

# Analysing the Effectiveness of Different AI-Based Tutoring Systems and their Impact on Education across Global Contexts: A Literature Review

Danyal Asgar<sup>1</sup>

*Received June 03, 2025*

*Accepted September 28, 2025*

*Electronic access October 31, 2025*

This literature review examines the role of AI-based tutoring systems in education across the globe, with particular emphasis on how they affect student academic performance, motivation, and knowledge retention. Drawing on 48 peer-reviewed studies from diverse educational contexts and countries, it compares the effectiveness of AI tutors against human tutors and traditional classroom instruction. The review identifies strong correlation with major benefits, such as improved results in STEM subjects, enhanced student motivation through adaptive learning, and better knowledge retention. However, the analysis also incorporates mixed and negative findings - including evidence that over-reliance on AI can reduce critical thinking, creativity, and independent problem-solving, a phenomenon known as cognitive offloading. Some studies reported only modest gains compared to traditional instruction, and others found no significant improvement in performance. The review further addresses ethical considerations such as data privacy, algorithmic bias, and equity of access. The findings suggest that while AI-based tutoring systems can offer strong, personalised learning experiences, their impact is highly context-dependent, varying by subject, age group, and implementation quality. Effectiveness is maximised when AI is used to complement, rather than replace, human instruction, ensuring both improved learning outcomes and the development of higher-order cognitive skills. The review clearly recognises AI's potential in education; however, it does not get into the technical aspects of AI systems, like how they are built, or their application in non-STEM subjects. Future studies should focus on analysing the impact of these AI systems on different age groups, conducting RCTs, and including more perspectives from teachers and learners on how these platforms have affected them.

**Keywords:** Artificial Intelligence, Intelligent Tutoring Systems, AI in Education, Student Engagement, Educational Outcomes, Data Privacy, Adaptive Learning, Algorithmic Bias

## Introduction

The incorporation of AI into education represents one of the greatest pedagogical shifts of the 20th and 21st centuries. Broadly defined as the capacity of machines to replicate key aspects of human thinking, such as learning, reasoning, and problem solving, AI is increasingly influencing how educators and researchers rethink foundational ideas about how teaching and learning should be structured.

In recent years, AI technologies, particularly those embedded in Intelligent Tutoring Systems (ITS), adaptive learning platforms, and natural language interfaces, have been adopted to enhance personalised instruction, automate administrative tasks, and support data-driven teaching interventions<sup>1</sup>.

AI in education is further being used to address long-standing challenges in teaching, such as providing individualised feedback at scale and meeting the diverse needs and learning styles of different students. AI-based systems can now tackle these

problems by offering one-on-one tutoring experiences that adapt dynamically to a student's pace, learning style, and level of mastery<sup>2</sup>, capabilities historically limited to human tutors. According to Baker & Inventado (2014), AI in education is also beginning to be used for tracking student engagement in real time, predicting learning outcomes, and recommending resources<sup>3</sup>.

From a theoretical standpoint, the promise of AI in education aligns closely with several foundational learning theories:

- Constructivist Theory<sup>4-6</sup>, for instance, argues that learners actively construct knowledge through interaction and feedback; AI tutoring systems, particularly those that adjust in real time to learner input, enable exactly such feedback loops.
- Cognitive Load Theory<sup>7-9</sup> is also highly relevant, suggesting that student knowledge retention depends on factors such as information difficulty and amount; adaptive platforms do exactly this, altering extraneous load and tailoring content complexity to a student's current working memory

<sup>1</sup> Independent Student Researcher

---

capacity.

- Furthermore, Self-Determination Theory<sup>10-12</sup> suggests that motivation increases when learners feel a sense of autonomy, competence, and relatedness - three elements that personalised AI tutors are increasingly designed to support.

These theoretical frameworks have been discussed in the context of AI in education by Holmes et al. (2019), Luckin et al. (2016), and Chen et al. (2020), who emphasise that AI systems can be purposefully designed to support these principles in practice<sup>1,13,14</sup>.

The COVID-19 pandemic further accelerated the adoption of AI tools in education. As institutions worldwide moved to online learning overnight, the need to address concerns related to AI-based tutoring systems became even stronger. Holmes et al. (2019) point out the many challenges around the implementation of these AI systems, particularly concerning data privacy, equity, and student-teacher relationships<sup>13</sup>.

Taking these factors into consideration, there is an increasing need to critically assess the effectiveness of AI-based tutoring systems and its impact on students.

## Purpose of the Literature Review

This literature review aims to evaluate research findings on the effectiveness of AI-powered tutoring systems and to assess their influence on student academic performance and engagement. It also intends to look into the optimal conditions for these AI technologies to work, and critically identify their risks, flaws, and ethical issues.

Additionally, the review also aims to explore how the success of these methods can vary between countries and school systems.

### Key Research Questions:

1. What are the distinguishing traits and functionalities of AI-based tutoring systems?
2. How effective are AI-based tutoring systems in improving student performance in comparison to conventional teaching methods?
3. How do AI-based tutoring systems affect student engagement and motivation?
4. What are the conditions under which AI-based tutoring systems are most successful?
5. What are the main obstacles and moral dilemmas surrounding the application of AI in education?
6. How do different global contexts affect the efficacy of AI-based teaching systems?

## Methodology of Collecting Data

To identify relevant studies for this literature review, academic databases such as Google Scholar, ERIC, and ResearchGate were searched using a combination of key terms including “AI tutoring systems,” “intelligent tutoring systems,” “adaptive learning platforms,” and “artificial intelligence in education.”

The initial search yielded 176 articles published between 2005 and 2025. After removing duplicates and screening titles and abstracts, 94 articles were shortlisted for full-text review. Following closer examination, 48 studies met the inclusion criteria and were included in the final synthesis.

Inclusion criteria consisted of peer-reviewed articles or credible research reports that:

- Focused on AI-based or intelligent tutoring systems,
- Examined student outcomes such as academic performance, engagement, or retention,
- Included clear descriptions of the AI tool or system used.

Exclusion criteria included:

- Non-peer-reviewed opinion pieces,
- Papers focused solely on general EdTech tools without an AI component,
- Studies without any measurable outcomes or relevance to tutoring systems.

Each included study was assessed for relevance and credibility. Studies based on randomised controlled trials (RCTs), quasi-experimental designs, or large-scale observational research were prioritised. No formal scoring tool (e.g., GRADE) was used for quality appraisal, but methodological robustness was noted qualitatively in the analysis.

From each selected study, the following data was extracted:

- Author(s) and year of publication
- Country/region of the study
- Educational level or student demographic
- Type of AI system used (e.g., adaptive learning platform, chatbot tutor, etc.)
- Study design and methodology
- Reported outcomes (e.g., performance gain, engagement impact, retention)

- Any noted limitations, biases, or contradictory findings.

To analyse and summarise the findings, a thematic coding approach was applied. Initially, themes such as “academic outcomes,” “student engagement,” “motivation,” “accessibility,” and “ethical concerns” were identified. These themes were emergent, derived from recurring topics across the reviewed literature, not predetermined. Coding was done manually and iteratively by the author to reduce bias, with repeated reviews of the extracted data to ensure consistency.

No secondary reviewer was available for inter-coder reliability; however, multiple review passes were conducted to minimise subjectivity.

This method enabled the review to synthesise findings across global contexts and educational levels, while also identifying gaps, contradictions, and emerging trends in the application of AI-based tutoring systems.

In addition to peer-reviewed journal articles and academic books, this study includes a small number (3) of reliable non-academic sources to provide up-to-date context and implementation details on government plans for AI systems.

Two reputable news sources, Reuters and China Media Project, were consulted for coverage of recent AI tutoring deployments and policy changes in China. These were selected based on their established reputations for fact-checking and reliability.

Official data and policy information were obtained from the United Arab Emirates Government’s official AI portal. This source was included due to its authoritative nature and relevance to national AI-in-education initiatives.

All non-academic sources were cross-checked where possible against multiple outlets or official statements to mitigate bias and ensure factual accuracy. These materials were used strictly for descriptive and contextual purposes and were not treated as primary evidence when evaluating causal claims.

## Key Concepts

Artificial Intelligence (AI), as explained by Russell & Norvig (2021), refers to systems or machines that can react to stimuli to perform specific tasks<sup>15</sup>, and can iteratively improve themselves based on the information they collect. Typically, these systems attempt to mimic human intelligence.

AI-based tutoring systems, a subset of educational technology, use AI to deliver personalised feedback, instruction, and learning pathways based on individual student data<sup>2</sup>. These can be categorised into a few categories, the main ones being Intelligent Tutoring Systems (ITS), Adaptive Learning Platforms, or AI-driven Chatbots.

For example:

- Cognitive Tutor is an intelligent tutoring system developed by Carnegie Mellon University; it observes how students

solve math problems and provides practice questions and feedback accordingly.

- Similarly, Duolingo is an adaptive language learning platform that changes teaching method and style based on user mistakes and progress.
- Jill Watson is an AI chatbot used at Georgia Tech University during online classes, which assists students with course material information.

Learning is defined as “a relatively permanent change in knowledge or behavior resulting from experience”<sup>16</sup>. In educational contexts, it often refers to a student’s ability to retain, apply, and build upon newly acquired knowledge or skills.

Engagement refers to the cognitive, emotional, and behavioural investment a learner commits to their academic tasks<sup>17</sup>. It includes sustained attention, participation, curiosity, and enthusiasm during the learning process. AI systems attempt to boost engagement through interactive design, real-time feedback, and adaptive pacing.

Motivation in this review is understood through the lens of Self-Determination Theory<sup>10</sup>, which highlights intrinsic motivation (driven by interest or enjoyment) and extrinsic motivation (driven by rewards or outcomes). AI systems may influence motivation by offering autonomy, competence-based progression, and goal-setting features.

Student Learning Outcomes refer to the knowledge gained, skills developed, values formed, and ability to apply learnings outside of a classroom environment. In layman’s terms, student learning outcomes are what a student is expected to learn, feel or do after completing a learning course. Baker and Inventado (2014) emphasise that AI tools not only improve education, but help students understand their lessons better, stay motivated, and remember what they have learnt for a long period of time<sup>3</sup>.

## Historical Evolution

With the creation of SCHOLAR, an Intelligent Tutoring System that taught geography using semantic networks, the exploration of artificial intelligence in education began as early as the 1960s<sup>18</sup>. This early system aimed to simulate one-on-one tutoring by adapting its questions and responses based on student input.

Furthermore, as noted in Anderson et al. (1995), the 1980s and 1990s marked a period in which AI and cognitive science began to merge, giving rise to sophisticated tutoring systems like SOPHIE and the Cognitive Tutor<sup>19</sup>. These systems incorporated models that tracked student understanding and offered customised instruction, particularly in STEM education.

With the advent of machine learning and natural language processing in the 2000s, AI tools began to scale in both complexity and reach. Adaptive learning platforms like ALEKS, Knewton

and DreamBox Learning emerged, offering personalised pathways for learners. More recently, AI-driven virtual assistants, automated essay scoring systems, and real-time learning analytics have become part of mainstream educational technology<sup>1</sup>.

The global shift to remote learning during the COVID-19 pandemic further accelerated the adoption of AI-based solutions, demonstrating their utility in maintaining instructional continuity, offering on-demand support, and collecting learning data for decision-making<sup>13</sup>.

level mathematics and chemistry courses using the ALEKS (Assessment and Learning in Knowledge Spaces) system compared to those in traditional formats<sup>24</sup>.

However, as VanLehn (2011) points out, AI tutors are most effective when used to support and enhance traditional instruction - not replace it<sup>21</sup>. Their success relies on thoughtful implementation, good system design, and active teacher involvement. When teachers use AI-generated insights to guide and personalise their instruction, students benefit the most.

### A Timeline of AI Tutoring Evolution

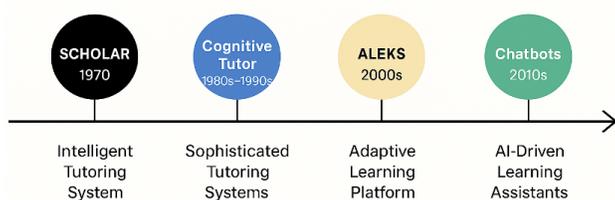


Fig. 1 Evolution of AI Tutoring Systems (adapted from Woolf, 2010)

## Effectiveness of AI Tutoring Systems against Traditional Tutoring

A key question in educational research is whether or not AI tutoring systems can outperform traditional teaching. Numerous studies have shown that, in certain contexts, students using AI tutoring systems tend to perform better on assessments than those in traditional classroom instruction settings<sup>20,21</sup>.

In a meta-analysis conducted by Steenbergen-Hu & Cooper (2014) analysing close to 40 studies, it was found that AI tutoring systems helped students outperform those in traditional classrooms, particularly in math and science; the scores reported were close to those achieved through one-on-one human tutoring<sup>20</sup>. This suggests that AI tutors, by providing personalised instruction, instant feedback, and adaptive practice, can be a strong alternative or complement to conventional teaching. This conclusion was also supported by Kulik & Fletcher (2016), in which 50 studies were analysed<sup>22</sup>.

The Cognitive Tutor, an ITS developed at Carnegie Mellon University, demonstrated this effect clearly. In a study involving 470 students, those using the Cognitive Tutor scored 15-25% higher on standardised tests and 100% higher on tests targeting specific algebra problem-solving skills compared to peers in traditional classes<sup>23</sup>. Similarly, a study by Hagerty and Smith (2005) reported higher success rates among students in college-

## Impact on Student Engagement and Motivation

Student engagement and motivation are crucial determinants of effective learning. AI technologies are believed to enhance engagement through adaptive feedback, real-time support, gamified elements, and multimodal content delivery<sup>13</sup>. In order for students to stay interested and learn better, AI should be tailored in a way to make learning more fun, interactive and personalised for students.

By tracking student behaviour in real time and determining when a student is disengaged or confused, AI systems can adapt the pace or material of a subject<sup>3</sup>. Some systems, for instance, involve using facial recognition and eye tracking software to assess students' levels of focus and attention, then making necessary adjustments to content or sending out alerts to keep students in their optimal learning environment - adequately challenged, yet not overwhelmed, maximising student engagement.

Gamification elements can also be used to raise motivation; incorporating game-like aspects into tutoring systems strongly correlates with increased time-on-task and improved overall student satisfaction<sup>25</sup>. Online platforms such as DreamBox Learning and Smart Sparrow have integrated such features to boost student motivation and engagement. Chatbots and similar systems also offer a conversational experience that helps lessen feelings of isolation, essential for such online learning environments.

AI systems are often viewed positively by students, especially when they are intuitive to use and offer prompt, personalised feedback. In addition, AI tutors can assist students in setting objectives, monitoring progress, and reflecting on their performance, supporting the development of self-regulated learning strategies<sup>26</sup>. This promotes positive long-term academic habits in addition to keeping students engaged in the short term.

## Impact of AI on Long-Term Learning Outcomes

Long-term learning outcomes refer to the lasting benefits and skills students gain from an educational experience - what they can retain, understand, and apply long after the course is over. This is something AI tutors can help with. Through utilising

## Student Engagement



**Fig. 2** AI Features Enhancing Student Engagement (adapted from Winne & Baker, 2013)

specific teaching scenarios and techniques, AI systems can help students maximise their gained knowledge in the long term.

By repeating important ideas at the right time, it is easier for students to recall different topics. Apps like Duolingo and MATHia work exactly this way - they help students practice certain subjects just when they are about to forget something<sup>27</sup>. At the same time, if students make any mistakes while working on new tasks, these systems can identify that the student has not fully mastered the earlier content, hence focusing more on consolidating those topics before moving on to new ones.

AI can also help students practice using what they've learned in real-life situations. For example, in a virtual science lab, students can test out ideas and learn from what happens. This helps students understand how to use their knowledge outside the classroom<sup>28</sup>, thereby increasing knowledge retention.

Students can learn how to study better and become more independent when shown how they're doing, and given tips on how to improve<sup>29</sup>. Many AI tools show students these statistics, helping them think about how they learn.

### Challenges and Ethical Considerations

Although AI-based tutoring systems offer many benefits, they also bring with them a range of important challenges and ethical issues that must be discussed to ensure responsible implementation; specifically, those related to student data privacy, over reliance on technology and access disparities.

The first key issue is that these systems often rely on the

collection of detailed learner data, including behaviour, performance, how they spend their time, and sometimes even biometric indicators. While this does enable personalisation, it also raises serious concerns about student privacy and data security<sup>30</sup>. Institutions must implement clear policies regarding data usage, storage, and consent to safeguard student rights.

Another major concern is the presence of algorithmic bias. If trained on data containing inherent biases, or a lack of diversity, AI tutors can provide feedback with unequal recommendations, based on factors such as race, gender, or language background<sup>31</sup>. Hence, the government, AI educators, and developers must ensure that AI is trained on fair, transparent and diverse data. Such data must be gathered from a wide range of groups, and checked for biases by external auditors and reviewers. For effective implementation, AI systems require a joint effort and each stakeholder plays an important role in monitoring, regulating and improving ethical practices.

There are concerns that increased reliance on AI may lead to the devaluation of human teachers, reduce human interaction in learning environments, or even render humans obsolete in such environments. However, research emphasises that AI should augment, not replace, teachers. Educators play a crucial role in interpreting AI feedback, providing emotional support, and fostering critical thinking<sup>13</sup>. Balanced integration is key to maintaining the human aspect of education. In addition to this, students may feel isolated or disengaged if they interact mostly with machines. Physical health problems, like eye strain and headaches due to prolonged screen time, or poor posture for using laptops and tablets for extended hours can also be caused by engaging more with these digital AI systems<sup>32</sup>, although this is more so an effect of general device overuse.

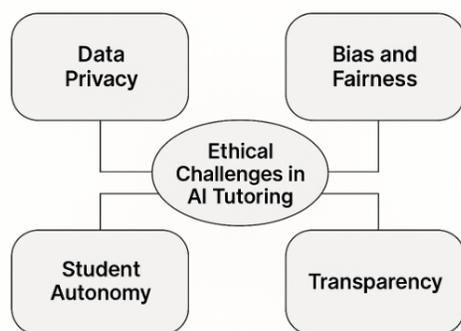
Access to AI-based tutoring systems often requires digital infrastructure, including stable internet and compatible devices. In many low-resource settings, this presents a barrier to adoption and exacerbates existing educational inequalities. Efforts must be made to ensure that technological advancements benefit all students, regardless of socioeconomic status<sup>33</sup>.

### Comparison of Educational AI Implementation in Different Countries

In an effort to modernise education and better prepare students for a rapidly changing future, countries around the world are increasingly investing heavily into AI tutoring solutions and programs. Yet despite the widespread enthusiasm, the effectiveness and integration of these AI-based tutoring systems differs largely, depending on the country and the local educational context. Differences in national education strategies, technological infrastructure, cultural attitudes, and resource availability shape how AI is implemented and perceived in classrooms globally.

In China, the integration of AI in education is driven by strong

## Ethical Challenges in AI Tutoring



**Fig. 3** Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioural Scientist*, 57(10), 15101529.

government support and investment. AI technologies are being deployed at scale, with platforms like Squirrel AI providing personalised tutoring to millions of students. Large-scale studies have shown that students in China using AI-powered tutoring systems improved math scores by an average of 20% compared to traditional instruction<sup>34</sup>. The Chinese Ministry of Education has endorsed AI as a means to modernise education and close learning gaps, with AI literacy now included in national curricula<sup>35,36</sup> and university teachings<sup>37</sup>.

Similarly, South Korea has begun implementing AI based systems, in which students have personalised AI tutors with access to online learning platforms, allowing teachers to focus more on social and emotional aspects for their students.

As per Stanford's Global AI Vibrancy Ranking<sup>38</sup>, the United States leads in their overall AI standing. Yet this is not the case for the education sector; due to the fact that educational governance in the United States is decentralised, with significant power left to individual states and school districts, the adoption of AI in education varies widely. The Cognitive Tutor was highly effective in urban districts<sup>23</sup>, but follow-up studies in rural and underfunded schools highlighted lower adoption rates and inconsistent outcomes due to teacher resistance and technological barriers<sup>21</sup>. Institutions like Carnegie Learning and Khan Academy have made notable strides in adaptive learning technologies, yet access still remains uneven, particularly due to disparities in funding, technological infrastructure, and local policy priorities<sup>13</sup>. In some respects, the U.S. highlights the challenges of implementing new technologies within a highly independent system.

In the United Arab Emirates, substantial investment in digital infrastructure has enabled schools to pilot AI-driven learning platforms at scale. Although not exactly an AI tutoring system, the Ministry of Education has also integrated AI into school curricula, and promotes AI literacy through initiatives like the UAE

AI Camp and the establishment of specialised institutions, such as the Mohamed bin Zayed University of Artificial Intelligence. As part of its "UAE National Strategy for AI 2031" vision, the government envisions AI to be central in cultivating innovative, future-ready learners, and will take steps to ensure that AI is integrated heavily in the UAE ecosystem<sup>39</sup>. This heavy focus on AI could suggest that further AI-based tutoring systems could be launched soon in educational institutes in the country.

Unfortunately, in many low and middle-income countries, access to AI tools remains limited due to infrastructural challenges and costs. Nevertheless, mobile-first platforms and AI chatbots (e.g. using WhatsApp) are emerging as promising solutions to reach underserved learners, particularly in sub-Saharan Africa and South Asia<sup>33</sup>.

Global effectiveness of AI tutoring also depends on how well the tools are adapted to local teaching norms, languages, and cultural expectations. International collaboration is essential to ensuring global relevance and impact<sup>1</sup>.

## Synthesis and Discussion

The synthesis of the reviewed literature reveals a growing consensus on the potential of AI-based tutoring systems to enhance student learning across a range of contexts. AI has shown significant positive impacts on academic performance, especially in subjects like mathematics and science, where structured knowledge domains align well with algorithmic models, and these systems can foster engagement through adaptive feedback, gamification, and interaction, supporting long-term outcomes by promoting knowledge retention.

Across studies, several recurring patterns emerge, with successful AI implementations sharing essential traits including integration with classroom instruction, timely and constructive feedback, and personalised teaching.

Research also highlights that AI's effectiveness is context-dependent. Factors such as infrastructure, teacher involvement, policy support, and alignment with curricular goals significantly influence the success of AI-based systems. Countries with centralised strategies and high technological readiness (e.g. China) report more comprehensive deployment, whereas decentralised systems (e.g. the USA) exhibit more innovation but less consistency.

Despite its promise, several gaps remain; there is a lack of studies assessing the sustained impact of AI on learning; ethical considerations - particularly around bias, privacy, and access - remain under-researched in practice; and most AI systems are optimised for STEM fields, with limited adaptation to humanities or interdisciplinary learning.

Future research should prioritise inclusive, and interdisciplinary investigations. Policy frameworks should focus on equity, ethical design, and professional development for educators.

---

Collaboration among technologists, educators, and policymakers is essential to align AI development with educational values and societal needs.

## Comparative Insights and Contradictions in the Effectiveness of AI Tutoring Systems

While the reviewed literature generally presents overall positive impressions of AI tutoring systems, a deeper comparative review reveals substantial variation, methodological contradictions, and contextual dependencies that are often overlooked but must be discussed to gain a complete understanding of the effectiveness of AI educational systems.

This section synthesises findings from studies across educational domains, age groups, geographies, and experimental designs, while also discussing contradictions that challenge the narrative of AI as a universally effective educational solution.

### Variation by Subject Domain: STEM vs. Humanities

AI tutoring systems have demonstrated the greatest efficacy in STEM (Science, Technology, Engineering, and Mathematics) subjects, where structured problem-solving aligns well with algorithmic modelling. Steenbergen-Hu and Cooper's (2014) meta-analysis of 39 studies found statistically significant gains in math learning outcomes ( $g = .32$  to  $g = .37$ )<sup>20</sup>, with effect sizes being slightly lower than those of one-on-one human tutoring. These findings were reinforced by Kulik and Fletcher (2016), who analysed 50 studies and similarly noted robust effects in quantitative domains<sup>22</sup>.

This aligns with Cognitive Load Theory<sup>7</sup>, which suggests that structured, rule-based domains reduce extraneous load and allow AI systems to optimise germane cognitive processing, whereas open-ended domains require nuanced scaffolding that AI currently struggles to provide.

Results, however, have been mixed when used for writing, history, or reading comprehension. In their assessment of 146 papers on artificial intelligence in higher education, Zawacki-Richter et al. (2019) found that intelligent tutoring systems in non-STEM disciplines had restricted adaptive accuracy and were less effective in promoting deep, critical engagement with texts<sup>37</sup>. This inconsistency may stem from the open-ended, interpretive nature of humanities learning, which current AI systems, predominantly based on rule-based or decision-tree logic, struggle to replicate meaningfully.

This is further supported by studies conducted on the Cognitive Tutor system, reporting gains of up to 100% on domain-specific tests<sup>23</sup> - yet comparable systems for non-STEM subjects, like Writing Pal failed to produce statistically significant gains in essay-writing skills<sup>40</sup>, indicating subject-specific efficacy limits.

### Age-Group Differences in Learning Gains

It is uncertain to say definitively whether age plays a critical role in determining the impact of AI tutoring.

Evidence suggests that for younger students, especially in K-12 education, AI tutoring systems can provide adaptive learning support that helps improve engagement and learning outcomes modestly but positively. Younger learners with less prior knowledge and cognitive maturity tend to benefit from AI's ability to tailor instruction and provide repeated practice in an accessible way<sup>41</sup>. This was echoed in VanLehn's (2011) study, where ITSs proved more impactful for learners with low prior knowledge than for advanced students<sup>21</sup>.

These trends can be understood using Vygotsky's Zone of Proximal Development (1978)<sup>5</sup>, which holds that AI systems help bridge the gap between what students can accomplish on their own and what they can accomplish with assistance. Compared to adults who have established metacognitive methods, younger learners who lack considerable prior knowledge benefit more from this structured scaffolding.

However, at the same time, students at the college and university level have demonstrated more pronounced benefits from using AI tutoring systems. Studies in medical education and university STEM courses show AI tutoring can even outperform expert-led tutoring in some respects (e.g., faster conceptual knowledge acquisition and better retention)<sup>42</sup>. This may be due to older students' greater capacity for self-regulated learning and more intensive engagement with the AI systems' personalised feedback and adaptive challenges. These gains were consistent and unaffected by type of tutoring system<sup>20</sup>.

In comparison to older, adult learners, in contrast, college freshmen displayed higher improvements in mathematics test scores compared to traditional learning, of up to 20% using ALEKS<sup>24</sup>; a similar implementation among adult learners in community colleges showed no significant advantage over traditional instruction<sup>37</sup>. This suggests diminishing returns of AI systems as cognitive maturity and meta-cognitive strategies improve.

In sum, AI tutoring systems can be effective for both younger and older students, but a large number of studies state that the strongest and most consistent improvements in learning tend to be shown by students particularly in higher education settings. Adult learners display the least improvements, often not improving at all.

Therefore, a provisional/working conclusion can be made that AI systems are most effective when used by university level students. However, as direct comparison studies between age groups and across multiple subject domains are not widely available, any concrete conclusions cannot be made. More studies must be conducted in this area to validate for certain in which age group AI tutoring is most effective.

---

## Experimental Design Contradictions: RCTs vs. Observational Studies

A major challenge in the literature is the absence of rigorous, causal research designs. Although numerous studies report positive outcomes, few use randomised controlled trials (RCTs).

- Kulik & Fletcher (2016) analysed studies using quasi-experimental methods and found moderate effects<sup>22</sup>, but the absence of randomisation or control groups in many undermines causal interpretation.
- The Cognitive Tutor's reported substantial improvement in algebra skills is based on a quasi-experimental cohort design<sup>23</sup>. However, an RCT conducted by Pane et al. (2014) on Carnegie Learning's platform found only modest effects on student outcomes (effect size = 0.15)<sup>43</sup>, which contrasts sharply with earlier uncontrolled studies reporting higher gains. This raises concerns of confounding variables such as access to more experienced teachers or school funding levels skewing results.

More RCTs studying this were unable to be found, representing a gap in the literature. Future studies should focus on this area.

## Mixed and Negative Findings

Although positive results mostly dominate the literature, there are a small number of contradictory studies which reveal that AI may not in fact have as large an improvement in learning as initially revealed by most of the literature discussed above in this paper.

There are studies indicating that over-reliance on AI hints can worsen learning outcomes; students may use AI systems passively, allowing the algorithm to do the thinking, in a phenomenon dubbed "cognitive offloading"<sup>29</sup>.

A systematic literature review by Zhai et al. (2024) found that regular use of AI dialogue and tutoring systems was associated with a decline in students' cognitive abilities, diminished information retention, and an increased tendency for cognitive offloading<sup>44</sup>. Multiple studies cited by Zhai et al. warn that heavy dependence on AI tools can erode critical thinking and unassisted problem-solving, as students passively accept AI-generated outputs instead of engaging with the subject.

Krullaars et al. (2023) further argue that such over-reliance can reduce students' motivation and commitment to learning, and may even lead to a decline in the ability to evaluate sources and perform independent analysis<sup>45</sup>.

Constructivist concepts<sup>4,5</sup> emphasise that active learner engagement is integral for long-term knowledge retention. Passive acceptance of AI outputs in this case undermines self-regulated

learning processes<sup>46</sup>, negatively affecting metacognitive awareness and causing the opposite of the intended effect of these AI systems, worsening student learning instead.

Overall, although AI systems generally seem to improve test scores, there is also a major correlation between AI overuse and declines in problem solving ability and cognitive function. This reiterates, as mentioned previously in this paper, that for maximum effectiveness, AI systems must be used in complement with human tutors to allow for improved learning along with critical thinking, achieved when humans intervene and prevent over reliance on AI.

## Comparison of AI Tutoring System Effectiveness Globally

Unfortunately, there is a lack of RCTs to definitely compare the effectiveness of these systems across countries. However, based on the observational/quasi-experimental studies available, a more general comparison can be made.

From the findings, it can be concluded that effectiveness varies as much by study design and implementation context as by the tool itself: in Ghana, a quasi-experimental trial of the WhatsApp-based Rori tutor reported a large math gain ( $\sim 0.350.40$  SD) over 8 months<sup>47</sup>, likely amplified by mobile-first delivery, low baseline access to tutoring, and tightly scoped skills practice; in China, large observational and quasi-experimental studies of Squirrel AI show substantial improvements<sup>48</sup>, but effects are harder to interpret causally because of selection into programs and sustained policy support that improves fidelity; in the United States, randomised evaluations of Carnegie Learning/Cognitive Tutor generally find smaller, but credible, effects ( $\sim 0.100.25$  SD)<sup>43</sup> with gains concentrated in second-year implementations and schools with stronger teacher uptake; and in UK/LMIC pilots (e.g., adaptive or chatbot tools used via SMS/WhatsApp), studies often show engagement increases but mixed or null learning gains, with infrastructure, localisation, and teacher orchestration limiting impact<sup>49</sup>.

Overall, effects tend to be largest where access gaps are biggest, over-dependence is prevented, and implementations are tightly aligned to a narrow skill domain, while rigorous RCT contexts with varied classrooms yield more modest but reliable gains; policy backing and teacher integration appear to be consistent moderators of success.

## Comparison of Different AI Tutoring System Types and Their Effectiveness

The three main categories of AI tutoring systems (Intelligent Tutoring Systems (ITSs), Adaptive Learning Platforms, and AI Dialogue Systems (e.g., Chatbots)) vary significantly in structure, functionality, and educational impact. Each demonstrates differing levels of effectiveness depending on the subject domain, learner age, and implementation context.

**Table 1** Comparative Table of Findings from Multiple Studies

Study	Country	Age Group	Subject	AI Tool	Design	Effect Size	Result
Koedinger et al. (1997)	USA	HS	Math	Cognitive Tutor	Quasi-exp	0.30	Significant algebra gains
Pane et al. (2014)	USA	MS	Math	Carnegie	RCT	0.15	Modest gain
Hagerty & Smith (2005)	USA	College	Math	ALEKS	Quasi-exp	~0.25	First-year success
Roscoe et al. (2019)	USA	MS-HS	Writing	Writing Pal	Quasi-exp	No overall main effect of practice format	Benefits varied depending on prior literacy levels. Game-based formats were preferred by students but not consistently more effective for learning
Holmes et al. (2019)	UK	HS	Mixed	Various	Review	Mixed	Teacher pushback
Zhai et al. (2024)	China (multi)	Mixed	Mixed	Various Dialogue Tutors	Systematic Review	Negative correlation	Cognitive offloading, reduced retention and critical thinking
Krullaars et al. (2023)	Netherlands	HS/College	Mixed	Various AI Tutors	Correlational	Negative association	Decline in motivation and independent analysis
Henkel et al. (2024)	Ghana	Grades 3-9	Math	Rori What-sApp AI	Quasi-exp	~0.35	Significant Math Gains

1. **Intelligent Tutoring Systems (ITSs):** ITSs are highly structured systems that model student knowledge and provide real-time, personalised feedback during problem-solving tasks.

Studies such as Koedinger et al. (1997) and VanLehn (2011) have repeatedly demonstrated that ITSs like Cognitive Tutor and AutoTutor can outperform traditional classroom instruction, yielding moderate to large gains (effect sizes between 0.300,40) in structured areas like mathematics and science<sup>21,23</sup>.

While these systems can provide high effectiveness especially for low-performing students, they are limited by high development costs, having limited efficacy in open-ended domains, and being difficult to scale beyond tightly defined subjects<sup>37</sup>.

2. **Adaptive Learning Platforms:** These platforms (such as the ALEKS platform) use performance data to dynamically adjust the sequence and difficulty of learning materials for students, and are shown to be particularly effective for

practice and reinforcement.

These systems are easily scalable and suitable for homework support and self-paced learning, with Hagerty & Smith (2005) reporting ~25% score improvements using ALEKS in college-level math<sup>24</sup>. However, long-term retention and deep learning remain less studied, and these platforms are generally found to be weak at addressing misconceptions without human intervention<sup>50</sup>.

3. **AI Dialogue Systems:** These systems use natural language processing (NLP) to simulate conversational tutoring; examples include Jill Watson, Socratic, and domain-specific chatbots.

Recent studies like and Zhai et al. (2024) reveal mixed or negative outcomes - while such systems may improve short-term engagement, with high accessibility and low implementation costs, they often fail to support critical thinking or problem-solving, with some students becoming overly reliant on surface-level cues or AI-generated explanations<sup>44</sup>.

---

However, it must be noted that according to Henkel et al. (2024)<sup>47</sup>, in a study conducted in Ghana for students across 11 schools in grades 3-9, a WhatsApp based chatbot AI was able to yield significant improvement in student maths scores - this can be attributed to the study's quasi-experimental design, which may have skewed results, or, more likely, the researchers were able to prevent student cognitive offloading by limiting time spent with this chatbot to just two 30-minute sessions per week, striking a fine balance, allowing for maximum learning improvements without significant cognitive ability loss, proving the positive effects that these systems can cause if managed properly.

In conclusion, the most effective types of AI tutoring system are ITSs, which demonstrate the greatest gains for students - but this is restricted to structured, logic-based subjects. AI dialogue systems, on the other hand, exhibit promise in terms of engagement but are the least successful, potentially hindering crucial cognitive skills if not carefully monitored. Both adaptive platforms and ITSs consistently outperform traditional classroom instruction in improving student test results; nevertheless, all of the aforementioned AI platforms fall short when compared to one-on-one human tutoring, which can have effect sizes of 0.80 - 1.00<sup>51</sup>.

It can be argued that, as evidenced by RCTs, these AI platforms are not as effective as previously thought - however even then, they show consistent gains over traditional classroom instruction and so can still serve as a powerful tool in education, alongside human tutoring.

### Overall Patterns from Study Comparisons

- AI tools are effective in improving structured task performance (e.g., math drills) short-term, but less effective in conceptual or deep learning tasks.
- Lower-income students and geographies show limited benefit due to tech access, motivation, and lack of adult scaffolding.
- University students show more benefit than younger or adult learners.
- Novelty and gamification may increase engagement temporarily but don't guarantee comprehension or long-term retention.
- Government-driven systems (e.g., China) show more scalability; fragmented systems (e.g., U.S.) have inconsistent results.
- RCTs report flat or null effects, while self-reported or observational studies (e.g., Squirrel AI in China) report higher effectiveness.

- Ensuring over-reliance on these systems is prevented is key in achieving substantial results improvements.
- AI systems frequently fail in open-ended domains (e.g., writing, critical thinking, complex problem solving), where human reasoning and feedback are harder to replicate.
- ITSs are generally the most effective type of AI system in improving learning outcomes.
- One-on-one human tutoring remains unmatched in effectiveness by any of the AI systems.

### Limitations of the Review and Future Recommendations

Although this literature review covers many different aspects and effects of AI-based tutoring, in multiple settings, it does not address a few other crucial points, such as:

- The technical insights of how AI systems are created and operate.
- Detailed insights into teacher training, readiness, or challenges in implementing AI.
- Teachers' and students' lived, real-world experiences using AI tools.
- Long-term impacts of AI-based learning, from early age potentially into adulthood

Despite promising results, the evidence base around AI-based tutoring systems has critical gaps. Future work should prioritize the aforementioned areas to provide a more holistic, comprehensive understanding of AI's role in global education. Furthermore, future studies should address the impact of these systems on cheating, with different age groups, and using RCTs.

It should be noted that a large number of the studies reporting positive effects are correlational or quasi-experimental. Without random assignment, causality cannot be firmly established, and observed gains may be influenced by confounding factors such as teacher quality, student motivation, or institutional resources.

### Conclusion

This literature review has explored the multifaceted role of AI-based tutoring systems, with a particular focus on their effectiveness and impact on student learning, by analysing various studies. It can be concluded that AI technologies hold transformative potential in revolutionising education by significantly improving student engagement and academic achievement.

AI solutions such as Intelligent Tutoring Systems and adaptive learning platforms have demonstrated clear, measurable gains

---

in student achievement, particularly in structured subjects such as maths. These systems offer real-time, data-driven insights that support learners in maximising their academic performance and long-term learning outcomes. Beyond these metrics, these systems support the development of autonomous learning habits and increased student motivation, and can also ease pressure on educators by handling administrative tasks.

Nonetheless, challenges such as algorithmic bias, data privacy, teacher integration, and ethical concerns must be carefully managed to ensure fair and effective deployment. This requires more than just advanced technology; it demands thoughtful design, inclusive implementation, and continued research. Ensuring that AI serves as a complement - not a replacement - to human instruction is critical.

As AI continues to evolve, its role in education must be shaped by academic goals, equity considerations, and a commitment to supporting all learners. With the right balance of innovation, regulation, and educational vision, AI can serve as a powerful ally in addressing the diverse and dynamic needs of students worldwide.

## Acknowledgment

I would like to sincerely thank my computing teachers Miss Sameer and Miss Jamil, and my English teacher Miss Hunter for their guidance throughout the writing of this literature review. Their mentorship played a key role in helping the quality of my writing along with my knowledge of the subject, improving both the quality and depth of this work.

## About the Author

Danyal Asgar is a 17-year-old high school student based in Dubai. He aspires to study Artificial Intelligence at a top university in the US. His talent has led him to develop various agentic AIs and deploy them in real use, create an AI-powered education platform with hundreds of users, and intern at various top multinational firms. He enjoys badminton, and is an artist who has exhibited at various galleries across Dubai.

## References

- 1 R. Luckin, W. Holmes, M. Griffiths and L. Forcier, *Intelligence unleashed: An argument for AI in education*.
- 2 B. Woolf, *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*.
- 3 R. Baker and P. Inventado, *Educational data mining and learning analytics*.
- 4 J. Piaget, *The origins of intelligence in children*.
- 5 L. Vygotsky, *Mind in society: The development of higher psychological processes*.
- 6 R. Mayer, *Multimedia learning*.
- 7 J. Sweller, *Cognitive load during problem solving: Effects on learning*.
- 8 F. Paas and J. Merrinboer, *Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach*.
- 9 T. Jong, *Cognitive load theory, educational research, and instructional design: Some food for thought*.
- 10 E. Deci and R. Ryan, *Intrinsic motivation and self-determination in human behavior*.
- 11 R. Ryan and E. Deci, *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being*.
- 12 C. Niemiec and R. Ryan, *Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice*.
- 13 W. Holmes, M. Bialik and C. Fadel, *Artificial intelligence in education: Promises and implications for teaching and learning*.
- 14 L. Chen, N. Chen and J. Lin, *Artificial intelligence in education: A review*.
- 15 S. Russell and P. Norvig, *Artificial intelligence: A modern approach*.
- 16 M. Driscoll, *Psychology of learning for instruction*.
- 17 J. Fredricks, P. Blumenfeld and A. Paris, *School engagement: Potential of the concept, state of the evidence*.
- 18 J. Carbonell, *AI in CAI: An artificial-intelligence approach to computer-assisted instruction*.
- 19 J. Anderson, A. Corbett, K. Koedinger and R. Pelletier, *Cognitive tutors: Lessons learned*.
- 20 S. Steenbergen-Hu and H. Cooper, *A meta-analysis of the effectiveness of intelligent tutoring systems on college students academic learning*.
- 21 K. VanLehn, *The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems*.
- 22 J. Kulik and J. Fletcher, *Effectiveness of intelligent tutoring systems: A meta-analytic review*.
- 23 K. Koedinger, J. Anderson, W. Hadley and M. Mark, *Intelligent tutoring goes to school in the big city*.
- 24 G. Hagerty and S. Smith, *Using the ALEKS system to increase student success in a college mathematics course*.
- 25 Y. Kim, H. Park and Y. Baek, *Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning*.
- 26 P. Winne and R. Baker, *The potentials of educational data mining for researching metacognition, motivation and self-regulated learning*.
- 27 P. Pavlik and J. Anderson, *Using a model to compute the optimal schedule of practice*.
- 28 J. Kay, P. Reimann, E. Diebold and B. Kummerfeld, *MOOCs: So many learners, so much potential*.
- 29 P. Winne and A. Hadwin, *Studying as self-regulated learning*.
- 30 S. Slade and P. Prinsloo, *Learning analytics: Ethical issues and dilemmas*.
- 31 R. Binns, *Fairness in machine learning: Lessons from political philosophy*.

- 
- 32 W. H. Organization, *Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*.
  - 33 U.N.E.S.C.O., *Artificial intelligence and education: Guidance for policy-makers*.
  - 34 N. Maslej, L. Fattorini, R. Perrault, Y. Gil, V. Parli, R. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. Niebles, Y. Shoham, R. Wald and J. Clark, *The AI Index 2024 Annual Report*.
  - 35 Reuters, *China to rely on artificial intelligence in education reform bid*.
  - 36 C. M. Project, *AI joins China's primary schools*.
  - 37 O. Zawacki-Richter, V. Marn, M. Bond and F. Gouverneur, *Systematic review of research on artificial intelligence applications in higher education where are the educators?*
  - 38 N. Maslej, L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlaschi, J. Lotufo, A. Rome, A. Shi and S. Oak, *The AI Index 2025 Annual Report*.
  - 39 U.A.E., *Artificial Intelligence Office. UAE National Strategy for Artificial Intelligence*, <https://ai.gov.ae/strategy/>, Retrieved from.
  - 40 R. Roscoe, L. Allen and D. McNamara, *Contrasting Writing Practice Formats in a Writing Strategy Tutoring System*.
  - 41 H. Lee and J. Lee, *The effects of AI-guided individualized language learning: A meta-analysis*.
  - 42 A. Olney, S. DMello, N. Person, W. Cade, P. Hays, C. Dempsey, B. Lehman, B. Williams and A. Graesser, *Efficacy of a computer tutor that models expert human tutors*.
  - 43 J. Pane, E. Steiner, M. Baird, L. Hamilton and J. Pane, *Continued progress: Promising evidence on personalized learning*.
  - 44 C. Zhai, S. Wibowo and L. Li, *The effects of overreliance on AI dialogue systems on students' cognitive abilities: A systematic review*.
  - 45 Z. Krullaars, A. Januardani, L. Zhou and E. Jonkers, *Exploring Initial Interactions: High School Students and Generative AI Chatbots for Relationship Development*.
  - 46 B. Zimmerman, *Becoming a self-regulated learner: An overview*.
  - 47 O. Henkel, H. Horne-Robbinson, N. Kozhakhmetova and A. Lee, *Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI-Math Tutor in Ghana*.
  - 48 W. Cui, Z. Xue and K. Thai, *Performance comparison of an AI-based adaptive learning system in China*, conference paper.
  - 49 Z. Pardos and S. Bhandari, *Learning Gain Differences Between ChatGPT and Human Tutor Generated Algebra Hints*.
  - 50 R. Pelnek, *Adaptive Learning is Hard: Challenges, Nuances, and Tradeoffs in Modeling*.
  - 51 B. Bloom, *The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring*.