

## Looking beyond the hype: understanding the effects of AI on learning

Elisabeth Bauer, Samuel Greiff, Arthur C. Graesser, Katharina Scheiter, Michael Sailer

### Angaben zur Veröffentlichung / Publication details:

Bauer, Elisabeth, Samuel Greiff, Arthur C. Graesser, Katharina Scheiter, and Michael Sailer. 2025. "Looking beyond the hype: understanding the effects of AI on learning." *Educational Psychology Review* 37 (2): 45.  
<https://doi.org/10.1007/s10648-025-10020-8>.

### Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

**CC-BY 4.0: Creative Commons: Namensnennung**

Weitere Informationen finden Sie unter: / For more information see:

<https://creativecommons.org/licenses/by/4.0/deed.de>





# Looking Beyond the Hype: Understanding the Effects of AI on Learning

Elisabeth Bauer<sup>1</sup>  · Samuel Greiff<sup>2</sup> · Arthur C. Graesser<sup>3</sup> · Katharina Scheiter<sup>4</sup> · Michael Sailer<sup>1</sup>

Accepted: 5 April 2025  
© The Author(s) 2025

## Abstract

Artificial intelligence (AI) holds significant potential for enhancing student learning. This reflection critically examines the promises and limitations of AI for cognitive learning processes and outcomes, drawing on empirical evidence and theoretical insights from research on AI-enhanced education and digital learning technologies. We critically discuss current publication trends in research on AI-enhanced learning and rather than assuming inherent benefits, we emphasize the role of instructional implementation and the need for systematic investigations that build on insights from existing research on the role of technology in instructional effectiveness. Building on this foundation, we introduce the ISAR model, which differentiates four types of AI effects on learning compared to learning conditions without AI, namely inversion, substitution, augmentation, and redefinition. Specifically, AI can substitute existing instructional approaches while maintaining equivalent instructional functionality, augment instruction by providing additional cognitive learning support, or redefine tasks to foster deep learning processes. However, the implementation of AI must avoid potential inversion effects, such as over-reliance leading to reduced cognitive engagement. Additionally, successful AI integration depends on moderating factors, including students' AI literacy and educators' technological and pedagogical skills. Our discussion underscores the need for a systematic and evidence-based approach to AI in education, advocating for rigorous research and informed adoption to maximize its potential while mitigating possible risks.

**Keywords** Artificial intelligence · Educational technology · Learning · Effectiveness · Evidence-based design

---

Extended author information available on the last page of the article

Published online: 24 April 2025

 Springer

## Introduction

Discussions about the impact of artificial intelligence (AI) on education are marked by both overenthusiasm and deep skepticism, reflecting varied perspectives on how AI might affect educational practices as well as student learning processes and outcomes. In this reflection paper, we critically discuss current research trends from the perspectives of a diverse group of authors with research backgrounds in AI in education, educational technology, learning analytics, cognitive science, and learning science. We propose a framework for systematizing future directions in research on AI-enhanced learning and highlight important connections to previous literature on the effects of AI- and technology-enhanced learning. In our discussion, we focus primarily on cognitive learning processes and outcomes, while noting that motivation and emotions are inextricably linked to cognitive learning.

There are many different approaches falling under the umbrella of AI, including rule-based learning systems that utilize symbolic AI, statistical probabilistic algorithms derived from data mining with machine learning, and neural networks with deep learning techniques, which use advanced multi-layered architectures to effectively capture complex data patterns (D'Mello & Graesser, 2023; Zapata-Rivera & Arslan, 2024). These AI approaches are already an integral part of many learning environments, and some have been researched for decades, especially in the field of AI in education (AIED; du Boulay et al., 2023). In the context of education, technologies with and without AI can be divided into technologies *for* education (i.e., educational technologies; e.g., software designed for educational purposes) and non-educational technologies *in* education that are designed for broader contexts (e.g., the Internet). A prominent example of educational AI systems are intelligent tutoring systems (ITS), as these computer learning environments help students master knowledge and skills through intelligent algorithms that facilitate fine-grained adaptation to students and instantiate principles of effective learning (Graesser et al., 2016). By contrast, non-educational AI tools such as translation tools (e.g., DeepL), writing assistants (e.g., Grammarly), and non-educational conversational agents (e.g., ChatGPT) have been developed for broader purposes but are also applied in educational settings like language learning (Vogt & Flindt, 2023).

Recently, generative AI has gained prominence, particularly through large language models (LLMs) like generative pre-trained transformers (GPT). These technological advances allow for highly naturalistic interaction sequences and offer a range of educational use cases that were previously difficult or impossible to implement. The launch of ChatGPT made these advances widely accessible, marking a disruptive event in the discussion of the future of education. Generative AI has been and continues to be intensively discussed on social media by the broad public, with education being among the most frequently referred contexts, for example, regarding the usefulness of generative AI for various teaching and learning scenarios (Fütterer, Fischer et al., 2023). The high interest motivated significant investments in research funding, such as the Dutch National Education

Lab AI, which was funded with €36 million through the EU Recovery and Resilience Facility and was awarded €80 million from the Dutch National Growth Fund for its 2022-2032 infrastructure and capacity building, as well as €63 million for the development and commercialisation of educational prototypes. In the USA, four National Research and Development Centers on generative AI in the classroom were established, each receiving approximately \$10 million from the Institute of Education Sciences, and the National Science Foundation funded two new research centers in addition to the three previously funded centers on AI in education, each receiving approximately \$20 million over 5 years.

There has also been a dramatic increase in the number of research publications on AI-enhanced education. In this article, however, we argue that the current publication boom partly reflects a tendency to overhype recent developments, driven by a neglect of previous theoretical and empirical insights about instructional mechanisms and learning, the way arguments are framed, and the study methods used. These new research trends depart from the traditional AIED research, which has focused on systematically establishing insights on instructional effectiveness with rigorous research methods. So, looking beyond the hype, how can we effectively harness the potential of current AI advances to improve key educational outcomes, while acknowledging relevant limitations? And, building on previous research, what research directions and approaches might be prioritized to achieve this goal?

This paper argues that research on AI-enhanced learning should prioritize cognitive learning processes and outcomes to maximize AI's potential for effective learning. As this research evolves, it is crucial to integrate theoretical, empirical, and methodological insights from prior research on the learning effects of digital technologies, including AIED. We reflect on lessons from different types of technology comparisons, highlighting their strengths and limitations in advancing research on AI-enhanced learning. To guide future research, we propose a model that categorizes four types of AI effects on learning and highlight promising research directions informed by theoretical and empirical insights. Additionally, we outline key conditions for successful AI integration in education, including student and teacher prerequisites and contextual factors. While we focus on generative AI and LLMs due to their growing prominence, many of our arguments broadly apply to other AI-driven learning technologies, which is why we generally refer to "AI" as an overarching term.

## Cognitive Learning Outcomes in the Age of Generative AI

Learning effectiveness in today's educational landscape is driven by optimizing cognitive outcomes, specifically, the acquisition of knowledge and the development of skills. Knowledge refers to declarative information that students understand and remember, while skills denote procedural know-how developed through practice (Anderson et al., 2019). In the twenty-first century, different sets of knowledge and skills are crucial: domain-specific knowledge and skills ensure mastery in particular fields, while transversal skills (e.g., problem-solving and critical thinking), applicable across various domains, enable individuals to adapt to dynamic, cross-disciplinary

problems (Greiff et al., 2014). Many contexts require combining knowledge and skills. For example, in an inquiry task about the benefits of sunscreen, students need natural sciences knowledge (e.g., knowledge about ultraviolet light) combined with skills such as literature search (see Stadler et al., 2024). With advancements in AI, especially generative AI, an increasing range of tasks can be outsourced to AI systems (e.g., many well-defined data analysis tasks), challenging educational systems and their targeted objectives. However, developing human knowledge and skills remains crucial, especially for tasks requiring deep understanding, ethical considerations, and creative problem-solving, which cannot be entirely outsourced to AI (e.g., qualitative interpretation requiring contextual understanding and theoretical framing).

Specifically, despite ubiquitous and widely-accessible information, *domain-specific knowledge and skills* remain essential learning objectives as they empower individuals to reason about a broad set of problems, to apply theoretical knowledge to practical problems, and to establish important foundations for expertise development in specialized fields, such as engineering, medicine, and the sciences (Greiff et al., 2014). In addition, domain-specific knowledge is needed for understanding AI outputs in various contexts, ranging from everyday activities to professional action in specialized fields. Especially AI systems that analyze and generate language using algorithms trained on large datasets, such as LLMs, provide outputs that can be misleading, biased, or incorrect. These models inadvertently learn and replicate biases from the vast, biased datasets they are trained on (Lee et al., 2024). For example, generated materials and other outputs may reflect cultural or gender biases (e.g., Kotek et al., 2023; Tao et al., 2024), and learners' input language can lead to biased outputs, as LLMs may contain language and dialect prejudices (Hofmann et al., 2024). In addition, the probabilistic nature of LLMs implies they generate responses based on likelihood, not verified information, leading to plausible sounding but sometimes incorrect statements. This phenomenon is often referred to as hallucinations, which some researchers consider as a misleading metaphor as LLMs are not designed to represent the world accurately but rather to produce text without an actual concern for truth (Hicks et al., 2024; Perković et al., 2024). Instead, these models generate text by probabilistically predicting word sequences based on patterns in their training data, a process which has been compared to a "stochastic parrot" (Bender et al., 2021). If learners uncritically rely on information from AI systems, they risk adopting biased or incorrect information, an issue also observed in other contexts, such as interactions with Internet sources (Miller & Bartlett, 2012).

This is one of the reasons why, in addition to domain-specific knowledge and skills, *transversal skills* become increasingly important. While there is no definite list of transversal skills for a successful twenty-first-century learner, suggestions include critical thinking, problem-solving, information literacy, technology literacy, collaboration, communication, and creativity (Fiore et al., 2018; Van Laar et al., 2020). Additionally, AI literacy is becoming an increasingly important transversal skill for learners (Ng et al., 2021). It encompasses an understanding of AI's fundamental concepts, capabilities, implications, and ethical considerations, along with the skills required to interact with AI systems effectively and critically evaluate their outputs (Yan et al., 2024). However, transversal skills are also crucial for navigating other challenging situations of the dynamic and technology-driven twenty-first century (Spector & Ma, 2019; Van Laar

et al., 2020), involving handling vast information with mixed quality from the Internet (e.g., interacting with fake news), keeping pace with various technological and scientific advancements (e.g., biotechnological developments like gene editing), and addressing complex socio-scientific issues (e.g., climate change).

Effectively teaching knowledge and skills and preparing individuals for complex real-world problem-solving requires engaging them in relevant cognitive learning activities. Exceeding *shallow learning* that is focused on memorization, students should also be involved in *deep learning* processes where they synthesize, evaluate, and integrate new and existing knowledge (Chi & Wylie, 2014; Graesser, 2015). In their ICAP model (which stands for interactive, constructive, active, and passive learning), Chi and Wylie (2014) propose four types of learning activities, ranging from shallow to deep learning: Shallow learning comprises activities that afford *passive* engagement, in which information is received without active processing, and activities that afford *active* engagement, which involves applying existing knowledge to materials that promote retention but not new insights. Deep learning comprises activities that afford *constructive* engagement, which involves generating ideas and outputs beyond the learned material, enhancing problem-solving and other transversal skills, and activities that afford *interactive* engagement, which entails collaborative idea generation, leading to novel inferences while fostering communication and collaboration skills. To effectively enhance the learning of both knowledge and skills, especially transversal skills crucial for the twenty-first century, educators need to initiate deep learning processes in their students. This can be facilitated by systematic research on how to harness the potential of new AI capabilities to effectively initiate and support deep learning processes, including identifying the limitations of different types of AI enhancements and their potential negative effects on cognitive learning processes and outcomes.

## Research on AI-Enhanced Learning

### Recent Publication Trends in AI-Enhanced Learning

While a balanced view on the potential of AI, particularly LLMs and generative AI, for learning is necessary, current publications often focus primarily on the potential advantages of these technologies for learning. In the following, our objective is to reflect on the current trends in research on AI-enhanced learning by highlighting both the strengths and limitations of different publication types and study approaches. Doing so, we identify highly promising directions, encourage methodological reflections, and inspire future research, ensuring that both positive and negative impacts are duly considered.

The publication boom following the disruptive event of ChatGPT's launch in November 2022 includes numerous discussion papers on the general functionalities, opportunities, and challenges of increasingly powerful AI systems (e.g., Abd-Alrazaq et al., 2023; Alasadi & Baiz, 2023; Grassini, 2023; Kasneci et al., 2023; Rasul et al., 2023; Yan et al., 2024). As a starting point, these papers offer valuable insights, especially to newcomers in the field, but research must eventually shift

toward focusing on evidence-based studies. Indeed, empirical research on LLMs and generative AI in education is beginning to gain momentum. Some initial studies focus on providing insights into the performance of this new generation of algorithms (e.g., Du et al., 2024; Meyer & Dannecker, 2024). Yet, publications with a focus on algorithm performance face the problem of quickly becoming outdated due to the fast pace of AI developments and the duration of typical peer-review processes. Empirical studies evaluating the instructional benefits of LLM-based interventions may provide more lasting educational value, at least if they increase our understanding of how these technological advances can and cannot enhance learning processes and outcomes (e.g., Fan et al., 2024; Stadler et al., 2024). Despite many examples of rigorous research, there are also many studies that fail to adequately acknowledge their limitations (a problem we will elaborate on below), thereby contributing to the hype about AI's effects on learning. A growing number of reviews and meta-analyses have begun to focus on synthesizing the primary studies on LLMs and generative AI applications (e.g., Deng et al., 2025; Wu & Yu, 2024). However, syntheses that exclusively focus on LLMs and generative AI usage are currently still rather constrained by the limited number of available studies, which may be the reason why some syntheses and meta-analyses with this specific focus apply rather lenient inclusion criteria. This dynamic can easily lead to a "garbage in, garbage out" problem unless strict methodological and conceptual inclusion criteria are applied during study selection (Borenstein et al., 2021).

Indeed, many recent primary studies and research syntheses have faced criticism for methodological issues that compromise the interpretability and validity of their findings. Unlike traditional AIED research, which has systematically established insights on instructional effectiveness through rigorous methodologies, these newer research trends often seem to prioritize rapid exploration over methodological robustness. While continued research on AI-enhanced learning is essential, it must acknowledge limitations arising from specific study characteristics in order to avoid overgeneralizations and instead refine our understanding of AI's role in enhancing learning processes and outcomes. Some critical study characteristics include the measurement approaches and study designs used in primary studies, as well as the inclusion criteria employed in research syntheses.

Concerning the target variables and measurements, studies often focus on subjective variables like satisfaction or self-assessments of learning to make claims about the usefulness of the new AI-approaches, neglecting actual learning processes and outcomes. Furthermore, there are studies that use AI-enhanced performance as an indicator of learning effects, concluding that performance improvements *during* an AI-supported task (e.g., a writing task or programming task) suggest learning benefits compared to an unaided comparison condition. However, although performance during a supported intervention phase *can* suggest initial learning benefits, it cannot be considered as sufficient evidence for actual learning effects. This would require demonstrating subsequent performance or knowledge improvements that persist without AI support. Recognizing this distinction and understanding the potential limitations of performance measurements when interpreting such findings are crucial. Similarly, research syntheses must critically assess the nature and quality of

the primary studies' measurements as part of their selection and appraisal process to prevent conflating learning with mere performance enhancements.

Additionally, the study designs employed significantly influence the interpretations that can be drawn from the research. For example, pre-post designs without a comparison condition can indicate whether there has been a positive or negative change during the AI-enhanced intervention period. However, such designs do not clarify how the AI-enhanced instruction compares to other forms of instruction (with or without AI or other technologies), nor do they confirm whether changes are due to the investigated intervention or other factors (e.g., additional instruction or participant fatigue). Similarly, "no-intervention" control group designs (e.g., AI feedback versus no feedback) help determine whether an intervention has any effect. However, when the control group receives no instruction or additional support, the findings reveal only that the AI intervention is better than doing nothing, offering limited insight into its instructional quality. This approach can be useful when the effectiveness of a specific instructional approach is still debated. However, if an intervention's effectiveness is already well established, such a comparison becomes less informative. In this case, a more relevant question becomes how the AI intervention performs relative to other well-established methods, such as teacher-led instruction. Even if AI-enhanced instruction shows positive effects compared to no instruction, it may still underperform in comparison to alternatives like non-AI technology-enhanced instruction, teacher-led instruction, or peer instruction. Additionally, study designs that compare different interventions with and without AI or other technologies can lead to misinterpretations if the purpose of the comparison is unclear. It is essential to distinguish whether the goal is to evaluate if an instructional approach can be implemented using AI (technology comparison against no technology or a different technology) or whether the aim is to compare functionally different types of instructional methods (instructional comparison).

The specific comparisons made in the primary studies are also important to consider when integrating findings in research syntheses. Without careful differentiation, meta-analyses risk an "apples and oranges" issue, where studies with varying interventions and comparison conditions are mixed together to make broad claims about AI effectiveness. This can blur important distinctions between the effects of integrating AI and the effects of different types of instruction, potentially resulting in misleading generalizations.

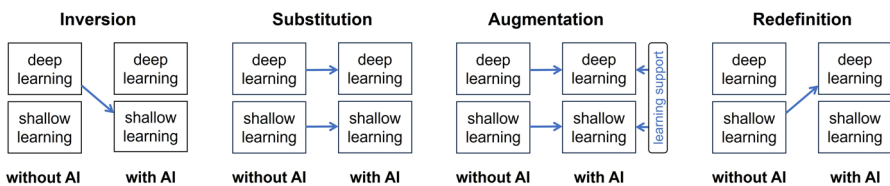
Obviously, not only the aforementioned study designs, but others as well, have inherent limitations that must be considered and weighed in terms of their pros and cons for the specific research questions addressed. Also, certain research contexts impose additional restrictions on which designs are suitable and ethically justifiable. For example, using a "no-instruction" control in educational settings is controversial unless adequate compensatory instruction can be provided. Nonetheless, it is crucial to be explicit about what conclusions can and cannot be drawn from each study to prevent overhyping the capabilities of the technology itself, considering that its effectiveness is contingent upon the quality of the instruction provided.

## A Model for Conceptualizing AI Enhancement

In many regards, the current research trends and the accompanying debate about AI in education echo earlier discussions about the impact of technological advancements on education more generally. The prime example is the media debate about whether media inherently influence educational outcomes. This debate featured arguments that media can significantly shape the learning process due to their unique capabilities (Kozma, 1991, 1994), contrasted with the view that educational outcomes are influenced not by the media themselves, but by the instruction they deliver (Clark, 1983, 1994a, 1994b). Similarly, Salomon (1979) emphasized that media can function as a delivery tool for content (learning from media) or as a cognitive tool that enhances learners' engagement and thinking (learning with media). While these perspectives differ, together they underscore the dual role of technology in education: instrumental, emphasizing that any positive learning effects should be attributed to the instruction rather than the technology itself; and transformative, highlighting the potential to afford known and new possibilities for enhanced instruction.

When speaking of AI enhancement of cognitive learning processes and outcomes, it is important to be aware that enhancement can be defined in different ways. For example, enhancement might be defined as substituting specific instructional actions previously performed by a teacher (substitution), as augmenting instruction with additional cognitive learning support (augmentation), or as redefining instructional tasks to offer students more options to engage in deep learning processes (redefinition; Puentedura, 2006, 2014; Sailer et al., 2024a). Depending on researchers' definition of enhancement in their respective research context, their understanding might influence the choice of comparisons they make. To offer an explicit terminology for the AI enhancement debate, classify possible implementations, and define different research directions, we suggest distinguishing four types of effects in the context of AI-enhanced learning: inversion effects, substitution effects, augmentation effects, and redefinition effects of AI in educational contexts, summarized in the ISAR model shown in Fig. 1.

The ISAR model builds on the SAMR model (Puentedura, 2006, 2014), which categorizes the extent to which digital technologies transform learning tasks, and the ICAP-inspired SAMR model (Sailer et al., 2024a), which additionally integrates the idea of cognitive processing depth from the ICAP model (Chi & Wylie, 2014). The ISAR model retains substitution, augmentation, and redefinition as key mechanisms,



**Fig. 1** The ISAR model of inversion, substitution, augmentation, and redefinition effects of AI in education

as these were empirically supported by a second-order meta-analysis on technology-enhanced learning (Sailer et al., 2024a). The ISAR model refines these mechanisms for AI-enhanced learning and continues to emphasize the importance of comparison conditions to assess when and where AI provides instructional benefits over non-AI-enhanced instruction (as illustrated later through different comparisons from meta-analyses on ITS). Beyond these three empirically supported mechanisms, the ISAR model introduces inversion effects, referring to reduced cognitive learning when learners over-rely on AI as demonstrated by initial research on generative AI.

*Inversion effects* in AI-enhanced learning occur when AI, intended to support deep learning, instead leads to reduced cognitive processing and learning outcomes, counteracting its intended benefits. This has been observed in studies on students using ChatGPT for constructive learning tasks (e.g., information search and writing), where generative AI use is linked to shallower processing and diminished learning outcomes (Fan et al., 2024; Stadler et al., 2024).

*Substitution effects* in AI-enhanced learning occur when AI provides instructional equivalence to non-AI alternatives without changing learners' cognitive processing depth and, therefore, without directly changing learning outcomes. However, substitution might improve efficiency and resource allocation by replacing specific aspects of human instruction. For example, meta-analyses comparing ITS with human tutoring or small-group instruction found no significant differences in cognitive learning effects (Ma et al., 2014; Steenbergen-Hu & Cooper, 2014; VanLehn, 2011), suggesting that while ITS do not surpass human tutoring, they offer comparable conditions for achieving cognitive outcomes.

*Augmentation effects* occur when AI enhances instruction by providing additional cognitive learning support compared to a non-AI alternative. Like substitution, augmentation does not alter the task itself, maintaining similar processing depth to the comparison condition. For example, meta-analyses found medium to large learning benefits for ITS, which provide feedback and targeted hints, compared to self-reliant learning without support, and small to medium gains when comparing ITS to non-adaptive or less adaptive computer-assisted learning (Ma et al., 2014; Steenbergen-Hu & Cooper, 2014). Thus, augmentation effects include AI-driven instructional support, enhancing learning beyond unsupported learning or lower-quality non-AI support.

*Redefinition effects* occur when AI transforms learning tasks to foster deeper (constructive or interactive) learning, provided that the non-AI comparison condition does not already support deep learning processes. Meta-analyses comparing ITS to active and passive instructional methods (e.g., teacher-centered instruction, text reading) found small to medium benefits for constructive learning with ITS (Ma et al., 2014; Steenbergen-Hu & Cooper, 2014). Thus, AI enables redefinition when it engages students in constructive or interactive learning processes, compared to conditions that involve passive and active learning processes with less emphasis on skill development and knowledge construction.

The assumption underlying the substitution, augmentation, and redefinition effects in the ISAR model is that the greater the instructional enhancement relative to a comparison condition, the greater the transformative potential for enhanced cognitive learning processes and outcomes through the affordances offered by the

AI system. However, implementations need to consider how to avoid undesired inversion effects. In the following, we provide examples of inversion effects and furthermore explore how AI affordances can enhance learning through substitution, augmentation, and redefinition while minimizing the risk of inversion. Doing so, we focus our considerations on generative AI and LLMs due to the high interest in these technologies.

## Effects of AI-Integration for Cognitive Learning Enhancement

### Inversion Effects of AI Undermining Deep Learning

A major concern associated with an inadequate use or implementation of AI in learning contexts is its potential to undermine the acquisition of knowledge and skills (Huber et al., 2024; Kasneci et al., 2023). This phenomenon is already known from workplace learning, specifically from contexts where AI-driven automation increasingly complements or replaces human actions (Rafner et al., 2021). Whether such effects are problematic depends on the specific conditions and underlying goals. Shifting toward system monitoring may be desirable for enhancing performance and efficiency if AI outperforms humans in a given task. In this context, hybrid intelligence aims to create synergies between human and AI capabilities to optimize problem-solving and task performance (Akata et al., 2020). However, replacing human actions with AI can be problematic if it leads to the loss of human skills critical for achieving high-quality outcomes, diminishing expertise previously acquired and maintained through regular practice. Over-reliance on algorithms may weaken decision-making and judgment skills, resulting in poor choices when AI support is unavailable (Sutton et al., 2018). Especially when AI minimizes human input, professionals may accept its outputs uncritically (Hoff, 2011). This concern is particularly relevant in highly specialized fields like finance and medicine, where professionals bear significant responsibility and their decisions can profoundly impact others (Levy et al., 2019; Mascha & Smedley, 2007).

Similarly, in educational contexts, student-initiated over-reliance on AI tools can undermine the development of skills such as critical thinking (Zhai et al., 2024). For example, when students use ChatGPT to generate complete assignments, rather than for assistance such as suggesting revisions for their own critical review, they miss essential learning opportunities. Another issue may be that students are not necessarily able to ask high-quality questions and seek help (Aleven et al., 2003; Graesser & Person, 1994), which is problematic if this is a key requirement for successful system interaction. Additionally, suboptimal design or integration of AI-tools by educators or instructional designers can limit student engagement in key learning activities, thereby hindering skill development. A study by Stadler et al., (2024) compared students' use of ChatGPT to traditional search engines during engagement in a scientific inquiry task on the socio-scientific issue of nanoparticles in sunscreen. The findings indicated that ChatGPT simplified task processing by reducing task-irrelevant extraneous load and load that is intrinsic to the learning task (see Sweller 2010); however, this did not improve learning effectiveness, but was accompanied by reduced cognitive processing

depth (germane load; see Sweller 2010) and reduced cognitive learning outcomes, as indicated by students' quality of arguments presented in a posttest justification task. Similarly, Fan et al. (2024) compared the effects of different support options, including support through ChatGPT, a chat with a human expert, a set of writing analytics tools, and no support, on students' revision processes during a writing task. An analysis of students' self-regulated learning behavior indicated that all support options increased students' engagement in elaboration, organization, and orientation processes during their revisions. However, while the ChatGPT group, compared to the other groups, showed improved task performance during the supported intervention phase, there were no differences in the posttest knowledge gain or knowledge transfer. A temporal process analysis of learners' metacognitive activities suggested that the ChatGPT group relied strongly on the AI support and showed relatively low metacognitive processing compared to the other support groups. The authors concluded that ChatGPT might promote "metacognitive laziness" where students refrain from engaging deeply in self-regulated learning processes.

How can inversion effects be mitigated? In the context of education, effective prevention strategies are yet to be explored. In the context of workplace learning, it has been suggested that the risks of over-reliance on potentially suboptimal outputs can be mitigated if professionals are actively engaged in reflective processes before or during the generation of AI outputs. One example is cognitive forcing functions that initiate human inputs, such as initial judgments, before AI outputs are generated (Buçinca et al., 2021). Additionally, providing well understandable explanations with AI outputs can help users to critically assess information (Vasconcelos et al., 2023). Such hybrid intelligence approaches were suggested to support continuous learning and upskilling, ensuring humans maintain and develop skills through human-AI collaboration (Järvelä et al., 2023; Rafner et al., 2021).

Considering the overall evidence, suboptimal use of AI tools during learning can lead to shallow learning processes despite instruction aimed at deep learning processes, resulting in an inversion of the intended goal of enhancing learning. While the studies of Stadler et al. (2024) and Fan et al. (2024) investigate a non-educational generative AI system in educational settings, similar outcomes could potentially occur with educational AI systems that either relieve students of essential learning processes or allow them to outsource critical learning processes. Accordingly, the design of AI-enhanced learning opportunities must prioritize promoting cognitive engagement and the development of essential domain-specific and transversal skills. In addition, research needs to continue advancing our understanding of inversion effects, especially how to avoid pitfalls in the context of learning with (educational or non-educational) generative AI systems.

### **Substitution Effects in AI-Enhanced Learning**

The assumption that leveraging AI for substitution of instructionally equivalent non-AI learning conditions does not inherently lead to greater learning gains aligns with long-standing debates in educational research, specifically the arguments from the media debate that simply introducing a new instructional medium does not

improve learning outcomes if the instructional method remains unchanged (Clark, 1983, 1994a, 1994b). This argument is well empirically supported, for example by a meta-meta-analysis that compared technology-enhanced learning with non-technology conditions in higher education (Sailer et al., 2024a). Similarly, we can assume that when AI replaces specific instructional functions without altering the cognitive processes involved, the effectiveness of the AI-enhanced instruction remains comparable to non-AI instructional methods used for the same purpose in the same context. However, substitution may indirectly enhance cognitive outcomes if AI increases efficiency and optimizes learners' resource allocation. In such cases, learners may redirect their spared resources toward more intensive practice or additional learning activities, though the extent of such benefits likely depends on individual factors such as motivation and self-regulation.

One example of potential AI-enhanced substitution is the use of AI-generated instructional videos and podcasts. Research suggests that AI-generated videos lead to cognitive learning outcomes that are comparable to using teacher recordings and teacher-generated video instruction (Leiker et al., 2023; Netland et al., 2025; Xu et al., 2024), making AI-generated educational videos a cost-efficient alternative. However, challenges remain, including lower perceived social presence when comparing videos of AI-generated agents to filmed human instructors (Netland et al., 2025; Xu et al., 2024). Besides videos and podcasts for content presentation, AI-generated quiz questions and flashcards can substitute teacher-created quizzes or those from textbooks while maintaining instructional equivalence (Almadhoob et al., 2024; Bachiri et al., 2023; Hutt & Hieb, 2024; May et al., 2025). However, AI offers the advantage of enabling the generation of quiz questions and flashcards on demand, facilitating self-assessment and repeated practice. Yet, ensuring that AI-generated questions align with the content of the learning materials and maintain high-quality formulations can pose a challenge (Hutt & Hieb, 2024; May et al., 2025). Intelligent textbooks like iTELL create interactive reading and writing tasks, offering another form of AI-based instruction (Crossley et al., 2024). Additionally, AI-enhanced question-answering through conversational agents can provide instant responses to students' clarification questions in- and outside of class that would otherwise be answered by a teacher during class (Almadhoob et al., 2024; Hicke et al., 2023; Nazar et al., 2024). This offers advantages in the accessibility and immediacy of answers while alleviating teacher workload, especially in educational contexts with many students (e.g., lectures in higher education).

When using generative AI to enhance content representation and practice tasks, common challenges include output accuracy, content alignment, and pedagogical quality, which are discussed as crucial across all applications (Hicke et al., 2023; Hutt & Hieb, 2024; Leiker et al., 2023; May et al., 2025; Nazar et al., 2024; Netland et al., 2025). Such issues might be mitigated by focusing on educational AI tools with built-in quality assurance mechanisms that ensure, for example, that AI-generated materials adhere to the course content. For instance, Jill Watson, a virtual teaching assistant, ensures content alignment by restricting AI-generated responses to instructor-approved course materials through retrieval-augmented generation (Kakar et al., 2024). Also, teacher oversight can be needed, keeping the human in the loop, to actively monitor AI-generated content and ensure that AI outputs remain

accurate, relevant, and pedagogically sound in educational settings. Generally, a frequent concern in the context of using AI to achieve substitution effects is “dehumanize” learning (e.g., Ghosh, 2024), highlighting that AI-enhanced instruction should not fully replace social interactions, which are fundamental to learning and development. This is underlined by findings, such as reduced social presence perceived when watching AI-generated videos (Xu et al., 2024) and meta-analytic findings suggesting that ITS are particularly effective when they are used to supplement human instruction (Sun et al., 2021). Similar to Holmes and Miao (2023), we recommend that AI should complement, rather than replace, human interaction in educational contexts.

While the examples discussed in this section focus on content representation and practice opportunities, AI could be used to partially substitute for cognitive learning support (e.g., feedback) or the initiation of deep learning activities that would otherwise be performed to the same extent by a teacher or tutor. However, because key affordances of AI in education lie in implementing these instructional approaches more intensively than is currently done in many educational contexts, the following sections discuss these approaches from the perspectives of augmentation (AI-enhanced cognitive support) and redefinition (AI-enhanced learning activities for deep learning).

### **Augmentation Effects Through AI-Enhanced Cognitive Support**

One of the most commonly discussed affordances of AI in education is the opportunity to augment or automate the cognitive support provided to the learners. Cognitive support refers to instructional strategies and tools that help learners in processing learning activities more effectively, enhancing their understanding, retention, and application of knowledge. Cognitive support is a key mechanism in various educational contexts but also a broad umbrella term grounded in various educational paradigms (see NASEM, 2018). Traditionally, cognitive support is provided by teachers, meaning that AI-enhanced cognitive support qualifies as substitution if it replicates the support typically offered without AI. However, when AI expands the variety or frequency of support beyond what is traditionally available, this can be characterized as augmentation. Specifically, AI technologies can enhance cognitive support by providing additional feedback and scaffolding that optimize cognitive processing of various learning activities. Additionally, in digital learning environments, AI can enhance the quality of feedback and scaffolding compared to non-adaptive or less adaptive technology-driven cognitive support.

Feedback is a cognitive support measure that informs learners about their performance in relation to learning goals and highlights ways for improvement (Hattie & Timperley, 2007). Effective feedback significantly impacts learning outcomes by offering insights into task performance, facilitating self-assessment, and guiding future efforts. AI systems can augment teacher feedback by providing real-time data analytics and visualizations of student performance, exemplified by teacher dashboards that facilitate monitoring and assessment through visualizing relevant learner variables (Knoop-van Campen et al., 2023; Xhakaj et al., 2017). Likewise, learners

can benefit from visual feedback tools, such as student-facing dashboards, which facilitate students' self-assessment by providing real-time overviews of individual or collaborative learning activities and outcomes (Breideband et al., 2023; Jivet et al., 2017; Long & Alevan, 2017). However, the effectiveness of both teacher- and student-facing dashboards depends on the dashboard usability and audience characteristics, as designing dashboards that provide accessible, relevant, and actionable information can be challenging, and users may struggle to translate insights into meaningful actions because of their knowledge and skills (Jivet et al., 2017; Matcha et al., 2020). In contrast to visual feedback, instructional feedback provides students with verbal information about their performance, ranging from simple feedback on task performance to elaborate feedback that presents a formative assessment with suggestions for improvement (Narciss et al., 2014). For cognitive learning outcomes, elaborate instructional feedback providing detailed guidance on task processing and self-regulation was found to be particularly helpful (Wisniewski et al., 2020). Further, computer-generated animated agents can use non-verbal facial cues, paralinguistic cues (e.g., intonation in speech), and gestures as an additional form of feedback (Johnson & Lester, 2016). To keep the human in the loop and respond to concerns about dehumanizing learning through replacing human interaction, AI-based feedback can be complemented by AI-enhanced peer feedback as a scalable alternative to AI-enhanced teacher feedback; however, even with AI support, the quality of peer feedback processes and outcomes may still vary (Banihashem et al., 2024; Bauer et al., 2023).

Scaffolding is a second main type of cognitive support and facilitates learners' processing of a learning task within their zone of proximal development, which is the zone of task difficulty where learners need guidance to succeed (Wood et al., 1976). Scaffolding approaches include providing additional structures to facilitate task processing as well as adjusting task difficulty and task sequencing to enable learners to achieve performance levels that would be out of reach without the support. Scaffolding that structures a learning task can take various forms, such as modeling and worked examples (Van Gog & Rummel, 2010), prompts and hints (D'Mello & Graesser, 2023), scripts and roles (Fischer et al., 2013), and reflection phases (Mamede & Schmidt, 2017). Additionally, scaffolding can adjust task difficulty and task sequencing, thereby creating personalized learning paths, which is a technique that is often employed by ITS systems but also considered beneficial in other contexts, such as simulation-based learning (Fischer et al., 2022; Holstein et al., 2018). Scaffolding can target various aspects of cognitive learning processes, such as the activation of domain-knowledge (Sommerhoff et al., 2023) and the application of skills, such as collaboration (Vogel et al., 2017) and self-regulation (Azevedo et al., 2022). Meta-analyses show that scaffolding enhances cognitive outcomes in digital learning environments (Belland et al., 2017; Chernikova et al., 2020). Effective scaffolding adapts to learners' needs and fades out as they become more competent to promote autonomy and self-regulated learning (Pea, 2004). Moreover, different scaffolds are most effective for different learners; for example, worked examples may not be as beneficial to high-knowledge learners as reflection phases (Chernikova et al., 2020).

AI technologies offer diverse approaches to implementing adaptive cognitive support. Rule-based systems deliver feedback and scaffolding based on predefined rules but lack flexibility, while machine learning and deep learning systems offer personalized feedback by analyzing complex data patterns (Sailer et al., 2024b; Zapata-Rivera & Arslan, 2024). Log data analysis identifies student behavior patterns, enabling timely interventions (Lim et al., 2023). Natural language processing techniques analyze written responses to provide personalized cognitive support for improving skills, such as argumentation (Butterfuss et al., 2022; Zhu et al., 2020). Additionally, natural language processing techniques, combined with speech synthesis, can analyze and generate spoken language (e.g., in computer-generated animated agents; Fink et al., 2024), though this process may still experience time lags compared to text-based communication (Dekel et al., 2024). Analytic AI systems evaluate performance through data-driven insights, identifying patterns and discrepancies to provide consistent predefined feedback and scaffolds (e.g., Bauer et al., 2025; D’Mello et al., 2024). Generative AI creates personalized feedback and scaffolds based on interactions and performance data, enabling dynamic support systems; however, these systems may lack explainability and, if not designed in agreement with instructional principles, may also generate content that varies in accuracy, specificity, and pedagogical quality (Banihashem et al., 2024). While adequate prompting determines the immediate output quality, fine-tuning with domain-specific data can further enhance alignment with domain knowledge, instructional accuracy, and educational needs.

The validity and accuracy of AI-enhanced cognitive support are crucial for its effectiveness, as low-quality cognitive support can lead to learner disengagement, ultimately hindering the intended augmentation. To ensure effective implementation, AI-driven cognitive support must align with established learning theories and undergo rigorous validation procedures. Advancing our understanding of how to optimize AI for augmentation requires research that builds on prior insights into cognitive support mechanisms and systematically compares different variations of AI-enhanced support to identify the most effective approaches while mitigating potential limitations.

### **Redefinition Effects Through AI-Enhanced Learning Activities for Deep Learning**

The most transformative potential of AI in education may lie in redefinition effects, where AI is used to redesign learning tasks in ways that encourage students to engage in deep (constructive or interactive) rather than shallow (passive or active) learning processes. Although research has begun to explore ways to redefine education with (generative) AI, new approaches are likely to emerge in the future, requiring ongoing discussion and investigation.

A straightforward option for generative AI to enhance deep learning activities in education is by assisting teachers with planning instructional approaches that afford deep learning processes. Instructional materials such as interactive simulations can be generated with the help of AI (Bewersdorff et al., 2025) and allow students’ knowledge construction through the exploration of complex concepts (Kali &

Linn, 2008). Further, generative AI can support the preparation and implementation of instructional approaches, such as problem-based learning, scenario-based learning, and game-based learning, enhancing cognitive engagement (Huber et al., 2024; Kasneci et al., 2023). Teachers can use a write-curate-verify approach for generating materials, by writing the prompts, curating the output, and verifying the output, which keeps the human in the loop to ensure high-quality results (Bai et al., 2024; Ninaus & Sailer, 2022). However, like students, teachers may also be at risk of over-relying on AI outputs and adopting them uncritically, particularly if they have limited awareness of AI's potential pitfalls. Therefore, teacher characteristics, such as AI literacy, play a crucial role in ensuring effective teacher-AI collaboration. Specialized interfaces that enhance output explainability (e.g., through annotations) and promote teachers' critical reflection and revision of AI-generated content might help mitigate these issues.

Furthermore, students can benefit from constructive learning activities that involve interacting with generative AI, when these activities promote deep learning but would be impractical or too difficult to implement without AI (e.g., a simulated conversation with a historic character). Constructive learning activities, where students actively generate formative outputs, foster deep understanding and skill development and can, in some cases, be enhanced by generative AI. For example, learning-by-design engages students in iterative cycles of designing, building, testing, and reflecting to deepen their learning (Kolodner et al., 2003a). This approach fosters domain-specific knowledge and transversal skills, such as problem-solving, critical thinking, and collaboration (Kolodner et al., 2003a; Kolodner et al., 2003b). Meta-analyses show medium- to large-sized positive effects on student achievement in K-12 STEM education (Delen & Sen, 2023). Design-based learning was also used in other learning contexts, such as engineering (Arastoopour et al., 2016), computer science and programming (Jun et al., 2017), and to train technology skills in teacher education (Yeh et al., 2021). Generative AI can assist learners in creating text elements, visuals, or code, for example, to design games and other creative outputs (Huber et al., 2024). While doing so, learners can practice relevant transversal skills, such as their technology skills, problem-solving, and creativity (e.g., through iterative prompting), and deepen their understanding of relevant domain-specific knowledge. Using such constructive tasks may increase learner motivation, which could potentially help mitigate inversion effects. Additionally, flipped classroom approaches could be useful because the constructive activities take place during class, which facilitates teacher interventions that guide students' use of AI, while passive knowledge transfer takes place outside of class time (Akçayır & Akçayır, 2018; Strelan et al., 2020).

Interactive learning activities involving communication enable deep learning and skill development through collaborative knowledge construction. Especially in the absence of suitable human interaction partners, interactive learning activities can be facilitated through generative AI that uses chat-based interfaces (e.g., ChatGPT). However, while affording naturalistic interactivity, these AI systems, especially in the case of non-educational applications used in education (e.g., ChatGPT), can face challenges in aligning with other learning principles (see NASEM, 2018). Specifically, they show high immediate adaptivity to human input but, without additional system components, have limited "memory" of learners (e.g., about learner characteristics and previous learning progress), which hinders conversational coherence across

multiple interactions. Additionally, high-quality cognitive support, such as feedback, depends on detailed information about the learning content and suitable pedagogical approaches that may not be available without targeted system design. Insights for dealing with these issues can be derived from research on ITS with conversational agents. This research showed that technology-driven natural language interactions can foster deep conceptual learning by requiring explanation and reflection, and help develop essential communication skills (Rus et al., 2023). However, in addition to a (chat-based) user interface, further key components include a learner model tracking students' knowledge and misconceptions, a domain model representing relevant subject information, and a tutoring model determining instructional strategies (D'Mello & Graesser, 2023). A common form of conversational ITS involves dialogue-based interactions with a simulated tutor (e.g., AutoTutor; Graesser, 2016). Alternatively, in a triologue, the student can be the tutor, explaining content to a simulated tutee with optional help from a simulated teacher (Graesser et al., 2017). Recent ITS developments increasingly use LLMs for enhanced interactivity. For example, the Socratic Playground for Learning (Zhang et al., 2024) uses GPT-4 with assisted prompt engineering for multi-turn dialogues. Although it lacks a traditional domain model and learner model, limiting content quality assurance and tracking of learners' task mastery, the Socratic Playground for Learning uses a Socratic tutoring model to foster critical thinking and reflection by iteratively posing questions and guiding learners. Similarly, the Ruffle & Riley system (Schmucker et al., 2024), a triologue-based tutoring platform, employs AutoTutor's Expectation Misconception Tailoring as a tutoring model to structure dialogues and address student misconceptions. Instead of using a traditional domain model, it generates tutoring workflows from existing content and relies on chat logs and real-time responses rather than building a comprehensive learner model. These examples illustrate how LLMs can be integrated with ITS components and pedagogical principles, but also highlight challenges in balancing flexibility, content quality, pedagogical quality, and personalized instruction. In this context, research must focus on developing components like tutoring models in ways that prevent inversion effects in interactions with conversational agents. This includes optimizing pedagogical behaviors, such as how the agent formulates and asks questions, to enhance learning effectiveness. Nevertheless, natural language-based learning systems represent a powerful approach to realizing potential redefinition effects through AI-enhanced learning activities for deep learning.

## Conditions for Successful AI Integration in Education

To effectively integrate AI in education, several factors must be considered, including the prerequisites of students and teachers (e.g., knowledge, skills, beliefs, motivation) and contextual factors (e.g., infrastructure, regulations). These factors are comparable to those influencing the integration of other digital technologies in education (Lachner et al., 2024; Sailer et al., 2021), which we will summarize below, but may partially be further specified for the context of AI.

Prerequisites for both students and teachers include their knowledge and skills specific to the learning content and transversal skills such as critical thinking and problem-solving (Greiff et al., 2014). These skills may influence the quality of interactions with AI systems, such as effective prompt-writing and assessing AI outputs. Especially knowledge and skills related to digital technologies, including AI literacy, are crucial for raising awareness about problems such as biases in training data and outputs (Ng et al., 2021). For teachers, technological pedagogical content knowledge and technology-related teaching skills are essential for effectively integrating AI tools into diverse learning scenarios and teaching situations (Lachner et al., 2024; Mishra et al., 2023). Additionally, students' and teachers' motivation and beliefs may significantly impact effective AI-enhanced learning. As discussed earlier, over-reliance on AI due to high trust can decrease critical reflection of AI outputs (Buçinca et al., 2021). Conversely, low trust and technology-acceptance (including perceived usefulness and ease-of-use) may prevent students and teachers from benefiting from the opportunities of AI enhancement (Fütterer, Scherer et al., 2023; Viberg et al., 2023). Teachers with high self-efficacy and positive attitudes toward technology may be more likely to experiment with AI technologies, increasing the likelihood of incorporation into their teaching practices (Backfisch et al., 2021; Scherer et al., 2019).

Contextual factors such as opportunities for teacher professional development, institutional infrastructure, access to technology, and regulations may be critical for implementing AI-enhanced learning (Sailer et al., 2021). Ongoing professional development may ensure that teachers remain updated with the latest AI advancements and are equipped to integrate these tools effectively into their teaching practices (Lindner et al., 2019; Williams et al., 2021). Individual access to technology is essential for both teachers and students, ensuring they have the necessary devices and resources to engage with AI tools (Crompton, 2017). This access is vital to avoid amplifying the digital divide within and across countries (UNESCO, 2023). Inclusive education can be facilitated through institutional access to technology, which is, however, also not a given (Liu et al., 2024). Institutional infrastructure, including reliable Internet access, sufficient hardware, and technical support, is crucial for seamless technology integration in educational settings (Liu et al., 2020; Sailer et al., 2021). Political and institutional regulations governing the ethical use of AI, data privacy, and security are necessary to protect all parties and may provide a structured approach to integrating AI in education (Liu et al., 2020). Lastly, user-friendly, ethical, and regulation-compliant AI applications may facilitate the effective use of AI in education.

In conclusion, understanding the opportunities and limitations of AI-enhanced learning in any educational context requires a holistic perspective on the various conditions necessary for effective and sustainable AI integration in education.

## Conclusion

This reflection paper has explored the potential of AI to transform instruction, focusing on cognitive learning effects. We have reflected on current trends in research publications on AI-enhanced learning and proposed the ISAR model to distinguish

inversion, substitution, augmentation, and redefinition effects of AI enhancement in the context of cognitive learning processes and outcomes. This distinction can guide productive research on AI-enhanced learning, avoiding overgeneralization by explicitly addressing the nature of the targeted effect. The ISAR model can also guide the design of AI-enhanced learning approaches, as illustrated in the examples provided. However, successful AI integration in education additionally requires that students and teachers have the necessary knowledge, skills, and motivation to interact effectively with AI systems. Robust digital infrastructure, equitable access to technology, and supportive policies are crucial for seamless implementation.

We advocate for systematic design and research focused on the cognitive learning effects of AI-enhanced education. These should consider possible inversion, substitution, augmentation, and redefinition effects in order to effectively leverage AI-enhanced learning and address potential risks to cognitive outcomes. By prioritizing systematic research and evidence-based approaches, we can move beyond the hype and ensure that AI in education leads to meaningful improvements in student learning outcomes.

**Acknowledgements** The third author was funded in a cooperative agreement with the US Army DEVCOM Soldier Center (W912 CG-24-2-0001) and by grants with the Institute of Education Sciences of the US Department of Education (R305 A200413, R305 T240021). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies of these funding agencies.

**Author Contributions** EB: conceptualization, writing—original draft, writing—review and editing; SG: writing—review and editing; ACG: writing—review and editing; KS: writing—review and editing; MS: conceptualization, writing—review and editing.

**Funding** Open Access funding enabled and organized by Projekt DEAL. US Army DEVCOM Soldier Center, W912 CG-24-2-0001, Arthur C. Graesser, Institute of Education Sciences of the US Department of Education, R305 A200413, Arthur C. Graesser, R305 T240021, Arthur C. Graesser

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abd-Alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., ... & Sheikh, J. (2023). Large language models in medical education: opportunities, challenges, and future directions. *JMIR Medical Education*, 9(1), e48291. <https://doi.org/10.2196/48291>

- Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., ... Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, *53*(8), 18–28. <https://doi.org/10.1109/MC.2020.2996587>
- Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*, 334–345. <https://doi.org/10.1016/j.compedu.2018.07.021>
- Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, *100*(8), 2965–2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*(3), 277–320. <https://doi.org/10.3102/00346543073003277>
- Almadhoob, A. H., Saleh, A. S. K., & Akbar, F. (2024). QuizWiz: Integrating generative artificial intelligence in an online study tool. In *Proceedings of the 2024 7th International Conference on Big Data and Education* (pp. 87–96). <https://doi.org/10.1145/3704289.3704296>
- Anderson, J. R., Betts, S., Bothell, D., Hope, R., & Lebiere, C. (2019). Learning rapid and precise skills. *Psychological review*, *126*(5), 727–760. <https://doi.org/10.1037/rev0000152>
- Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3), 1492–1501. <https://www.ijee.ie/contents/c320316B.html>
- Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M. D., Wortha, F., & Cerezo, R. (2022). Lessons learned and future directions of metatutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*, *13*, Article 813632. <https://doi.org/10.3389/fpsyg.2022.813632>
- Bachiri, Y. A., Mouncif, H., spsampsps Bouikhalene, B. (2023). Optimizing learning outcomes and retention in MOOCs with AI-generated flashcards. In *International Conference on Smart Learning Environments* (pp. 242–247). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-99-5961-7\\_32](https://doi.org/10.1007/978-981-99-5961-7_32)
- Backfisch, I., Scherer, R., Siddiq, F., Lachner, A., & Scheiter, K. (2021). Teachers' technology use for teaching: Comparing two explanatory mechanisms. *Teaching and Teacher Education*, *104*, 103390. <https://doi.org/10.1016/j.tate.2021.103390>
- Bai, S., Gonda, D. E., & Hew, K. F. (2024). Write-curate-verify: A case study of leveraging generative AI for scenario writing in scenario-based learning. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/TLT.2024.3378306>
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback?. *International Journal of Educational Technology in Higher Education*, *21*(1), 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., et al. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology*, *54*(5), 1222–1245. <https://doi.org/10.1111/bjet.13336>
- Bauer, E., Sailer, M., Niklas, F., Greiff, S., Sarbu-Rothsching, S., Zottmann, J. M., ... & Fischer, F. (2025). AI-based adaptive feedback in simulations for teacher education: An experimental replication in the field. *Journal of Computer Assisted Learning*, *41*(1), e13123. <https://doi.org/10.1111/jcal.13123>
- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in stem education: A meta-analysis. *Review of Educational Research*, *87*(2), 309–344. <https://doi.org/10.3102/0034654316670999>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., ... & Nerdel, C. (2025). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences*, *118*, 102601. <https://doi.org/10.1016/j.lindif.2024.102601>

- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). Introduction to meta-analysis. *John Wiley & Sons*. <https://doi.org/10.1002/9780470743386>
- Breideband, T., Bush, J., Chandler, C., Chang, M., Dickler, R., Foltz, P., ... & D'Mello, S. (2023). The Community Builder (CoBi): Helping students to develop better small group collaborative learning skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 376–380). <https://doi.org/10.1145/3584931.3607498>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–21. <https://doi.org/10.1145/3449287>
- Butterfuss, R., Roscoe, R. D., Allen, L. K., McCarthy, K. S., & McNamara, D. S. (2022). Strategy uptake in writing pal: Adaptive feedback and instruction. *Journal of Educational Computing Research*, 60(3), 696–721. <https://doi.org/10.1177/07356331211045304>
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., Fischer, F., & DFG Research group COSIMA. (2020). Facilitating diagnostic competences in higher education—A meta-analysis in medical and teacher education. *Educational Psychology Review*, 32(1), 157–196. <https://doi.org/10.1007/s10648-019-09492-2>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459. <https://doi.org/10.3102/00346543053004445>
- Clark, R. E. (1994a). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. <https://doi.org/10.1007/BF02299088>
- Clark, R. E. (1994b). Media and Method. *Educational Technology and Development*, 42, 7–10. <https://doi.org/10.1007/BF02298090>
- Crompton, H. (2017). Moving toward a mobile learning landscape: Presenting a mlearning integration framework. *Interactive Technology and Smart Education*, 14(2), 97–109. <https://doi.org/10.1108/ITSE-02-2017-0018>
- Crossley, S., Choi, J. S., Morris, W., Holmes, L., Joyner, D., & Gupta, V. (2024). Using intelligent texts in a computer science classroom: Findings from an iTELL deployment. In Y. Shi, P. Brusilovsky, B. Akram, T. Price, J. Leinonen, K. Koedinger, & A. Lan (Eds.). *Proceedings of the educational data mining in computer science education workshop 2024* (pp. 11–19). [https://ceur-ws.org/Vol-3796/CSEDM-24\\_paper\\_8293.pdf](https://ceur-ws.org/Vol-3796/CSEDM-24_paper_8293.pdf)
- D'Mello, S. K., Duran, N., Michaels, A., & Stewart, A. E. (2024). Improving collaborative problem-solving skills via automated feedback and scaffolding: A quasi-experimental study with CPS-Coach 2.0. *User Modeling and User-Adapted Interaction*, 34(4), 1087–1125. <https://doi.org/10.1007/s11257-023-09387-6>
- D'Mello, S. K., spsampsps Graesser, A. (2023). Intelligent tutoring systems: How computers achieve learning gains that rival human tutors. In *Handbook of educational psychology* (pp. 603–629). Routledge. <https://doi.org/10.4324/9780429433726>
- Dekel, A., Shechtman, S., Fernandez, R., Haws, D., Kons, Z., & Hoory, R. (2024). Speak while you think: Streaming speech synthesis during text generation. In *ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 11931–11935). <https://doi.org/10.1109/ICASSP48485.2024.10446214>
- Delen, I., & Sen, S. (2023). Effect of design-based learning on achievement in K-12 education: A meta-analysis. *Journal of Research in Science Teaching*, 60(2), 330–356. <https://doi.org/10.1002/tea.21800>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Du Boulay, B., Mitrovic, A., & Yacef, K. (Eds.). (2023). *Handbook of artificial intelligence in education*. Edward Elgar Publishing. <https://doi.org/10.4337/9781800375413>
- Du, H., Jia, Q., Gehringer, E., & Wang, X. (2024). Harnessing large language models to auto-evaluate the student project reports. *Computers and Education: Artificial Intelligence*, 7, 100268. <https://doi.org/10.1016/j.caeai.2024.100268>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 00, 1–42. <https://doi.org/10.1111/bjet.13544>

- Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: Benefits and challenges. *Frontiers in Education*, 9, 1416307. <https://doi.org/10.3389/educ.2024.1416307>
- Fiore, S. M., Graesser, A., & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*, 2(6), 367–369. <https://doi.org/10.1038/s41562-018-0363-y>
- Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowsch, A., Neuhaus, B. J., et al. (2022). Representational scaffolding in digital simulations—Learning professional practices in higher education. *Information and Learning Sciences*, 123(11/12), 645–665. <https://doi.org/10.1108/ILS-06-2022-0076>
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1), 56–66. <https://doi.org/10.1080/00461520.2012.748005>
- Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific reports*, 13(1), 15310. <https://doi.org/10.1038/s41598-023-42227-6>
- Fütterer, T., Scherer, R., Scheiter, K., Stürmer, K., & Lachner, A. (2023). Will, skills, or conscientiousness: What predicts teachers' intentions to participate in technology-related professional development?. *Computers & Education*, 198, 104756. <https://doi.org/10.1016/j.compedu.2023.104756>
- Ghosh, S. (2024). Rehumanizing education in the age of technology. In Adarsh Garg, B V Babu, Valentina E. Balas (eds.) *Advances in technological innovations in higher education: Theory and practices* (pp. 1–15). CRC Press. <https://doi.org/10.1201/9781003376699>
- Graesser, A. C. (2015). Deeper learning with advances in discourse science and technology. *Policy Insights from Behavioral and Brain Sciences*, 2, 42–50. <https://doi.org/10.1177/2372732215600888>
- Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, 26(1), 124–132. <https://doi.org/10.1007/s40593-015-0086-4>
- Graesser, A. C., Forsyth, C. M., & Lehman, B. A. (2017). Two heads may be better than one: Learning from computer agents in conversational dialogues. *Teachers College Record*, 119(3), 1–20. <https://doi.org/10.1177/016146811711900309>
- Graesser, A. C., Hu, X., Nye, B., & Sottilare, R. (2016). Intelligent tutoring systems, serious games, and the generalized intelligent framework for tutoring (GIFT). In H. F. O'Neil, E. L. Baker, and R. S. Perez (Eds.), *Using games and simulation for teaching and assessment* (pp. 58–79). Abingdon, UK: Routledge. <https://doi.org/10.4324/9781315817767>
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137. <https://doi.org/10.3102/00028312031001104>
- Grassini, S. (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, 13(7), 692. <https://doi.org/10.3390/educs13070692>
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, (13), 74–83. <https://doi.org/10.1016/j.edurev.2014.10.002>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hicke, Y., Agarwal, A., Ma, Q., & Denny, P. (2023). AI-TA: Towards an intelligent question-answer teaching assistant using open-source LLMs. *arXiv preprint arXiv:2311.02775*. <https://doi.org/10.48550/arXiv.2311.02775>
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 38. <https://doi.org/10.1007/s10676-024-09775-5>
- Hoff, T. (2011). Deskilling and adaptation among primary care physicians using two work innovations. *Health Care Management Review*, 36(4), 338–348. <https://doi.org/10.1097/HMR.0b013e31821826a1>
- Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). Dialect prejudice predicts AI decisions about people's character, employability, and criminality. *arXiv preprint arXiv: 2403.00742*. <https://doi.org/10.48550/arXiv.2403.00742>
- Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing. <https://doi.org/10.54675/EWZM9535>
- Holstein, K., Yu, Z., Sewall, J., Popescu, O., McLaren, B. M., spsampsps Aleven, V. (2018). Opening up an intelligent tutoring system development environment for extensible student modeling. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June*

- 27–30, 2018, *Proceedings, Part I 19* (pp. 169–183). Springer International Publishing. [https://doi.org/10.1007/978-3-319-93843-1\\_13](https://doi.org/10.1007/978-3-319-93843-1_13)
- Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review*, 36(1), 1–20. <https://doi.org/10.1007/s10648-024-09868-z>
- Hutt, S., & Hieb, G. (2024). Scaling up mastery learning with generative AI: Exploring how generative AI can assist in the generation and evaluation of mastery quiz questions. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale* (pp. 310–314). <https://doi.org/10.1145/3657604.3664699>
- Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54(5), 1057–1076. <https://doi.org/10.1111/bjjet.13325>
- Jivet, I., Scheffel, M., Drachler, H., spsampsps Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In É. Lavoué, H. Drachler, K. Verbert, J. Broisin, spsampsps M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (Vol. 10474, pp. 82–96). Springer International Publishing. [https://doi.org/10.1007/978-3-319-66610-5\\_7](https://doi.org/10.1007/978-3-319-66610-5_7)
- Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26(1), 25–36. <https://doi.org/10.1007/s40593-015-0065-9>
- Jun, S., Han, S., & Kim, S. (2017). Effect of design-based learning on improving computational thinking. *Behaviour & Information Technology*, 36(1), 43–53. <https://doi.org/10.1080/0144929X.2016.1188415>
- Kakar, S., Maiti, P., Taneja, K., Nandula, A., Nguyen, G., Zhao, A., ... spsampsps Goel, A. (2024). Jill Watson: scaling and deploying an AI conversational agent in online classrooms. In A. Sifaleras spsampsps F. Lin (Eds.), *International Conference on Intelligent Tutoring Systems* (pp. 78–90). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-63028-6\\_7](https://doi.org/10.1007/978-3-031-63028-6_7)
- Kali, Y., & Linn, M. C. (2008). Designing effective visualizations for elementary school science. *The Elementary School Journal*, 109(2), 181–198. <https://doi.org/10.1086/590525>
- Kasneeci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneeci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Knoop-van Campen, C. A. N., Wise, A., & Molenaar, I. (2023). The equalizing effect of teacher dashboards on feedback in K-12 classrooms. *Interactive Learning Environments*, 31(6), 3447–3463. <https://doi.org/10.1080/10494820.2021.1931346>
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., ... & Ryan, M. (2003a). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice. *The journal of the learning sciences*, 12(4), 495–547. [https://doi.org/10.1207/S15327809JLS1204\\_2](https://doi.org/10.1207/S15327809JLS1204_2)
- Kolodner, J. L., Gray, J., & Fasse, B. B. (2003b). Promoting transfer through case-based reasoning: Rituals and practices in learning by design classrooms. *Cognitive Science Quarterly*, 3(2), 183–232.
- Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the ACM collective intelligence conference* (pp. 12–24). <https://doi.org/10.1145/3582269.3615599>
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61(2), 179–211. <https://doi.org/10.3102/00346543061002179>
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42, 7–19. <https://doi.org/10.1007/BF02299087>
- Lachner, A., Backfisch, I., & Franke, U. (2024). Towards an integrated perspective of teachers' technology integration: A preliminary model and future research directions. *Frontline Learning Research*, 12(1), 1–15. <https://doi.org/10.14786/flr.v12i1.1179>
- Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjjet.13505>
- Leiker, D., Gyllen, A. R., Eldesouky, I., spsampsps Cukurova, M. (2023). Generative AI for learning: investigating the potential of learning videos with synthetic virtual instructors. In *International*

- conference on artificial intelligence in education (pp. 523–529). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-36336-8\\_81](https://doi.org/10.1007/978-3-031-36336-8_81)
- Levy, J., Jotkowitz, A., & Chowers, I. (2019). Deskillling in ophthalmology is the inevitable controllable? *Eye*, 33(3), 347–348. <https://doi.org/10.1038/s41433-018-0252-7>
- Lim, L., Bannert, M., Van Der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., et al. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139, Article 107547. <https://doi.org/10.1016/j.chb.2022.107547>
- Lindner, A., Romeike, R., Jasute, E., & Pozdniakov, S. (2019). Teachers' perspectives on artificial intelligence. In *12th International conference on informatics in schools, "Situation, evaluation and perspectives"*, ISSEP.
- Liu, K., Tschinkel, R., & Miller, R. (2024). Digital equity and school leadership in a post-digital world. *EECNU Review of Education*. <https://doi.org/10.1177/20965311231224083>
- Liu, Q., Geertshuis, S., & Grainger, R. (2020). Understanding academics' adoption of learning technologies: A systematic review. *Computers & Education*, 151, 103857. <https://doi.org/10.1016/j.compedu.2020.103857>
- Long, Y., & Alevan, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, 27(1), 55–88. <https://doi.org/10.1007/s11257-016-9186-6>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology*, 106(4), 901. <https://doi.org/10.1037/a0037123>
- Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education*, 3(1), 15–25. <https://doi.org/10.1016/j.hpe.2017.01.003>
- Mascha, M. F., & Smedley, G. (2007). Can computerized decision aids do "damage"? A case for tailoring feedback and task complexity based on task experience. *International Journal of Accounting Information Systems*, 8(2), 73–91. <https://doi.org/10.1016/j.accinf.2007.03.001>
- Matcha, W., Uzir, N. A. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE transactions on learning technologies*, 13(2), 226–245. <https://doi.org/10.1109/TLT.2019.2916802>
- May, T. A., Fan, Y. K., Stone, G. E., Koskey, K. L., Sondergeld, C. J., Folger, T. D., ... & Johnson, C. C. (2025). An effectiveness study of generative artificial intelligence tools used to develop multiple-choice test items. *Education Sciences*, 15(2), 144. <https://doi.org/10.3390/educsci15020144>
- Meyer, L., & Dannecker, A. (2024). Comparative analysis of generative AI models in educational exercise performance. In *EDULEARN24 Proceedings* (pp. 5181–5190). <https://doi.org/10.21125/edulearn.2024.1273>
- Miller, C., & Bartlett, J. (2012). 'Digital fluency': Towards young people's critical use of the internet. *Journal of Information Literacy*, 6(2). <https://doi.org/10.11645/6.2.1714>
- Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and Generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>
- Narciss, S., Sosnovsky, S., Schnaubert, L. et al. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56–76. <https://doi.org/10.1016/j.compedu.2013.09.011>
- Nazar, A. M., Selim, M. Y., Gaffar, A., & Ahmed, S. (2024). Revolutionizing undergraduate learning: CourseGPT and Its Generative AI Advancements. *arXiv preprint arXiv:2407.18310*. <https://doi.org/10.48550/arXiv.2407.18310>
- Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, 224, 105164. <https://doi.org/10.1016/j.compedu.2024.105164>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ninaus, M., & Sailer, M. (2022). Closing the loop—The human role in artificial intelligence for education. *Frontiers in psychology*, 13, 956798. <https://doi.org/10.3389/fpsyg.2022.956798>
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423–451. [https://doi.org/10.1207/s15327809jls1303\\_6](https://doi.org/10.1207/s15327809jls1303_6)

- Perković, G., Drobnjak, A., & Botički, I. (2024). Hallucinations in LLMs: Understanding and Addressing Challenges. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 2084–2088). IEEE. <https://doi.org/10.1109/MIPRO60963.2024.10569238>
- NASEM (National Academies of Sciences, Engineering, and Medicine), 2018. How people learn II: Learners, contexts, and cultures. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24783>
- Puentedura, R. (2006). Transformation, technology, and education [Blog post]. <http://hippasus.com/resources/tte/>. Accessed 01.02.2025
- Puentedura, R. (2014). Building transformation: An introduction to the SAMR model [Blog post]. [http://www.hippasus.com/rpweblog/archives/2014/08/22/BuildingTransformation\\_AnIntroductionToSAMR.pdf](http://www.hippasus.com/rpweblog/archives/2014/08/22/BuildingTransformation_AnIntroductionToSAMR.pdf). Accessed 01.02.2025
- Rafner, J., Dellermann, D., Hjorth, A., Verasztó, D., Kampf, C., Mackay, W., & Sherson, J. (2021). Deskillung, upskilling, and reskilling: A case for hybrid intelligence. *Morals & Machines*, *1*(2), 24–39. <https://doi.org/10.5771/2747-5174-2021-2>
- Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56. <https://doi.org/10.37074/jalt.2023.6.1.29>
- Rus, V., Olney, A. M., & Graesser, A. C. (2023). Deeper learning through interactions with students in natural language. *Handbook of Artificial Intelligence in Education*, 250–272. <https://doi.org/10.4337/9781800375413.00021>
- Sailer, M., Maier, R., Berger, S., Kastorff, T., & Stegmann, K. (2024a). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences*, *112*, 102446. <https://doi.org/10.1016/j.lindif.2024.102446>
- Sailer, M., Ninaus, M., Huber, S. E., Bauer, E., & Greiff, S. (2024b). The end is the beginning is the end: The closed-loop learning analytics framework. *Computers in Human Behavior*, 108305. <https://doi.org/10.1016/j.chb.2024.108305>
- Sailer, M., Schultz-Pernice, F., & Fischer, F. (2021). Contextual facilitators for learning activities involving technology in higher education: The Cb-model. *Computers in Human Behavior*, *121*, 106794. <https://doi.org/10.1016/j.chb.2021.106794>
- Salomon, G. (1979). *Interaction of Media, Cognition, and Learning*. Jossey-Bass. <https://doi.org/10.4324/9780203052945>
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. <https://doi.org/10.1016/j.compedu.2018.09.009>
- Schmucker, R., Xia, M., Azaria, A., Mitchell, T. (2024). Ruffle spsampsRiley: Insights from designing and evaluating a large language model-based conversational tutoring System. In A. M. Olney, I. A. Chounta, Z. Liu, O. C. Santos, spsamps I. I. Bittencourt (Eds.) *Artificial Intelligence in Education. AIED 2024. Lecture Notes in Computer Science*. Springer, Cham. [https://doi.org/10.1007/978-3-031-64302-6\\_6](https://doi.org/10.1007/978-3-031-64302-6_6)
- Sommerhoff, D., Codreanu, E., Nickl, M., Ufer, S., & Seidel, T. (2023). Pre-service teachers' learning of diagnostic skills in a video-based simulation: Effects of conceptual vs. interconnecting prompts on judgment accuracy and the diagnostic process. *Learning and Instruction*, *83*, 101689. <https://doi.org/10.1016/j.learninstruc.2022.101689>
- Spector, J. M., & Ma, S. (2019). Inquiry and critical thinking skills for the next generation: From artificial intelligence back to human intelligence. *Smart Learning Environments*, *6*(1), 1–11. <https://doi.org/10.1186/s40561-019-0088-z>
- Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 108386. <https://doi.org/10.1016/j.chb.2024.108386>
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331–347. <https://doi.org/10.1037/a0034752>
- Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, *30*, 100314. <https://doi.org/10.1016/j.edurev.2020.100314>

- Sun, S., Else-Quest, N. M., Hodges, L. C., French, A. M., & Dowling, R. (2021). The effects of ALEKS on mathematics learning in K-12 and higher education: A meta-analysis. *Investigations in Mathematics Learning*, 13(3), 182–196. <https://doi.org/10.1080/19477503.2021.1926194>
- Sutton, S. G., Arnold, V., & Holt, M. (2018). How much automation is too much? Keeping the human relevant in knowledge work. *Journal of Emerging Technologies in Accounting*, 15(2), 15–25. <https://doi.org/10.2308/jeta-52311>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22, 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *arXiv preprint*. arXiv:2311.14096. <https://doi.org/10.48550/arXiv.2311.14096>
- UNESCO (2023). Global education monitoring report 2023: Technology in education—a tool on whose terms?. UN. <https://doi.org/10.54676/UZQV8501>
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *SAGE Open*, 10(1), 93–104. <https://doi.org/10.1016/j.chb.2019.06.017>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–38. <https://doi.org/10.1145/3579605>
- Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., ... & Kizilcec, R. F. (2023). Teachers' trust and perceptions of AI in education: The role of culture and AI self-efficacy in six countries. *arXiv e-prints*, arXiv-2312. <https://doi.org/10.48550/arXiv.2312.01627>
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511. <https://doi.org/10.1007/s10648-016-9361-7>
- Vogt, K., & Flindt, N. (2023). Artificial intelligence and the future of language teacher education: A critical review of the use of AI tools in the foreign language classroom. *The future of Teacher Education: Innovations across Pedagogies, Technologies and Societies*, 179–199. [https://doi.org/10.1163/9789004678545\\_008](https://doi.org/10.1163/9789004678545_008)
- Williams, R., Kaputsos, S. P., & Breazeal, C. (2021). Teacher perspectives on how to train your robot: A middle school AI and ethics curriculum. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 17, pp. 15678–15686). <https://doi.org/10.1609/aaai.v35i17.17847>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10–33. <https://doi.org/10.1111/bjjet.13334>
- Xhakaj, F., Aleven, V., Spsampsp McLaren, B. M. (2017). Effects of a dashboard for an intelligent tutoring system on teacher knowledge, lesson plans and class sessions. In E. Andr'e, R. Baker, X. Hu, M. M. T. Rodrigo, Spsampsp B. Du Boulay (Eds.), *Artificial intelligence in education* (Vol. 10331, pp. 582–585). Springer International Publishing. [https://doi.org/10.1007/978-3-319-61425-0\\_69](https://doi.org/10.1007/978-3-319-61425-0_69)
- Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2024). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*. Advance online publication. <https://doi.org/10.1111/bjjet.13530>
- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839–1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yeh, Y. F., Chan, K. K. H., & Hsu, Y. S. (2021). Toward a framework that connects individual TPACK and collective TPACK: A systematic review of TPACK studies investigating teacher collaborative

- discourse in the learning by design process. *Computers & education*, 171, 104238. <https://doi.org/10.1016/j.compedu.2021.104238>
- Zapata-Rivera, D., spsamsp Arslan, B. (2024). Learner modeling interpretability and explainability in intelligent adaptive systems. In *Mind, Body, and Digital Brains* (pp. 95–109). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-58363-6\\_7](https://doi.org/10.1007/978-3-031-58363-6_7)
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhang, L., Lin, J., Kuang, Z., Xu, S., Yeasin, M., & Hu, X. (2024). SPL: A socratic playground for learning powered by large language mode. *arXiv preprint*. arXiv:2406.13919. <https://doi.org/10.48550/arXiv.2406.13919>
- Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, 143, 103668. <https://doi.org/10.1016/j.compedu.2019.103668>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Elisabeth Bauer<sup>1</sup>  · Samuel Greiff<sup>2</sup> · Arthur C. Graesser<sup>3</sup> · Katharina Scheiter<sup>4</sup> · Michael Sailer<sup>1</sup>

✉ Elisabeth Bauer  
elisabeth.bauer@uni-a.de

<sup>1</sup> Learning Analytics and Educational Data Mining, University of Augsburg, Augsburg, Germany

<sup>2</sup> Centre for International Student Assessment, Technical University of Munich, Munich, Germany

<sup>3</sup> Department of Psychology and Institute of Intelligent Systems, University of Memphis, Memphis, TN, USA

<sup>4</sup> Digital Education, University of Potsdam, Potsdam, Germany