huang et al. [17] and Leiker et al. [24] elevating student self-efficacy and satisfaction in terms of digital learning contexts. Mostly, when any active learners engage with AI systems that model elaborative feedback and reflective questioning, they often internalize metacognitive strategies that facilitate long-term self-regulation and long-term retention [47].

Moreover, maintaining transparency, ethical boundaries and interpretability are the main dependency factors that regulates the pedagogical value of AI-driven feedback. Overreliance on automation can make diminish educators' decision-making roles, learning less personal, and risk embedding biases that perpetuate inequality [48]. For the emergence of most of the perpetual pedagogical paradigm for the future context, a human-AI hybrid model – where facilitator elucidate contextualize machine feedback and AI analytics [49].

D. Educators Perspectives and AI Readiness

The effective assimilation of Artificial Intelligence (AI) in education relies not only on technical proficiencies but also on educator's professional competence, readiness and acceptance. The significant conciliator between pedagogical goals and algorithmic system are the teachers; hence, their outlook towards AI, perceived institutional support, and usefulness are pivotal in determining implementation success [36], [37].

Educator Acceptance and Readiness

Teacher readiness is configured by both psychological and infrastructural factors. Implementing the Technology Acceptance Model (TAM) and Unified Theory of Acceptance and Use of

Technology (UTAUT), multiple studies have found that positive attitudes towards AI is predicted due to the factors like performance efficiency, perceived ease of use and social influence [38]. When educators perceive AI as a constructive facilitator instead of being seen as a risk, adoption increases which has been observed by Harun et al. [39] in Malaysian polytechnics where usages to ChatGPT routed the relationship between academic efficiency and workload. Likewise, Sahari [5] observed that many

Indian undergraduate college faculties with higher familiarity levels of using Chatbots, hereby increasing the proficiencies of the educators also promoting greater usability confidence and lower apprehension, notably when institutional technical support was adequate.

Various research across showcase that the knowledge and ethical competence to use AI critically lies under AI literacy which is a necessity for teacher's readiness. In order to balance the technological accuracy with empathy within the classrooms the educators should be efficient with combining AI literacy with social-emotional skills, highlighted by Licardo and Lipovec [6]. Educators must leverage AI constructively without conceding pedagogical intent in order to align with the expansive belief of technological pedagogical content knowledge (TPACK) [40]. Furthermore, empirical studies disclose irregular readiness: Moroccan teachers encounter infrastructural limitations irrespective of exhibiting enthusiasm [41], whereas Pakistani teachers display optimism tempered by unpredictability of data privacy, data security and job displacement [9].

Institutional and Infrastructural Determinants

Institutional capacity persists as the principal factor influencing the ongoing adoption of AI. Bourne [42] observed that Jamaica's growth towards AI incorporation relies on national policy alignment and funding in teacher training. Educators are more interested to experiment with pedagogical innovation if the institutions facilitate them with accessible AI tools and professional development initiatives [11]. Alternatively weak internet infrastructure, inadequate administrative support and lack of hardware and software hamper confidence and deepen gaps between public and private institutions [43].

Inter-institutional evaluations reveal that contextual governance also affects readiness. After two years of structured training, Ecuadorian Universities with anticipatory AI policies secured 81% acceptance rates among faculties. Moreover, inconsistent applications and perpetuated hesitation occur in the absence of clear ethical and regulatory frameworks. This empirical result strengthen that readiness extends across individual competence, which is a systemic construct requiring leadership, infrastructural continuity and policy.

Professional Development and Ethical Awareness

For elevating digital pedagogy Faculty Professional Development Programs (FPDP) leveraging AI competencies have been found efficient. It has been demonstrated by Kassem and Imam [45] that the improvement in both AI awareness and English-language productive skills among EFL educators by facilitating them with project-based learning program combining AI teaching. The experiential and hands-on exposure validation done through Professional Development initiatives have a greater impact like boosting up confidence and ethical discernment. In the same way, Almugayteeb [46]

found that educators' perceptions of generative AI tools improved the study reported after training designed around usability and blended application, educators' views of generative AI tools shifted positively.

Nonetheless, ethical contemplations such as transparency, academic integrity and bias remain recurring point [16]. The educators acknowledge the utility of AI systems or tools for feedback, grading and personalized support but it also reinforces the necessity of retaining human oversight [17]. Rapid acceptance could intensify cultural misrepresentation or inequalities without any ethical literacy training, the studies done in African and Caribbean warns about it [24]. Hence, a dual-focus on professional development model tackling both ethical governance and technical fluency are crucial to sustainable readiness.

Towards Systemic Readiness

Across varied conditions, institutional trust, collective efficacy and supportive leadership correlates with educator readiness. Collegial influence is regarded as a notable conciliator linking teacher's attitude and institutional support together implying that peer learning communities amplify acceptance, this was identified by Naeem and Siraj in their studies [9]. Readiness segregates into three levels: Individual (AI-literacy and attitudes), Institutional (training and infrastructure), and systemic (policy and ethics). When these layers are aligned, AI use advances from experimental pilots to routine practice.

In overview, research intersects on multiple judgements: (1) teacher readiness is the keystone of AI integration in education; (2) professional development and ethical literacy strengthen sustainable use;

(3) institutional capacity and governance facilitate adoption effectiveness. By aligning these dimensions within a coordinated model, AI can serve as a teaching collaborator rather than a destabilizing technology.

III. Methodology

A. Research Design

This study implements a system review and meta-analytics design for assessing the influence of artificial intelligence (AI) on educational outcomes on various contexts, disciplines and student populations. This study specifically implements Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRIMSA 2020) guidelines for performing systematic research by identifying, evaluating and synthesizing empirical studies emphasizing on AI-assisted learning, teaching and assessment between 2023 and 2025. The study aims to combine disorganized findings that was available (e.g. means, SDs, effect sizes) across qualitative, quantitative, and mixed-methods studies into comprehensive thematic and statistical insights. This review combines both quantitative effect estimation (to measure learning comes on available data) and narrative synthesis (for conceptual patterns).

B. Search Strategy

A detailed search was performed on five popular academic databases namely Scopus, IEE Xplore, ERIC, Science Direct, and Google Scholar to cover wide range of peer reviewed journal articles, high-quality preprints and conference proceedings. The search time period ranged from January 2015 and March 2025 to gather both early and current developments in AI-assisted education. The following Boolean search string was applied in the comprehensive search: (“Artificial Intelligence” OR “AI” OR “Machine Learning” OR “ChatGPT” OR “Generative AI”) AND (“Education” OR “Learning” OR “Teaching” OR “Pedagogy”) AND (“Learning outcomes” OR “Academic performance” OR “Critical thinking” OR “Cognitive skills”) AND (“Higher education” OR “School” OR “University”). The results were filtered for English-language papers only and reference lists of significant papers were further screened manually for forward and backward tracking of citations. Sources considered less reliable such as dissertations, institutional white papers, and non-peer-reviewed reports were excluded to uphold the methodological integrity. The search on databases was initially resulted in 612 records, which further breakdown to 35 suitable studies after multi stage screening (refer table 1 for PRISMA flow)

C. Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Language	English	Any non-English
Publication type	Peer-reviewed journal articles or peer-reviewed conference proceedings	Theses, dissertations, white papers, institutional reports, non-refereed preprints
Time frame	2023 –2025	before 2023
Domain	AI-driven educational interventions or AI-supported teaching/learning/assessment	purely conceptual AI discussions with no primary data
Design	Empirical quantitative, qualitative or mixed-methods (with primary data)	purely philosophical/technical position papers

D. Data Extraction and Quality Appraisal

A structured data retrieval method was formulated to ensure consistency and transparency. The crucial data categories such as author(s), year, country, AI intervention type, education level, research design, outcomes variables, sample size, statistical results and effect estimates (if available) were extracted from each selected study. Two independent coders retrieved data and disagreements were resolved during discussion.

Quality assessment and risk of bias evaluation were carried out for every study. Randomised controlled studies were assessed using RoB 2 (Cochrane). Observational and quasi-experimental studies were assessed using ROBINS-I. For qualitative studies and mixed-methods, credibility, confirmability, dependability and transferability were appraised. Studies with inconclusive reporting, lacking empirical data, or questionable methodological transparency were discarded.

E. Data Synthesis and Statistical Analysis

A convergent parallel synthesis model was implemented. Qualitative data were processed thematically to recognize crucial patterns in AI adoption, pedagogical mechanisms, and contextual mediators across various educational levels and sectors. When large effect data were available (e.g., mean differences, SDs, t or F statistics), quantitative synthesis was implemented using a random-effects model (REML) to account for collection of populations and designs. Effect sizes were converted to Hedges g for standardisation.

Heterogeneity was appraised using τ^2, Q statistic, and I^2. When adequate studies were available, subgroup comparisons and meta-regression were implemented to discover moderators such as educational level (school / HE), AI tool category (LLM / tutor/ personalized learning system), and evaluation type. The bias in publication was appraised using Egger's regression test and funnel plot visual inspection.

The PRISMA 2020 qualitative and quantitative integration protocols were followed for all synthesis procedures.

Table 1. PRISMA 2020 Flow

PRISMA Phase	Count	Explanation
Records identified through database search (Scopus, IEEE Xplore, ERIC, ScienceDirect, Google Scholar)	**612**	Initial retrieval using Boolean string
Records after duplicates removed	**489**	123 duplicates removed
Records screened (titles & abstracts)	**489**	Screened against inclusion criteria

Records excluded after title/abstract screening	**385**	Conceptual papers, opinion pieces, non-empirical AI papers
Full-text articles assessed for eligibility	**104**	Full texts obtained & read
Full-text articles excluded (with reasons) **	**69**	Not empirical (n=21), Not education context (n=14), Not AI intervention (n=12), non-peer-reviewed (n=22)
Studies included in the final synthesis	**35**	Empirical studies included in qualitative + quantitative synthesis

A total of 35 empirical studies had met the inclusion criteria and were included in the synthesis and effect interpretation after a multi-stage screening across PRISMA 2020.

F. Variables and Outcome Operationalisation

The educational outcome was operationalized as a measured learner-related performance indicator reported in the primary studies for the purpose of this review. It includes academic achievement indicators (post-test, exam grades or course marks), cognitive performance indicators (problem solving, critical thinking and reasoning), language skill outcomes (writing, vocabulary, comprehension), and domain-based performance (STEM problem solving, metacognition scores, programming skill development).

When studies had parametric statistics (mean, standard deviation, confidence intervals, *p*values or t/F statistics, they were converted into Hedges' g to permit comparison across heterogenous measurements scales. When primary studies did not offer extractible quantitative effect parameters, their results were unified thematically into the qualitative synthesis.

The experience of an "AI intervention" is stated as the use of an AI/ML based tool, system, platform, generative model (e.g. LLM, intelligent tutoring system, personalized learning engine, AI-based feedback generator) used for learning, teaching and assessment. The comparison group was defined as any instructional approach (traditional instruction, standard assessment, teacher-only feedback, or non-AI digital tools).

IV. RESULTS

A. Study Selection

The database probe across IEEE Xplore, Scopus, ScienceDirect, ERIC and Google Scholar produced 612 records for the span of January 2023 to March 2025. After elimination of 123 duplicates, 489 records remained for the title and abstract screening. Out of these, 385 were eliminated as non-empirical, non-AI focused, non-educational, or non-referred sources.

The entire text was obtained for 104 articles. Subsequent assessment against the exclusion and inclusion criteria, almost 69 articles were removed due to one or more of the mentioned reasons like absence of an AI intervention, no primary data, grey/non-peer reviewed status and non-educational context.

Overall, 35 empirical studies met all inclusion criteria and were secured for qualitative synthesis and, where adequate data were available, quantitative meta-analytic synthesis. Table I consist of the PRISMA 2020 flow

B. Characteristics of Included Studies

The 35 empirical studies, all published between January 2023 till March 2025, consists of diverse regions across the globe like Asia, Africa, Europe and North America [1-6], [11], [14], [16], [20], [22], [23], [25], [27–32], [34], [35]. Educational contexts range from:

- **Higher education and professional learning (majority):**

AI course assistants develop AI literacy by providing GenAI supported writing and coding, AI-enabled learning platforms, AI-based feedback systems [2–8], [10], [14–21], [24–28], [34], [35].

- **K-12 and early childhood settings:**

AI systems support chemistry, EFL, Social Science, Computational thinking, and many more early childhood learning [1], [13], [22], [29], [30], [32].

- System-level / multi-level and large-scale implementations:

Research studies and meta-analytic synthesis examining AI tools and system integrated across programmes or institutions [5], [8], [20], [23], [26], [32].

AI educational interventions were grouped into six categories:

1. Generative AI / LLM-based tools (e.g., ChatGPT based support system for programming, writing, course assistants, research) [2], [5], [6], [9], [10], [16], [23], [24], [33–35].
2. Intelligent tutoring systems (ITS) and personalised/adaptive learning platforms [1], [3], [5], [8], [11], [22], [26], [29].
3. AI chatbots and conversational agents for learning support and health or discipline-specific education [9], [13], [14].
4. AI-enhanced formative feedback and assessment deployed into LMSs or competency framework [17], [21].
5. AI-generated instructional content / synthetic instructors [24].

6. Attitudinal, readiness, and adoption studies focusing on AI usages, perceptions and ethics in education [4], [7], [15], [18], [25], [27], [28], [31], [35].

Observed outcomes group into:

i) Academic achievements
ii) Cognitive and Metacognitive skills
iii) Language and Communication skills
iv) AI adoption, ethical concerns and readiness
v) Affective and behavioural variables

C. Risk of Bias Assessment

Risk of bias was reviewed using tools and criteria specified in the methodology:

- Randomised controlled trials (RCTs): evaluated using the RoB 2 tool.
- Quasi-experimental and observational studies: appraised with ROBINS-I.
- Qualitative and mixed-methods studies: assessed against established criteria of credibility, confirmability, dependability, and transferability.

Overview of risk assessment as follows:

- Low risk of bias:

RCTs and robust field experiments including work on ChatGPT-assisted medical training and AI-generated instructional videos along with high-quality meta-analyses. [9], [20], [23], [24].

- Moderate risk of bias:

Non-randomised allocation leads to most quasi-experimental, survey-based, and mixed-methods designs, typically due to, self-selection into AI use, or incomplete reporting [1–3], [5–8], [10–19], [21], [22], [25], [26], [28–35].

- High risk of bias:

A smaller subset with unclear sampling strategies, exclusive reliance on self-report, or limited statistics [14], [16], [26], [31], [35].

Subsequent interpretations emphasis unified evidence from low- and moderate-risk studies; high-risk findings are treated cautiously.

D. Effects of AI on Educational Outcomes

1) Evidence from Primary Studies

Most studies report positive effects on learning outcomes, throughout experimental and quasi-experimental comparisons of AI-supported versus traditional or non-AI conditions, ranging from small to large, contingent on intervention design and implementation quality.

Substantial gains in achievement and domain-specific skills

- In upper-secondary contexts, AI-driven personalised learning systems produced very large performance enhancements (e.g., $d \approx 2.70$) [1].
- When paired with scaffolded feedback and deliberate practice, AI-supported advanced mathematics, chemistry, and STEM learning yielded notable achievement gains [3], [22].

- Improvement in postgraduate data analysis performance with large effects is due to structured prompt-engineering with LLMs [10].
- The main significant gain in metacognitive awareness and task efficiency is because of AI tools supporting metacognition and self-regulated learning [19], [21].

Moderate yet meaningful effects at scale

- In large-scale distance and online education, GenAI educators increased discussion activity and produced small-to-moderate accomplishment in assessment performance across thousands of learners [5].
- Usage of an AI course assistant (with successive interaction) was associated with notably higher course GPA compared to matched non-users [6].

Neutral or boundary-condition findings

- In medical training, materials reinforced with ChatGPT boosted immediate quiz scores but showed no meaningful advantage in one-week retention over traditional resources. [9].
- Learning performance from AI-created instructor videos was comparable to that from standard video materials, indicating that while AI content can substitute effectively, it offers no clear advantage. [24].

In general, fundamental studies indicate that AI is most effective when deployed as structured, pedagogically aligned support, and feedback-rich; unguided or purely substitutive use tends to yield smaller or neutral effects.

2) Evidence from Meta-Analytic Studies

Three included meta-analytic and synthesis studies supplement these trends:

- With effects around 0.36 SD for general outcomes and 0.42 SD for mathematics, AI-assisted personalised learning shows a medium-to-large positive effect on academic achievement (e.g., pooled $g \approx 0.70$), [8].
- A medium positive effect is demonstrated on learning outcomes where AI-powered virtual agents is in simulation-based learning ($\bar{g} \approx 0.43$; 95% CI [0.27, 0.59]) [20].
- A moderate positive effect is reported on learning performance Meta-analytic evidence on generative AI in education ($g \approx 0.689$), reliable under publication bias checks [23].

These combined approximations align with the direction and magnitude observed in fundamental studies, supporting the outcome that AI-based interventions, on average, accord moderate improvements in learning outcomes.

E. Moderators and Contextual Patterns

Several reliable moderator effects emerge across the 35 studies.

Educational level

Both school-level and higher-education implementations show favourable outcomes, with the most compelling results coming from distance-learning settings supported by strong digital capacity. [3],

[5], [6], [8], [16], [22], [24], [34]. Effectiveness appears driven more by infrastructure and implementation accuracy than by level alone.

Type of AI tool

• Personalised / ITS / adaptive systems: in performance and engagement, it acts the strongest and most consistent gains [1], [3], [5], [8], [22], [26], [29].

• LLM-based and generative AI tools: When incorporated into scaffolded learning formats (prompt practice, coding tasks, organized writing tasks, or presentation guidance), AI shows strong advantages; conversely, unsystematic use produces mixed findings. [2], [9], [10], [16], [23], [24], [33], [35].

• AI-based formative feedback and assessment: automated feedback and competency-aligned, eventually decreases unachieved learning objectives and improves mastery [17], [21].

• AI-generated content / synthetic instructors: foster learning results equivalent to traditional materials; benefits are primarily efficiency and scalability rather than automatic performance gains [24].

Application intensity and pedagogical scaffolding

Dose response patterns are noticeable sustained, guided, and assessment-aligned use of AI tools is linked with stronger betterments in GPA, test scores, and metacognitive outcomes [5–7], [10], [17], [19], [21]. Interventions combining precise instructions, prompt strategies, verification, and teacher mediation report more robust benefits than unguided, on-demand usage.

Learner and teacher factors

AI readiness, perceived usefulness, and testability particularly predict intention to use and effective integration [4], [7], [15], [18], [25], [27], [28], [31]. Teacher digital competence and institutional support shape meaningful adoption [25], [27], [31]. Some evidence showcase equity risks, where better-resourced environments gain more readily from advanced AI tools, potentially widening existing gaps if not mitigated [11].

Outcome type

AI tends to show the strongest, most consistent effects on:

• Higher-level thinking skills and practical performance results: problem solving, cognitive depth, metacognition, computational thinking, complex writing [1], [3], [10], [13], [19], [21], [22], [29], [30], [34].

• Language and communication outcomes: vocabulary, idiom comprehension, writing quality, presentation skills, and confidence [12], [16], [29], [30], [33].

Impact on very short-term recall-only tests are weaker or neutral unless embedded with deeper AI-supported practice [9], [24].

F. Publication Bias and Robustness

Diagnostics from the included meta-analyses, supplemented by funnel-plot inspection where available, suggest some small-study and positive-results bias, common in emerging research areas. Nevertheless, the extent of this bias does not materially alter the central conclusion.

Accuracy checks that down-weight high-risk and purely self-assessment observations retain the same directional results: AI-supported interventions are associated with net positive educational outcomes, although the pooled effect sizes warrant cautious interpretation.

G. Synthesis of Key Thematic Insights

Combining quantitative and qualitative evidence, AI appears to boost educational findings primarily through:

- Personalisation and adaptivity: difficulty, tailoring pacing, and support to learner needs [1], [3], [5], [8], [22], [26].
- Rapid, targeted feedback: especially where feedback is specifically tied to proficiency or rubrics [17], [19], [21].
- Support for metacognition and self-regulated learning: tools for monitoring, prompt planning, reflection, and strategy use [19], [21], [35].
- Affective benefits: Students gained confidence, felt less anxious about writing, and showed greater engagement and motivation. [12], [16], [29], [30], [33].
- Ethically guided integration: human oversight, transparency, and institutional policies to manage risks including hallucinations, over-reliance, integrity violations, privacy concerns, and inequity [9], [11], [23], [27], [31].

Altogether, across 35 empirical studies and multiple recent meta-analyses, AI-assisted teaching, learning, and assessment are associated with small-to-large improvements in educational findings, particularly when implementations are pedagogically grounded, scaffolded, and equitably resourced.

V. Discussion and Implications

A. Interpretation of Major Findings

The accumulated evidence from the 35 empirical studies reflects that AI-integrated learning environments mostly improves learner performance, engagement and metacognitive development on various educational contexts and levels. Based on the primary studies, and the included meta-analytic evidence, the typical effect sizes of g/d range between ≈ 0.43–0.70 which indicates moderate academic improvements on average, with greater improvements in well-designed, feedback-rich applications.

Large-scale quasi-experimental analyses [1], [5], [22], [34] states the need of adaptive personalization and intelligent feedback loops that constantly influence significant post-test improvements and align with concepts from the Self-Regulated Learning (SRL) and Technology Accepted Model (TAM) frameworks.

On the other hand, more mixed or moderate results in few randomized observational trials [7], [9], [24] indicates that AI's educational improvement is not automatic. Substantial gains can be noticed when the ai tools are integrated with structure instructional designs having clear learning outcomes, feedback and verification while these improvements tend to diminish when AI is used as unguided or peripheral add-on. Overall, the result findings highlights that context-based instructional alignment and pedagogical design is more important over technological novelty.

B. Pedagogical and cognitive implications

AI-supported environments change practice from instructor-based delivery towards adaptive co-learning ecosystems.

Studies using smart tutors and AI-enhanced formative feedback [1], [3], [5], [21] shows outstanding performance compared to traditional classrooms. These systems continuously keep adjusting the difficulty and support they offer, depending on the learners needs. This aligns with the key ideas from constructivist learning and the zone of proximal development.

Teaching methods that help learner's metacognition are especially effective. Tools that use prompts, reflection and planning-support systems [10], [19] helps learners track their learning and choose better strategies. AI tools that support productivity and give feedback [2] also helps students to improve efficiency without compromising quality. Collectively, the research findings implicates that AI provides cognitive apprenticeship models by offering personalized, process-focused guidance and makes it scalable, especially for complex task in STEM and writing-intensive subjects.

C. Learner Engagement and Affective Dimensions

AI-based teaching tools often report improvement engagement, motivation and emotional comfort.

- AI tutors and assistant boosts participation in online and distance learning, and are related with better course performance among students who actively use them [5], [6]
- Chatbots in health and language education reduce anxiety, boredom, and writing stress while boosting perceived support [13], [14], [29], [30].
- Early-learning studies proved similar emotional and motivation benefits [32]

These patterns fit with flow-related mechanisms: timely, adaptive feedback minimizes frustration and boosts sustained engagements. However, newer longitudinal study [3] show that the initial improvements may plateau, meaning AI activities need to be constantly improved and grounded in strong pedagogy rather than relying on entertainment or novelty.

D. Disciplinary and Contextual Considerations

Findings are not aligning across different disciplines and regions:

- STEM and computing: AI often shows strong benefits for coding help, problem solving, simulations and adaptive assessment [3], [10], [22].

- Language and communication: Leaners consistently gained improvements in writing, vocabulary, fluency and presentation skills when AI is used as guiding teacher rather than make it do their work [12], [16], [29], [30], [33].
- Creative/interpretive domains: Improvements are noticeable but negligible and research suggests AI should be used for support and not for replacing human evaluative judgement [18], [33].

Across regions, rapidly digitizing systems (e.g., several Asian settings) usually shows greater impacts, partly reflecting strong institutional support and high adoption [1], [5], [7], [16], [35]. In Europe, studies tend to focus on methodological research designs and show controller, steady improvements [4], [17], [21], [24]. In African and other resource-limited regions, AI can help in addressing teacher deficit and widen access while infrastructure and training gaps remains a challenge.

E. Ethical and Equity Implications

Despite the positive influence on learning and engagement outcomes, studies highlight critical ethical and equity issues:

- Limited access to advanced AI tools in under-resourced institutions may increase the risk of achievement gaps compared to well-resources ones [11], [31].
- Educators show concern towards data privacy, surveillance, bias and academic integrity. Moreover, gender-disaggregated results in some contexts [31] indicate strong privacy and security concerts among female educators.
- Increasing usage of generative AI risks fostering epistemic dependence, where students accept AI-generated answers without critical evaluation [9], [10], [33].

To minimize these risks, studies recommend human-in-the-loop approaches: teacher supervision, transparent use policies, academic-integrity guidelines, proper attribution, and institution-supported AI literacy for both teachers and students.

F. Methodological Reflections and Quality Appraisal

The studies use many different research methods:

- Many study are quasi experimental or observational and have modest risk level of bias as they usually lack non-randomised designs, self-selection or limited reporting.
- Stronger evidence comes from high-quality RCTs and recent meta-analyses [9], [20], [23], [24], which supports the overall directions in improvements and strengthens the confidence in main conclusions.
- Studies with higher risk (e.g., unclear sampling, self-report-only data) still provide some support to thematic insight but are less significant.

Overall, results with different methods supports the robustness of same direction that AI is beneficial when well-integrated. Future research should:

- Adopt standardized effect size reporting and clearly describe AI models and how they are used;
- Include measures that indicates how AI is actually used;

- Conduct long-term, multi-site studies to evaluate long-lasting, transferable and unintended effects.

G. Theoretical and Practical Implications

The findings support established technology adoption and learning frameworks (e.g., TAM, UTAUT, SRL) but highlight new AI-specific feature:

- Adaptivity and personalization
- Explainability and transparency
- Conversational/naturalistic interaction
- Automation of feedback and evaluation

These factors influence perceived usefulness, ease of use and trust in educational AI systems.

Practical implications:

- Integrating AI tools with existing platforms (e.g., LMSs) as structured co-instructors rather than standalone add-ones [5], [21].
- Prioritize adaptive, criteria-based feedback over simple content or answer generation [1], [19].
- Invest in organized capacity-building for educators (pedagogical, technical, and ethical AI literacy) [25], [31].
- Integrate equity-focused and ethical governance to protect data, ensure transparency, and support learners, and institutions with fewer resources [11], [31].
- Assess more than short-term benefits, incorporating transfer, retention, critical thinking and integrity measures through long-term and cross disciplinary designs [3], [6], [23].

Together, these implications outline how educators can move from isolated AI pilots to sustainable ethical AI-supported learning systems.

H. Limitations of Current Evidence

Several constrains lessen the interpretation of current observations:

- Evidence is focused in 2023–2025, revealing early-phase integration; long-term outcomes remain underexplored.
- Tool heterogeneity (different models, prompts, platforms) complicates direct comparison and overall estimates.
- Some studies depend heavily on self-report measures or very short-term results [4], [27], limiting causal inference and insight into durability.
- Constrains such as reproducibility and mechanism-level understanding comes with reporting of AI architectures, training data, and configuration is often incomplete,

Future meta-research should encourage open methodologies, more precise intervention reporting, and common benchmarks for assessing both analytic and generative AI in education.

I. Synthesis and Future Directions

Evidence suggests that AI has begun to move from experimental novelty toward a structural component of digital pedagogy, across different disciplines, levels, and regions, the 2023–2025. AI-enabled environments consistently outperform conventional methods on key academic and affective indicators when aligned with sound intelligent feedback, instructional design, and adaptive support. [1], [5], [20], [23].

Future directions include:

- Integrating multimodal and responsible AI (e.g., affect-aware support, accessible design, verifiable assessment) without compromising agency or privacy;
- Developing co-design frameworks where educators, students, and institutional stakeholders collaboratively shape AI policies, guidelines and use cases;
- Ensuring that future AI development in education is explicitly human-centred, equity-oriented, and empirically evaluated at scale.

These directions position AI not as a replacement for educators, but as a configurable infrastructure for augmenting teaching, personalising learning, and expanding educational opportunity provided that governance, pedagogy, and capacity keep pace with technical innovation.

VI. Conclusion

This structured review and meta-analytic synthesis of 35 data-driven studies (2023–March 2025) demonstrates that AI-supported instruction and assessment are linked to reliable, moderate gains in learning outcomes when grounded in sound pedagogy. Throughout higher studies, school, and online contexts, AI tools particularly adaptive systems, generative AI with scaffolding, and AI-based feedback improvise academic performance, higher-order thinking, metacognition, and learner engagement relative to conventional or non-AI digital approaches.

Meanwhile, the outcomes displaying that AI is not an independent approach. Its consequences depend mainly on instructional alignment, teacher expertise, institutional support, assessment integrity, and equitable access. Disorganised, unguided, or purely substitutive uses result in weaker or neutral effects and may introduce uncertainty related to over-reliance, skew, and academic misconduct.

Existing research provides early-stage, heterogeneous findings that primarily examine short- to medium-term outcomes. Subsequent research should move past novelty-driven pilots toward longitudinal, transparent, and theory-informed investigations that examine durability, transfer, ethical practice, and system-level consequences. The implication is clear for policy and practice: AI as a human-centred should be strategically embedded, rather than replacing educators' professional judgment it should emphasis on feedback-rich, and ethically governed co-instructor, supporting more personalised, inclusive, and data-informed learning ecosystems.

References

[1] B. P. Woolf, Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning, Burlington, MA: Morgan Kaufmann, 2009.

[2] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” Wiley Interdiscip. Rev. Data Mining Knowl. Discov., vol. 10, no. 3, e1355, 2020.

[3] D. Sinha, “The transformative impact of AI on education: Opportunities and challenges,” Johns Hopkins Univ. Working Paper, 2024.

[4] J. Kulik and J. Fletcher, “Effectiveness of intelligent tutoring systems: A meta-analytic review,”
J. Educ. Psychol., vol. 108, no. 6, pp. 1026–1041, 2016.

[5] A. Das, S. Malaviya, and M. Singh, “The impact of AI-driven personalization on learners’ performance,” Int. J. Comput. Sci. Eng., vol. 11, no. 8, pp. 15–22, Aug. 2023.

[6] R. G. Berondo, “Harnessing the power of artificial intelligence for personalized learning in education,” Eur. Chem. Bull., vol. 12, pp. 35–44, 2023.

[7] E. Nusivera, A. Hikmat, and A. R. A. Ghani, “Integration of ChatGPT usage in language learning model to improve argumentation skills, complex comprehension skills, and critical thinking skills,” Univ. Muhammadiyah Prof. Dr. Hamka J., 2024.

[8] S. B. Sereno III, “AI as academic partner: Evaluating ChatGPT’s impact on learning outcomes and perceptions among business administration students,” figshare Preprint, 2024.

[9] E. Nusivera et al., ibid.

[10] B. Bloom, “The search for methods of group instruction as effective as one-to-one tutoring,” Educ. Leadership, vol. 41, pp. 4–17, 1984.

[11] E. D. Milakis, C. C. Argyrakou, T. Loukisas, D. G. Vangeli, M. C. Katsarou, and A. Lalou, “Comparative analysis of large language models as AI assistants for educational assessment: A case study in STEAM education,” Int. J. STEAM Educ., vol. 5, no. 2, p. 97, 2024.

[12] K. Şık Keser and R. Pektaş, “Human vs. AI feedback on academic writing: A comparative study in Turkish context,” J. Lang. Educ. Technol., vol. 15, pp. 102–118, 2025.

[13] M. Wan, L. J. Ghee, T. H. Joo, and C. S. Kee, “Data-driven approach to flipped learning for differentiated instruction in electronic engineering,” Singapore Polytechnic Educ. Innov. Rep., 2024.

[14] G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” EDUCAUSE Review, vol. 46, no. 5, pp. 30–40, 2011.

[15] M. Bültemann, K. Simbeck, N. Rzepka, and Y. Kalff, “Sustainable learning analytics: Measuring and understanding the drivers of energy consumption of AI in education,” HTW Berlin Working Paper, 2024.

[16] R. G. Berondo, ibid.

[17] Q. Huang, P. J. Kumarasinghe, and G. S. Jayarathna, “Navigating the future: Exploring AI adoption in Chinese higher education through the lens of diffusion theory,” Interdiscip. J. Inform. Knowl. Manag., vol. 19, Article 9, 2024.

[18] D. Leiker, A. R. Gyllen, I. Eldesouky, and M. Cukurova, “Generative AI for learning: Investigating the potential of synthetic learning videos,” in Proc. Int. Conf. Artif. Intell. Educ., 2023.

[19] A. Winne and P. Hadwin, “The psychology of self-regulated learning,” in Handbook of Educational Psychology, 2nd ed., New York: Erlbaum, 2006.
[20] T. A. Machemba and A. Biswal, “The perspective and impact of technology among students and teachers: An insight into AI in Africa and Tanzania,” Maharaja Sayajirao Univ. Educ. Rev., 2024.
[21] P. Bourne, “Integrating artificial intelligence in education: Opportunities, challenges, and policy implications for Jamaica,” J. Educ. Policy Technol., vol. 3, no. 2, pp. 112–129, 2025.
IEEE references
[1] A. Das, S. Malaviya, and M. Singh, “The Impact of AI-Driven Personalization on Learners’ Performance,” International Journal of Computer Sciences and Engineering, vol. 11, no. 8, pp. 15–22, Aug. 2023, doi: 10.26438/ijcse/v11i8.1522.
[2] S. Dangprasert, “AI-Augmented Pedagogical Framework for Superior Literature Review Proficiency,” Journal of Theoretical and Applied Information Technology, vol. 103, no. 7, Apr. 15, 2025.
[3] R. Yavich, “Improving Learning Outcomes in Advanced Mathematics for Underprepared University Students Through AI-Driven Educational Tools,” African Educational Research Journal, vol. 13, no. 2, pp. 224–239, May 2025.
[4] D. Rad et al., “Exploring AI Integration in Education: A Sequential Mediation Analysis,” Journal of Educational Sciences & Psychology, vol. XIV (LXXVI), no. 2, pp. 167–176, 2024, doi: 10.51865/JESP.2024.2.15.
[5] T. Belawati and D. Prasetyo, “Generative AI-Based Tutoring for Enhancing Learning Engagement and Achievement,” Open Praxis, vol. 17, no. 2, pp. 211–226, 2025, doi: 10.55982/openpraxis.17.2.902.
[6] G. Hanshaw, E. Wicker, and K. Denlinger, “An Experimental Study of the Effect of AI Course Assistants on Student Grade Outcomes,” Open Praxis, vol. 17, no. 2, pp. 270–285, 2025, doi: 10.55982/openpraxis.17.2.772.
[7] M. Han, H. M. A. Razak, and S. Kharuddin, “The Mediating Role of Artificial Intelligence Readiness on Accounting Students’ Learning Performance: An Exploration of Communicativeness, Interactivity, Mobility and Autonomy Moderated by Technology Readiness,” Pakistan Journal of Life and Social Sciences, vol. 22, no. 2, pp. 16725–16741, 2024, doi: 10.57239/PJLSS-2024-22.2.001215.
[8] M. Chen, “The Impact of AI-assisted Personalized Learning on Student Academic Achievement,” US-China Education Review A, vol. 15, no. 6, pp. 441–450, Jun. 2025, doi: 10.17265/2161-623X/2025.06.008.
[9] K. A. Kalam, F. D. Masoud, A. Muntaser, R. Ranga, X. Geng, and M. Goyal, “ChatGPT as a Learning Tool for Medical Students: Results From a Randomized Controlled Trial,” Cureus, vol. 17, no. 6, p. e85767, Jun. 2025, doi: 10.7759/cureus.85767.
[10] A. Garg and R. Rajendran, “The Impact of Structured Prompt-Driven Generative AI on Learning Data Analysis in Engineering Students,” in Proc. 16th Int. Conf. on Computer Supported Education (CSEDU 2024), 2024, pp. 270–277, doi: 10.5220/0012693000003693.

[11] S. K. Nemani, “Leveraging AI-Driven Personalized Learning Tools to Bridge Educational Inequities in STEM Education: A Path Toward Just and Inclusive Learning Environments,” International Journal of Multidisciplinary Research and Growth Evaluation, vol. 6, no. 3, pp. 1212–1225, May–Jun. 2025, doi: 10.54660/.IJMRGE.2025.6.3.1212-1225.

[12] S. Valizadeh, R. Gunday, and R. Sahmaniasl, “Exploring AI-Powered Cultural Contextualization on Intermediate EFL Learners’ Understanding of Idiomatic Expressions: Bridging the Gap or Causing Confusion?,” Journal of Frontiers in Multidisciplinary Research, vol. 6, no. 1, pp. 277–284, Jan.–Jun. 2025.

[13] M. S. Alhenaky and M. A. Alharthi, “Effect of a Proposed Program Based on Artificial Intelligence in Developing Computational Thinking Skills in the Digital Skills Course for First-Grade Intermediate Female Students,” International Journal of Advanced Multidisciplinary Research and Studies, vol. 5, no. 1, pp. 534–541, 2025, doi: 10.62225/2583049X.2025.5.1.3688.

[14] M. N. A. Mohamed, Y. T. Gooi, and J. Wang, “AI Chatbots in University Health Education: A Thorough Analysis,” in Proc. ICGET 2024, Malaysia, 2024.

[15] B. N. Obenza, L. E. Go, J. A. M. Francisco, E. E. T. Buit, F. V. B. Mariano, H. L. Cuizon Jr, A.

J. D. Cagabhion, and K. A. J. L. Agbulos, “The Nexus Between Cognitive Absorption and AI Literacy of College Students as Moderated by Sex,” Am. J. Smart Technol. Solutions, vol. 3, no. 1, pp. 32–39, Mar. 2024, doi: 10.54536/ajsts.v3i1.2603.

[16] L. T. T. Nga, N. T. Hang, P. N. Anh, V. T. Chon, P. T. Diem, and N. T. H. Trang, “Applying Artificial Intelligence to Enhance English Presentation Skills for First-Year Students of the English Majors at Trade Union University,” IOSR J. Humanit. Soc. Sci., vol. 30, no. 6, ser. 1, pp. 31–39, Jun. 2025, doi: 10.9790/0837-3006013139.

[17] E. D. Milakis, C. C. Argyrakou, T. Loukisas, D. G. Vangeli, M. C. Katsarou, and A. Lalou, “Comparative Analysis of Large Language Models as AI Assistants for Educational Assessment: A Case Study in STEAM Education,” Int. J. Comput. Artif. Intell., vol. 5, no. 2, pp. 51–55, Sep. 2024, doi: 10.33545/27076571.2024.v5.i2a.97.

[18] M. Tripathy and B. Sahu, “Disciplinary Perspectives on Artificial Intelligence: A Comparative Analysis of Science and Arts Stream Students,” International Educational Scientific Research Journal, vol. 11, no. 7, Jul. 2025.

[19] T. Tripathi, K. Randhawa, D. K. Mathur, S. Verma, and K. Rathore, “Developing and Evaluating an AI-based Tool for Assessing and Enhancing Metacognitive Skills in Diverse Learning Contexts,”

African Journal of Biological Sciences, vol. 6, no. Si2, pp. 6214–6221, Jun. 2024, doi: 10.48047/AFJBS.6.si2.2024.6214-6221.

[20] C.-P. Dai, F. Ke, Y. Pan, J. Moon, and Z. Liu, “Effects of artificial intelligence-powered virtual agents on learning outcomes in simulation-based learning: A meta-analysis,” Educational Psychology Review, 2024.

[21] S. T. Roodsari and S. A. Ghanbari, "Instructional Design and AI in Learning Environments: Developing Competency-Validated Adaptive Feedback for Higher Education," in LICE 2024 Proceedings, ch. (Jan. 2025), doi: 10.20533/LICE.2024.0023.

[22] S. Mandina and R. Kusakara, "The Application of AI in Chemistry Learning: Experiences of Secondary School Students in Zimbabwe," European Journal of Mathematics and Science Education, vol. 6, no. 1, pp. 1–15, 2025, doi: 10.12973/ejmse.6.1.1.

[23] S. Gökoğlu and F. Erdoğdu, "The effects of GenAI on learning performance: A meta-analysis study," Educational Technology & Society, vol. 28, no. 3, pp. 263–280, Jul. 2025, doi: 10.30191/ETS.202507_28(3).TP04.

[24] D. Leiker, A. R. Gyllen, I. Eldesouky, and M. Cukurova, "Generative AI for Learning: Investigating the Potential of Learning Videos with Synthetic Virtual Instructors," in Communications in Computer and Information Science (AIED 2023), 2023, doi: 10.1007/978-3-031-36336-8_81.

[25] M. Lamrabet, H. Fakhar, N. Echantoufi, K. El Khattabi, and L. Ajana, "Revolutionizing Teachers' Professional Development: The Critical Role of AI-Based Tools from Initial Training to Lifelong Learning—A Case Study," International Journal of Information and Education Technology, vol. 15, no. 4, pp. 696–706, Apr. 2025, doi: 10.18178/ijiet.2025.15.4.2277.

[26] R. G. Berondo, "Harnessing the Power of Artificial Intelligence for Personalized Learning in Education," European Chemical Bulletin, vol. 12, no. 10, pp. 1243–1251, 2023.

[27] M. Naeem, M. Siraj, and S. Farooq, "Teacher and Administrator Perceptions of Using Artificial Intelligence in Education," Journal of Education & Social Sciences, vol. 11, no. 2, pp. 71–88, Oct. 2023, doi: 10.20547/jess1122311205.

[28] Q. Huang, P. J. Kumarasinghe, and G. S. Jayarathna, "Navigating the Future: Exploring AI Adoption in Chinese Higher Education Through the Lens of Diffusion Theory," Interdisciplinary Journal of Information, Knowledge, and Management, vol. 19, Article 9, 2024, doi: 10.28945/5277.

[29] S. E. S. D. Ibrahim, "Using Artificial Intelligence to Develop Preparatory School Pupils' EFL Vocabulary Aspects," Journal of Education in the 21st Century, vol. 7, no. 45, pp. 129–146, 2025.

[30] S. E. S. D. Ibrahim, "Using Artificial Intelligence to Reduce Preparatory School Pupils' Writing Apprehension," Journal of Education in the 21st Century, vol. 7, no. 45, pp. 151–175, 2025.

[31] S. Sahari, "Measuring the Familiarity, Usability, and Concern towards AI-Integrated Education of College Teachers at the Undergraduate Level," Sch J Arts Humanit Soc Sci, vol. 12, no. 5, pp. 166–176, May 2024, doi: 10.36347/sjahss.2024.v12i05.002

[32] Y. Doğan, V. Batdı, Y. Topkaya, S. Özüpekçe, and H. V. Akşab, "Effectiveness of Artificial Intelligence Practices in the Teaching of Social Sciences: A Multi-Complementary Research

Approach on Pre-School Education," Sustainability, vol. 17, no. 7, p. 3159, Apr. 2025, doi: 10.3390/su17073159.

[33] W. Nouari, S. Mouas, M. E. A. Hraki, and F. K. Berrahal, "Innovating Narratives or Stifling Creativity? Assessing the Impact of GPT on Science Fiction Writing Skills Among Master Students in Algeria," Educational Administration: Theory and Practice, vol. 30, no. 6, pp. 1494–1507, 2024, doi: 10.53555/kuey.v30i6.5526.

[34] A. S. Abdelmagid, A. Y. Al-Mohaya, A. M. Ibrahim, A. A. Teleb, and N. M. Jabli, "The Impact of Artificial Intelligence Applications on Developing Levels of Cognitive Depth of Information among Postgraduate Students," International Journal of Learning, Teaching and Educational Research, vol. 24, no. 4, pp. 316–333, Apr. 2025, doi: 10.26803/ijlter.24.4.15.

[35] M. M. Harun, N. M. Nasri, and R. Derahman, "The Mediating Role of ChatGPT in the Relationship Between Academic Workload, Learning Facilities, and Academic Efficiency," International Journal of Scientific Research and Engineering Development, vol. 8, no. 1, Jan.–Feb. 2025, doi: 10.5281/zenodo.14716248.

[36] B. P. Woolf, Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Burlington, MA, USA: Morgan Kaufmann, 2009.

[37] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," Wiley Interdiscip. Rev. Data Mining Knowl. Discov., vol. 10, no. 3, e1355, 2020.

[38] D. Sinha, "The transformative impact of AI on education: Opportunities and challenges," Johns Hopkins Univ., Working Paper, 2024.

[39] J. Kulik and J. Fletcher, "Effectiveness of intelligent tutoring systems: A meta-analytic review,"
J. Educ. Psychol., vol. 108, no. 6, pp. 1042–1053, 2016.

[40] E. Nusivera, A. Hikmat, and A. R. A. Ghani, "Integration of ChatGPT usage in language learning model to improve argumentation skills, complex comprehension skills, and critical thinking skills," [Journal details], 2024.

[41] S. B. Sereno III, "AI as academic partner: Evaluating ChatGPT's impact on learning outcomes and perceptions among business administration students," figshare, Preprint, 2024.

[42] B. S. Bloom, "The search for methods of group instruction as effective as one-to-one tutoring," Educ. Leadership, vol. 41, pp. 4–17, 1984.

[43] K. Şık Keser and R. Pektaş, "Human vs. AI feedback on academic writing: A comparative study in Turkish context," J. Lang. Educ. Technol., vol. 15, pp. 102–118, 2025.

[44] M. Wan, L. J. Ghee, T. H. Joo, and C. S. Kee, "Data-driven approach to flipped learning for differentiated instruction in electronic engineering," Singapore Polytechnic, Educ. Innov. Rep., 2024.

[45] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Rev., vol. 46, no. 5, pp. 30–40, 2011.

[46] M. Bültemann, K. Simbeck, N. Rzepka, and Y. Kalff, "Sustainable learning analytics: Measuring and understanding the drivers of energy consumption of AI in education," HTW Berlin, Working Paper, 2024.

[47] P. H. Winne and A. F. Hadwin, "The psychology of self-regulated learning," in Handbook of Educational Psychology, 2nd ed. Mahwah, NJ, USA: Erlbaum, 2006.
[48] P. Bourne, "Integrating artificial intelligence in education: Opportunities, challenges, and policy implications for Jamaica," J. Educ. Policy Technol., vol. 3, no. 2, pp. 112–129, 2025.

Appendix

Sl no.	Authors (Year)	Country	Level	Design	Sample	Focus / Key finding
1	A. Das, S. Malaviya, M. Singh (2023)	India (ICFAI Univ., Dehradun; plus US industry co-author)	Upper-secondary	Quasi-experimental, mixed methods	n=100 (Exp=50, Ctrl=50); exp group d≈2.70	AI-driven personalizat boosts post-test scor engagement correlates (r≈0.63, p<.01).
2	S. Dangprasert (2025)	Thailand (KMUTNB, Bangkok)	Higher education / research skills	Mixed methods (pre/post, survey, interviews)	n=30 researchers	"CRAFT" AI-augmen literature reviews: reduction; ~92% prec multivariate gains.

3	R. Yavich (2025)	Israel	Higher education (advanced mathematics)	Quasi-experimental with pre/post; mixed methods	n=50 (Exp=25, Ctrl=25) within underprepared subgroup of N=100 first-years	AI tools (SOWISO/ChatGPT/W significantly improve g effects with scaffolded feedback.
4	D. Rad et al. (2024)	Romania (6 universities)	Higher education	Cross-sectional survey; sequential mediation	n=675 students	Psychological influe perceived benefits mediated by rapid A career/education per positive paths.
5	T.	Indonesia (Universitas Terbuka)	Higher education (distance)	Quasi-experimental (GAI-tutor vs human tutor), Mann-Whitney U with Cliff's Δ	n=37,743	GAI-assisted classes discussion activity an assignment scores; to significantly higher (p

	Belawati, D. Prasetyo (2025)				student records across 4 courses (500 AI-assisted classes vs 500 control)	
6	G. Hanshaw, E. Wicker, K. Denlinger (2025)	United States (LAPU)	Higher education (fully online)	Quasi-experimental, non-equivalent groups with PSM; Mann-Whitney U, Wilcoxon, permutation tests	2,090 student-course combos (1,338 students) across 99 courses	Using the AI course a associated with sign GPAs (before PSM: U after PSM: Wilcoxo paired mean Δ≈0.377
7	M. Han, H.	China (data) / Malaysia (affiliation)	Higher ed (Accounting)	mixed methods; SEM	n=985	AI readiness **mediates** communicative / interac

	Mustafa A. Razak, S. Kharuddin (2024)			+ focus groups	undergrad accounting students	/ autonomy features — performance
8	M. Chen (2025)	UK (affiliation)	mixed levels (systematic review)	systematic review + meta-analysis (2019–2024)	45 studies	AI personalised lear (medium-large) impr SD academic; ~0.42 S
9	K. Kalam et al. -2025	USA	Higher ed (medical school)	randomized controlled trial	n=33	ChatGPT group ha higher initial scores materials; no signi advantage after 1 weel
10	A. Garg	India	Higher ed (engineering PG)	pre/post experiment (paired t)	n=20	structured prompt trai produced significant analysis (d≈0.89; p<.0

	& R. Rajendran (2024)					
11	S. K. Nemani (2025)	USA	Mixed (STEM; K-12 & HE discussed)	Systematic literature review + bibliometric analysis + educator survey	n=150 educators	AI-driven personalize performance & engage skews toward private s widening inequi access/ethics addresse
12	S. Valizadeh, R. Gunday, R. Sahmaniasl (2025)	Turkey	Secondary / language schools (EFL, B1–B2)	Quasi-experimental + mixed methods (pre/post tests, surveys, interviews)	n=60 (30 AI; 30 control)	AI-powered cultural significantly imp comprehension; larg Δ≈+25 pts; large effec

13	M. S. Alhenak y, M. A. Alharthi -2025	Saudi Arabia	Middle school (Grade 7, girls)	Quasi-experimental (pre/post; exp vs control)	n=48 (24/24)	AI-based chatbot p Digital Skills cour improved computatio control (post-test p=0. large).
14	M. N. A. Mohame	Malaysia	Higher ed (university	Mixed-methods	n not stated (cross-	AI chatbot (Dialogflo use, increased motivati

	d; Y. T. Gooi; J. Wang (2024)		health education)	(questionnaires + chatbot interaction logs)	sectional survey with chatbot sessions)	boredom; students im recognition and understanding; recon content & interactivity
15	B. N. Obenza et al. -2024	Philippines	Higher ed (multiple HEIs)	Non-experimental correlational (PLS-SEM; moderation)	n=372 college students (Region XI)	Cognitive absorption AI literacy (effect ≈ statistically signific moderator (≈ **0.011**); **37.8%** of variance.
16	L. T. T. Nga, N. T. Hang,	Vietnam	Higher ed (1st-year English majors)	Quasi-experimental (exp vs control; pre/post) + survey	n=66 (33 exp with ChatGPT/Quillbot/4English; 33 control)	AI-assisted practice improved presentation pronunciation, confide in exp vs control; with p<.05.

	P. N. Anh, V. T. Chon, P. T. Diem, N. T. H. Trang (2025)					
17	E. D. Milakis, C. C. Argyrakou, T. Loukisas, D. G. Vangeli, M. C. Katsarou, A.	Cyprus/Greece/UK	K-12 (grade 8 STEAM scenario; educator evaluation)	Embedded case study; expert rubric; one-way ANOVA, Tukey HSD; ICC=0.876	4 LLMs (ChatGPT-4o, Claude 3.5 Sonnet, Gemini, LLaMA-3 70B) rated by 5 STEAM educators	Claude scored h ($M \approx 3.06$); ChatGPT st understanding/applica weak on **differentiatio** Significant difference (e.g., $F(3,16)=7.152$, p

	Lalou (2024)					
18	M. Tripathy, B. Sahu (2025)	India	Higher ed (UG Science vs Arts)	Cross-sectional comparative survey (GAAIS); independent z-test	n=200 (100+100)	Science students sho more positive attitude students (z=2.874 Implication: discipli literacy/curriculum.
19	T. Tripathi, K. Randhaw	India	Mixed: secondary, higher ed, adult learning	Mixed-methods, quasi-experimental	n=120 (60 treatment / 60 control)	AI-based metacognitio MAI scores (t=4.57, p accuracy, time, errors; (all UEQ factors p<0.001)
	a, D. K.					

	Mathur, S. Verma, K. Rathore -2024			(6-week AI tool vs control)		
	C.-P. Dai, F. Ke, Y.					AI-powered virtua simulation-based lea **medium positive effe** CI [0.27, 0.59]); mattered: **interventi technology used, representation**; educa level/domain generally

20	Pan, J. Moon, Z. Liu (2024)	Multi (meta-analysis)	Mixed (K-12 & HE)	Meta-analysis (22 studies, 49 ES)	—	
				Quasi-experimental field study embedded in	250 students,	

21	S. T. Roodsari , S. A. Ghanbari (2025)	Germany	Higher education (Leipzig & TU Dresden)	LMS (AI formative-feedback system aligned	~2,000 submissions (45 resubmissions analyzed for t-test)	AI-generated, comp feedback **significa** unachieved learnin (p<.001; **Cohen’s c** large gains when fee tied to objectives.
				to competencies)		

	S. Mandina, R.			Mixed methods; quasi-experimental (pre/post; ANCOVA) + interviews/observations		AI tools (ITS, v assessment) led to **larg** ANCOVA **F(1,117)=1** **η^2=0.617**; ~**+20** improvement; engag gender or prior-knowl effects.

22	Kusakara (2025)	Zimbabwe	Secondary (Chemistry)		n=120 (60 AI; 60 control)	
23	S. Gökoğlu, F. Erdoğdu -2025	Multi	Mixed (K-12 & HE)	**Meta-analysis** of GenAI interventions	31 studies, n=2,646	GenAI has a **moderat** on learning performar Significant moderator **duration**, **GenAI tool** no critical publication
	D.			RCT (AI-		Both groups had signi gains (p<.001); r **difference** between A traditional video (p

24	Leiker, A. R. Gyllen, I. Eldesouky, M. Cukurova (2023)	Multi (UK/US/NL)	Adult/professional online learning	generated synthetic instructor video vs. traditional) + surveys	n=83	synthetic instructor v viable substitute f produced videos.
	M. Lamrabet, H.					

	Fakhar, N.					Higher knowledge, **mastery of emergin** predict teachers' perc of integrating AI-ba initial training th learning; strong relia and good sampl (KMO=0.922).
	Echantoufi, K. El Khattabi,			Quantitative survey; EFA/CFA;		
	L. Ajana		Teacher education (preservice + in-service)	ordinal logistic regression		
	-2025				n=482 (244	

25		Morocco			future teachers; 238 in-service)	
26	R. G. Berondo (2023)	Philippines	Mixed levels (elementary–college)	Mixed-methods; pre/post + surveys/interviews	n=100 students	AI-supported person **significantly** improve vs. pre-test (paired t reported effect size ≈ report higher engagem re: teacher training, pr

27	M. Naeem, M. Siraj, S. Farooq (2023)	Pakistan	Mixed (school–HE; teachers & admins)	Cross-sectional survey; PLS-SEM (SmartPLS 4)	n=385	"Influence from colle links from academic l support, and percepti **AI integration**; practi norms & institutional
	Q. Huang, P. J.					**Perceived Relativ** ($\beta \approx 0.48$, $p<.001$) a ($\beta \approx 0.39$, $p=.005$) sign intention to compatibility/complex not significant; mo (SRMR<.08; **$R^2 \approx 0.70$**

28	Kumarasinghe, G. S. Jayarathna (2024)	China (Hangzhou)	Higher ed (management undergrads)	Quantitative; PLS-SEM with CTA; random survey	n=312	
						AI-supported instructi **improved** EFL voca (and related writing vs. traditional teac difference large (e.g.

29	S. E. S. D. Ibrahim (2025)	Egypt	Preparatory (middle school, EFL)	Quasi-experimental (exp vs control; pre/post)	n=60 (30/30)	p<.001; η²≈0.35).
30	S. E. S. D. Ibrahim (2025)	Egypt	Preparatory (middle school, EFL)	Quasi-experimental (exp vs control; pre/post)	n=60 (30/30)	AI-based program r **apprehension** and e performance; exper outperformed control (significant; large prac

31	S. Sahari (2024)	India	Higher ed (undergrad college teachers)	Mixed methods; survey + CFA & χ^2	n=441 (5 states)	Familiarity, usability toward AI-integrated e **gender** and **professi** shows positive links (f
						usability & concerns teachers report hi females show higher p concerns.

	Y.			Multi-Complementary Approach (McA): meta-analysis + meta-thematic		
	Doğan,			+ experiment (pre/post, control)		
	V. Batdı, Y.				Meta-analysis (2005–2025; I^2=82.82; REM) +	Across the synthesis a trial, AI applications preschool learni complementary phas favor of AI use in education.
	Topkaya, S.				classroom experiment	
	Özüpekçe, H. V.					

32	Akşab -2025	Türkiye	Pre-school (social sciences)			
	W. Nouari, S. Mouas, M. E. A. Hraki, F. K.			Quasi-experimental + rubric evaluation (EG with GPT vs CG)	n=42 (21/21),	GPT-supported group imaginative writi exploration, science plot development; stu can catalyze creativity

33	Berrahal -2024	Algeria	Higher ed (MA, English)		one-semester SF writing task	
					n=35 master's students (AI platforms group n=17 vs. traditional Blackboard n=18) in "Computers in Education," King Khalid Univ.	

	Abdelmagid et al., 2025		Postgraduate (Higher Ed)	Quasi-experimental (two groups, pre–post)		Cognitive depth (rec strategic thinking reasoning), reliability

34		Saudi Arabia				
	Harun, Nasri & Derahma n, 2025			Quantitative cross-sectional; PLS-SEM	Polytechnic Sultan Idris Shah students; survey (sample size not stated in paper; G*Power used for planning)	

35		Malaysia	Higher Ed			Academic workload, l → (mediator) ChatGPT efficiency