Educational Psychology Review (2024) 36:31
https://doi.org/10.1007/s10648-024-09855-4

META-ANALYSIS

Effects of Artificial Intelligence-Powered Virtual Agents on Learning Outcomes in Computer-Based Simulations: A Meta-Analysis

Chih-Pu Dai[1] · Fengfeng Ke[2] · Yanjun Pan[3] · Jewoong Moon[4] · Zhichun Liu[5]

Accepted: 24 January 2024 / Published online: 1 March 2024

Abstract
Computer-based simulations for learning offer affordances for advanced capabilities and expansive possibilities for knowledge construction and skills application. Virtual agents, when powered by artificial intelligence (AI), can be used to scaffold personalized and adaptive learning processes. However, a synthesis or a systematic evaluation of the learning effectiveness of AI-powered virtual agents in computer-based simulations for learning is still lacking. Therefore, this meta-analysis is aimed at evaluating the effects of AI-powered virtual agents in computer-based simulations for learning. The analysis of 49 effect sizes derived from 22 empirical studies suggested a medium positive overall effect, $\overline{g} = 0.43$, SE = 0.08, 95% C.I. [0.27, 0.59], favoring the use of AI-powered virtual agents over the non-AI-powered virtual agent condition in computer-based simulations for learning. Further, moderator analyses revealed that intervention length, AI technologies, and the representation of AI-powered virtual agents significantly explain the heterogeneity of the overall effects. Conversely, other moderators, including education level, domain, the role of AI-powered virtual agents, the modality of AI-powered virtual agents, and learning environment, appeared to be universally effective among the studies of AI-powered virtual agents in computer-based simulations for learning. Overall, this meta-analysis provides systematic and existing evidence supporting the adoption of AI-powered virtual agents in computer-based simulations for learning. The findings also inform about evidence-based design decisions on the moderators analyzed.

Keywords Artificial intelligence · Computer-based simulation · Machine learning · Pedagogical agents · Simulation-based learning · Virtual agents

Extended author information available on the last page of the article

Introduction

Backgrounds and Definitions of Terms

Learning can occur when learners are actively engaged in iterative problem-solving processes for knowledge construction and skills application (e.g., Kolodner, 1992; Wood et al., 1976). To maximize opportunities for active knowledge construction and skills application, education researchers and practitioners leverage simulation-based learning, especially computer-based simulations for learning (e.g., de Jong & van Joolingen, 1998).

Simulations, as one way to present simulated experience for realistic problem solving, have been used for activate learning and knowledge construction for learners. Computer-based simulations for learning, including virtual reality, virtual laboratories, simulation games, or medical simulations, are a common form of simulations to foster such a learning experience. Despite the potentials of computer-based simulations for learning, scholars (e.g., de Jong & van Joolingen, 1998) have pointed out that learners can struggle when discovering, experimenting with, and constructing knowledge in computer-based simulations for learning. The multimodal forms of computer-based simulations for learning can also increase learners' cognitive load (Sweller et al., 2019). As such, integrating virtual agents in computer-based simulations for learning is considered a viable approach to support learners (Dai & Ke, 2022) while maintaining *desirable difficulties* (Bjork & Bjork, 2011) during the learning experiences in computer-based simulations.

Multiple studies have focused on exploring the use of virtual agents in computer-based simulations for learning (see Castro-Alonso et al., 2021; Peng & Wang, 2022). Recent virtual agents have undergone a transformative evolution, emerging as more powerful learning tools through the infusion of AI. With the increasing demand to evaluate the impacts of integrating various AI technologies into virtual agents to enhance learning in computer-based simulations, a warranted meta-analysis that synthesizes literature from different sources to examine the effects is essential.

With an aim to examine the effects and to derive design implications of AI-powered virtual agents in computer-based simulations for learning, we begin by providing detailed definitions in the following sections for key terms used in this meta-analysis—*AI*, *virtual agents*, *computer-based simulations for learning*, and *AI-powered virtual agents in computer-based simulations for learning*.

Definition of AI

In education, AI has been used in intelligent tutoring systems, learning analytics, classroom assistance, or learning diagnosis and assessment. AI technologies in education include subsets of machine learning, deep learning, or natural language processing. In the current study, we focus on the AI that simulates human character for learning interactions. Embryonic definition of AI holds that “every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” (McCarthy et al., 2006, p. 12). McCarthy et al. (2006) proposed seven AI prospects: automatic computers, how can a computer be programmed to use a language, neuron nets, theory of the size of a calculation, self-improvement, abstractions, and randomness and creativity. Thereby, we define AI in this meta-analysis as the computational-engineered machine intelligence that drives automatic human-like verbal and/or nonverbal interactions with the learners to achieve learning objectives.

Definition of Virtual Agents

Virtual agents can be used alternatively with educational virtual agents, pedagogical agents, virtual humans, virtual beings, intelligent agents, chatbots, or conversational agents (e.g., Cassell, 2001; Dai & Ke, 2022; Kim & Baylor, 2016; Russell & Norvig, 2021). Virtual agents can be embodied or unembodied (Araujo, 2018; Sinatra et al., 2021) and be adopted for behavioral, cognitive, and social learning purposes. While we focused on educational virtual agents, we used the term of virtual agents in this meta-analysis rather than pedagogical agents or other similar terms to broaden the scope for identifying studies on AI-powered virtual agents in computer-based simulations for learning. Research on virtual agents has been prolific. Virtual agents have been employed to motivate learners, stimulate situated interests for learners, support learners for cognitive and metacognitive learning activities, and provide learners with feedback for decision making, and assess learners based on a competency model (Ke et al., 2020). Correspondingly, research on virtual agents can focus on the cognitive and metacognitive functions, agent appearances and associated impacts on learners, and multimedia design principles.

Educational virtual agents, or virtual agents, are simulated, life-like characters in computer-based environments that can be used to interact with humans and facilitate learning (Mascarenhas et al., 2018). In general, virtual agents serve the purpose to communicate/interact with humans, verbally and/or nonverbally, via texts, animations, or voices in natural language and authentic manners. The existing studies suggest that virtual agent-facilitated learning not only provides authentic, adaptive, and personalized learning experiences, but also offers the opportunities for deep learning via constructive and iterative learning processes (Kulik & Fletcher, 2016).

Definition of Computer-Based Simulations for Learning

Simulation is an experience or activity that hypothetically imitates a compelling real-world situation or setting (Alinier, 2007; Tun et al., 2015). Simulation can differ in fidelity that is determined by the “verisimilitude of an experience” (Tun et al., 2015, p. 161). Simulation has been widely used in the aviation or engineering industry (Tun et al., 2015), clinical medical education (Decker et al., 2008), or teacher education (Ke et al., 2020; Dai, 2023; Dai et al., 2024). “A simulated learning experience imitates the working environment and requires the learner to demonstrate procedural techniques, decision making, and critical thinking" (Decker et al., 2008, p. 75). In computer-based simulations for learning, it can be delivered via immersive environments (e.g., virtual, or mixed reality, simulation games) and other computer-based environments (e.g., intelligent tutoring systems; VanLehn, 2011).

In the current meta-analysis, we define computer-based simulations for learning as computer-based environments that facilitate learning experience and produce compelling learning environments and learning scenarios for the purpose of fostering higher-order thinking, or complex operational knowledge and skills development (see "*AI, Virtual Agents, and Computer-Based Simulations for Learning*" section for details about the categories of computer-based simulations for learning in this meta-analysis). We aimed to be inclusive in our search terms and definition for computer-based simulations for learning. Hence, we followed the proposition of prior research (Merchant et al., 2014) to include both VR and simulation games, in addition to computer-based environments as computer-based simulations for learning.

Definition of AI-Powered Virtual Agents in Computer-Based Simulations for Learning

The potentials for learning could be maximized when integrating AI-powered virtual agents in computer-based simulations for learning (Kim & Baylor, 2016; Sinatra et al., 2021). In this section, we define "AI-powered virtual agents in computer-based simulations for learning," building on the definitions of AI, virtual agents, and computer-based simulations for learning introduced in earlier sections of this paper. The definition of AI-powered virtual agents has not been universal in the literature. In the current meta-analysis, AI-powered virtual agents in computer-based simulations for learning are virtual characters designed to improve learning by playing different roles in computer-based simulations such as mentors, social companions, virtual instructors, feedback or hints providers in the forms of texts, verbal speech, or with multimodality. AI-powered virtual agents leverage different computational-engineered machine intelligence to interact with learners, such as predefined rule-based mechanisms, modeling or knowledge-based technologies, and/or natural language processing (NLP)/machine learning (ML)-based algorithms. For example, Shiban et al. (2015) adopted AI-powered virtual agents in their computer-based simulation for math learning. The AI-powered virtual agents adopted were humanlike and acting as feedback providers in the forms of multimodality (i.e., texts and gestures) with rule-based mechanisms (see Fig. 1).

Another example (see Fig. 2) is the AI-powered virtual agent implemented in Le and Wartschinski (2018)'s study developing learners' reasoning skills. The AI-powered virtual agent was humanlike (i.e., through self-introduction in natural language with a humanlike name) but in the forms of text-based interactions acting as a mentor. The authors used NLP/ML algorithms to drive the interactions for adaptive and personalized learning.

Research Problems, Prior Reviews, and Purposes of the Current Meta-Analysis

For the past decades, research on virtual agents was fruitful (e.g., Baylor & Kim, 2009; Graesser et al., 2008; Nye et al., 2014). Implemented in computer-based simulations for learning, virtual agents have generally demonstrated benefits for

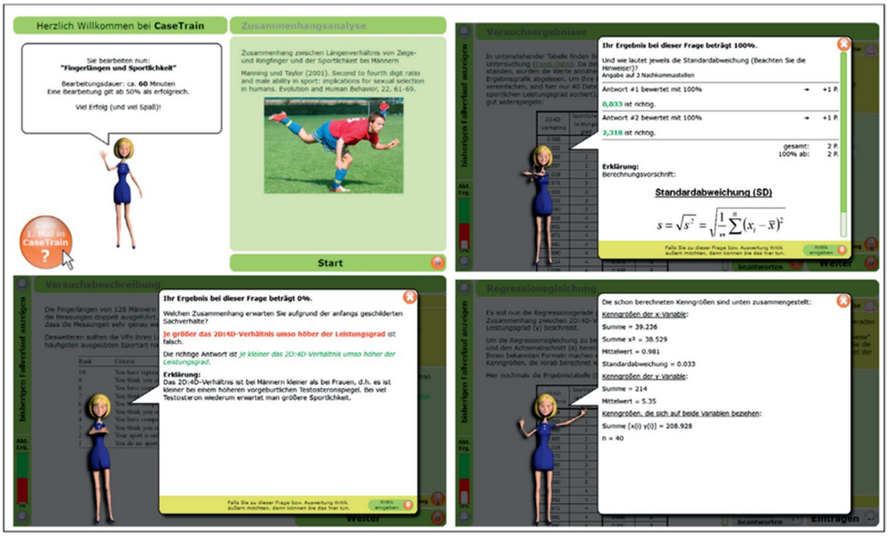

Fig. 1 An example of AI-powered virtual agent from Shiban et al., (2015, p. 8)

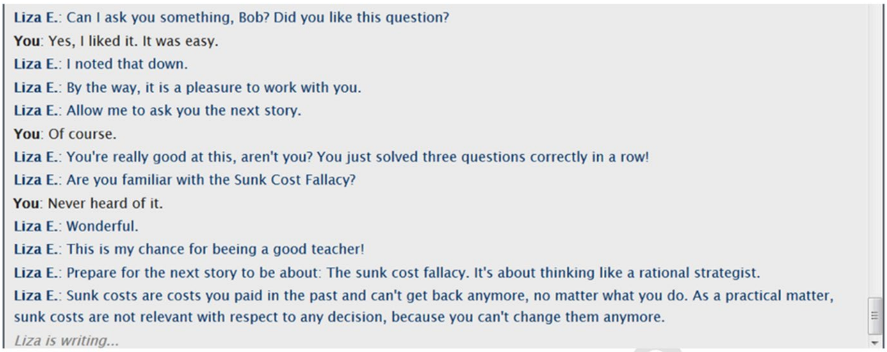

Fig. 2 An example of AI-powered virtual agent from Le and Wartschinski (2018, p. 48)

learners (Kulik & Fletcher, 2016). However, critical challenges of computer-based simulations for learning facilitated by virtual agents have been accounted for in the literature. For instance, as discussed by Veletsianos and Russell (2013), the design of the agent-learner interactions often failed to meet the learning goals due to limited interactions with the learners; further, the effectiveness of the learning process can be hindered by agents' inadequate interpretation of learners' input. In other words, the misalignment of agents' knowledge and abilities and learners' input caused inefficient interactions, as Veletsianos and Russell (2013) noted, "misclassification of user input can lead to agent responses that do not fit within the context or structure of the conversation" (p. 385). Despite reported issues in the literature, decades of development and advancement in AI have made natural language-based interactions and learning adaptivity more approachable and feasible (Dai & Ke, 2022). AI, replicating human intelligence, are computational systems designed to think and act rationally and effectively in a situation (such as problem solving, reasoning, and planning) (Seshia et al., 2022; Russell & Norvig, 2021). AI has been fast-growing and used in education to support teachers' classroom practices, ubiquitous learning, as well as exploration-oriented learning environments such as simulations and games (Roll & Wylie, 2016).

In particular, infusing AI in virtual agents for computer-based simulations for learning has grown over the past decades. Computer-based simulations for learning inherently accommodate virtual agents since humanlike interactions and authentic problem solving are essential in both applications. AI-powered virtual agents have been used to offer interactive demonstrations, navigational guidance, attentional guides with gaze and gestures, feedback, collaborative learning experiences, and story-rich problem-solving tasks (Rickel, 2001). Nevertheless, the underlying paradigms by which the agents are used to support learning have been contentious. For example, Schroeder and Gotch's (2015) study found that facilitation- and scaffolding-oriented intelligent virtual agents have reached a peak in the late 2000s and early 2010s, whereas AI agents acted as information providers in computer-based learning environments have burgeoned. Recently, AI-powered virtual agents for inquiry-based and experiential learning are also advocated (Dai & Ke, 2022; Hwang et al., 2020; Lippert et al., 2020; Rickel, 2001). It is argued that integrating AI-powered virtual agents in computer-based simulations for learning may enrich learning with authentic contexts, prompt exploration-based constructive learning, and engage learners in dialogue-based sensemaking interactions (Woolf et al., 2013) and discourse-rich collaborative problem solving (Graesser, 2016).

Given these divergent viewpoints, the design and development of AI-powered virtual agents require transdisciplinary breakthroughs to overcome the challenges such as constrained and oversimplified interactions (Luck & Aylett, 2000; Rickel, 2001). For example, Veletsianos and Russell (2013) revealed that virtual agents can provide responses with inadequate syntax or the absence of context in conversational AI. Cognitive, psychological, educational, and technological models need to work coherently to design and develop AI-powered virtual agents with high fidelity and a fine-grained architecture (e.g., Dai, 2024; Rickel, 2001) to achieve effective experiential learning. To this end, examining the AI components of virtual agents in computer-based simulations for learning is warranted. We built on prior meta-analyses on virtual agents in computer-based simulations for learning by endorsing and integrating the aforementioned varying applications with different learning paradigms. We extended the previous research efforts by systematically synthesizing the effects of *different AI technologies infused in virtual agents* in computer-based simulations for learning.

To elaborate, despite the acclaimed advantages of virtual agents, the overall effects of using various AI technologies in virtual agents implemented in computer-based simulations for learning are ambiguous in the literature. It is also unclear what design and implementation characteristics influence the effects of AI-powered virtual agents in computer-based simulations for learning. The current meta-analysis focuses on examining the overall effects and the salient design and implementation characteristics of using AI-powered virtual agents in computer-based simulations for learning.

Multiple meta-analyses on virtual agents (Castro-Alonso et al., 2021; Davis, 2018; Guo & Goh, 2015; Heidig & Clarebout, 2011; Peng & Wang, 2022; Schroeder et al., 2013), simulation-based learning or computer-based simulations for learning (Chernikova et al., 2020; Kulik & Fletcher, 2016; Merchant et al., 2014) have been conducted. Despite these former informative syntheses, limitations were presented in these prior meta-analyses. First, none of these existing meta-analyses provided empirical evidence governing the efficacy of virtual agents in computer-based simulations for learning infused with *AI technologies*. They either focused on virtual agents, simulation-based learning, or computer-based simulations for learning. For meta-analyses on virtual agents, multimedia design rather than the effects of AI elements was examined (e.g., Castro-Alonso et al., 2021; Davis, 2018; Guo & Goh, 2015). Second, the study characteristics in relation to these *AI-powered* virtual agents in computer-based simulations for learning were thus ambiguous in these prior meta-analyses; the examination of the moderators was inconclusive as well.

To address the gap in the prior related meta-analyses, the current meta-analysis extended prior meta-analyses by examining the effects of AI-powered virtual agents in computer-based simulations for learning. Our meta-analysis was unique in that it highlighted the combinations and dynamics of AI technologies, virtual agents, and computer-based simulations for learning. Specifically, we sought to address the following research questions (RQs):

RQ1. What is the overall effect of AI-powered virtual agents on learning outcomes in computer-based simulations for learning?

RQ2. Do the study characteristics (i.e., *intervention length, education level*, *domain*, *publication type*, *the role of AI-powered virtual agents*, *AI technologies in AI-powered virtual agents*, *the modality of AI-powered virtual agents*, *the representation of AI-powered virtual agents*, and *learning environment*) moderate the overall effect of AI-powered virtual agents in computer-based simulations for learning?

Potential Moderators

In studying AI-powered virtual agents in computer-based simulations for learning, several related moderators have been of interest in the literature: *Intervention length, Education level, Domain, Publication type, The role of AI-powered virtual agents, AI technologies, The modality of AI-powered virtual agents, Representation*, and *Learning environment.* We selected these potential moderators based on the existing research and the inconclusive nature of these moderators in the literature (e.g., Castro-Alonso et al., 2021; Johnson & Lester, 2016; Kim & Baylor, 2016), as well as our research purposes and questions. We elaborated on the rationales in detail in the following sections.

Intervention Length

Researchers and practitioners have been keen on understanding under what conditions and to what extent the intervention length (sometimes called “dosage”) affects learning in computer-based simulations for learning, especially when featuring design elements (e.g., virtual agents) that may demand cognitive resources from the learners (e.g., Chernikova et al., 2020; Jackson & McNamara, 2013; Tokac et al., 2019). Chernikova et al.’s (2020) meta-analysis on simulation-based learning in higher education reported that the longer the duration of simulation-based learning is, the higher the effect sizes of complex skills acquisition will be. Jackson and McNamara’s empirical study (2013) also found that learning performance was improved over time in an intelligent tutoring system. In contrast, Tokac et al.’s (2019) meta-analysis found that the length of game-based intervention does not explain the overall effect sizes in PreK-12 settings. Similarly, Merchant et al. (2014) did not find a significant association between learning outcomes and the duration of each session or time spent in computer-based simulations for learning or virtual worlds. On the one hand, while virtual agents have been frequently used in digital multimedia learning environments (Castro-Alonso et al., 2021), educational theories suggested that attention allocation, important for short-term cognitive rehearsal, plays a role in digital multimedia learning environments (Schweppe & Rummer, 2014). As indicated by the discussion of the role played by attention allocation, shorter interventions of digital multimedia learning environments can be more effective than longer ones (Kulik & Fletcher, 2016). On the other hand, research of knowledge construction based on experiential learning maintains that it takes time to constructively internalize the knowledge and skills learned (e.g., Carroll, 1963; Kolb, 1984). Due to the indeterminate state of the intervention length in the literature, we selected it as a potential moderator for further investigation.

Education Level

Learner’s stage of development and education has been a moderator in the evaluation of virtual agents in educational settings in prior meta-analyses (Castro-Alonso et al., 2021; Schroeder et al., 2013). Empirical studies on the associations between using AI-powered virtual agents, different educational levels, and learning outcomes were lacking. Using meta-analysis to synthesize the effects of learner’s stage of development and education became a viable way to inform the effects of AI-powered virtual agents across education levels. However, the results of the learner’s stage of development and education level were inconsistent in the literature. Earlier meta-analysis on the effectiveness of virtual agents indicated that the agents were more useful for K-12 students than for postsecondary students (Schroeder et al., 2013). In contrast, a recent meta-analysis (Castro-Alonso et al., 2021) reported that the postsecondary student group demonstrated a small positive effect size, but the overall heterogeneity moderated by education level was nonsignificant (Castro-Alonso et al., 2021). We built on these prior studies to further examine the moderating effect of education level in the contexts of virtual agents integrated with AI in computer-based simulations for learning.

Domain

The consideration of academic areas or subject domains as a potential moderator has been of interest in the literature on virtual agents but remain inconclusive. As suggested by Pavlik Jr. et al. (2013), "comparing learning gains carefully across different…domains…is difficult if not impossible to do in a valid way" (p. 41). Nonetheless, previous meta-analyses governing virtual agents and computer-based simulations for learning have shed light on the moderator of *domain*. For example, Castro-Alonso et al. (2021) found a nonsignificant difference between STEM and non-STEM domains, but significant positive effects for the disciplines of biology, computing, and English, and significant negative effects for history learning. Another meta-analysis reported that virtual agents used for learning science and math yielded significantly higher effect size than the agents used for learning humanities (Schroeder et al., 2013). In the literature of computer-based simulations for learning, domains of medical education, complex and soft skills (e.g., problem solving, negotiation, social skills), and language learning have been the domains frequently studied (Chernikova et al., 2020; Dai & Ke, 2022; Merchant et al., 2014; Peterson, 2010). Based on the prior research, in the current meta-analysis, we focused on examining five domains: math, science, medical studies, soft skills, and language learning.

Publication Type

Publication type is a general moderator in a meta-analysis research, since studies with significant effects are more likely to be published as peer-reviewed papers (Easterbrook et al., 1991). Meta-analyses that only include journal papers are prone to be negligent of insignificant findings. It is suggested that reviews should include both published and unpublished (e.g., dissertation) papers (Easterbrook et al., 1991). Thus, we listed publication type as a moderator by coding both published and unpublished papers whenever applicable.

Role-Specific Functionality of AI-Powered Virtual Agents

AI-powered virtual agents can play different roles and fulfill diverse functions in supporting learning (Heidig & Clarebout, 2011; Peng & Wang, 2022). Depending on the role-specific functionality of AI-powered virtual agents, the designated outcome or intended learning may differ (Luck & Aylett, 2000). Different studies discussed various role-specific functionality of AI-powered virtual agents. For instance, Dai and Ke (2022) synthesized that virtual agents play diverse roles in learning, offering guidance, cues, and serving as a social companion for learners. Moreover, virtual agents can provide hints and/or feedback (Rickel, 2001; Wang et al., 2008), as well as deliver instructional content (Schroeder & Gotch, 2015). Some alternative designs employed virtual agents to serve mixed roles, such as scaffolding inquiry and problem solving, facilitating reflections, or acting as a teachable agent (Dai & Ke, 2022). The distinctive and mixed roles assumed by AI-powered virtual agents resulted from the integration of various underlying theories and principles of learning in their design, highlighting the diverse educational frameworks that drive their functionality. In addition, the integration of AI into virtual agents could determine the roles played by these AI-powered virtual agents. Earlier meta-analyses focused on various aspects of learning facilitation, such as information processing (Heidig & Clarebout, 2011) and companion strategies (i.e., motivator, expert, and mentor) (Kim & Baylor, 2016), while we extended prior meta-analyses by exploring the moderating effects of *AI-powered* virtual agents' role-specific functionality in computer-based simulations for learning.

AI Technologies in Virtual Agents

Dai and Ke (2022) reported that AI technologies can be classified into Scripted AI, Rule-based AI, Module-based AI, and NLP/ML. Scripted AI and Rule-based AI are the two types of preset AI with restricted responsiveness and dynamicity. Scripted AI used coding scripts for a list of responses or reactions to be executed linearly (Spronck et al., 2006). Rule-based AI identifies both knowledge and responses with already existing algorithms throughout the machine decision-making processes (Maroengsit et al., 2019). Module-based AI applies AI modeling techniques to drive the human–computer interactions. For example, a module-based AI may use a mix of knowledge modeling, student modeling, or agent modeling with Bayesian network models to reason for uncertainty in decision making (Dai & Ke, 2022; Dai et al., 2021). NLP is considered a subcategory of ML (Jordan & Mitchell, 2015). ML-driven virtual agents with NLP adopt advanced algorithms such as convolution or artificial neural networks, decision trees, or support vectors to propel virtual agent communication with human learners. Crucial to investigating the effects of AI-powered virtual agents in computer-based simulations for learning are the distinctive functions of AI technologies within virtual agents. According to Peng and Wang (2022), different AI technologies contribute vitally to various levels of personalization and adaptivity within virtual agents, the argument they framed as "AI degree." Unique to this meta-analysis, AI technologies, as a moderator, can assist in understanding the impact of virtual agents integrated with various AI technologies, each offering distinct capabilities for interacting with learners in computer-based simulations, on learning outcomes.

Modality of AI-Powered Virtual Agents

Communications between virtual agents and learners can be accomplished in a multimodal manner (Ke et al., 2020; Johnson & Lester, 2016). Johnson and Lester's (2016) narrative review maintained that spoke virtual agents can produce better learning outcomes compared to virtual agents with text-only interactions. Kim's (2005) meta-analysis found that there were no significant differences between virtual agents that "used text ($d=0.29$), spoke ($d=0.53$), and were animated ($d=0.52$)" (as cited in Noetel et al., 2022, p. 430). However, the methodological limitations (e.g., unclear search terms, being a pilot study) called for further examination. Also, research about the modality principle of multimedia learning (see Castro-Alonso & Sweller, 2022) shows differences between written text and voice. The modality we examined in this meta-analysis included printed text, voice, and multimodal interactions (Johnson & Lester, 2016; Ke et al., 2020; Kim, 2005).

Representation of AI-Powered Virtual Agents

The manner that the virtual agents were anthropomorphized in computer-based simulations for learning is another factor that potentially moderates the learner-agent interactions and hence the learning outcomes (Johnson & Lester, 2016). The moderating effects of virtual agents' representation can be examined from alternative perspectives. One perspective is related to the complexity of the learner-agent interactions and the consequential demand on learners' cognitive resources. Research has examined two dimensional versus three dimensional virtual agents (Castro-Alonso et al., 2021), and showed that two-dimensional space is more effective for learning than three-dimensional space. Moreover, Davis (2018) studied humanoid and character agent types in terms of agent gesturing. The author found that agent representation (i.e., humanoid or character) did not moderate the effects of learning retention, but humanoid agents significantly decreased cognitive load whereas character agents increased extraneous cognitive load. Another perspective for the representation of virtual agents is its social fidelity and trustworthiness. Humanlike virtual agents can provide verbal and nonverbal communications (Baylor & Kim, 2009; Johnson & Lester, 2016), thus activating learners' interactions with perceived social fidelity (Kim & Baylor, 2016). Domagk (2010) classified virtual agents into human (e.g., real human recorded or animated humanlike character) and nonhuman (animal or fictional character). Empirical investigation is warranted to explore how humanlike agent, fictional (cartoon or wizard) agent, mixed humanlike/fictional agents, and humanlike agent with text-only communication moderate the effects of AI-powered virtual agents in computer-based simulations for learning.

Learning Environment

In this meta-analysis, we focus on computer-based simulations for learning. Specifically, computer-based system and VR/Simulation game were reviewed as two distinct types of learning environments. Although computer-based simulation for learning was found to be overall effective in higher education settings (Chernikova et al., 2020), Merchant et al.'s (2014) findings reported that simulation games were more effective than virtual reality and virtual worlds. Despite these prior meta-analyses, the moderating effects of the learning environment have been inconclusive in the literature, especially between computer-based system and immersive VR/simulation game (Dai et al., 2023). The hypothesis is that computer-based simulation environments can be beneficial for learning as long as they offer useful affordances with the integration of AI-powered virtual agents. According to Al-Elq (2010), these affordances may include hands-on experience, opportunities for repeated practice, the ability for learners to make mistakes and learn from failures, and immediate feedback.

Method

Literature Search

We searched digital databases including APA PsycInfo ($n=572$), APA PsycNet ($n=575$), ERIC ($n=229$), and Web of Science ($n=6950$) using the following keywords based on the literature in AI-powered virtual agents in computer-based simulations for learning (Chen et al., 2020; Dai & Ke, 2022; Jordan & Mitchell, 2015; Merchant et al., 2014; Sinatra et al., 2021): (*pedagogical agent OR animated pedagogical agent OR virtual tutor OR virtual agent OR virtual humans OR embodied agent OR conversational agent*) ***AND*** (*artificial intelligen* OR intelligent tutoring systems OR machine learning OR deep learning OR machine intelligen* OR natural language process**) ***AND*** (*education OR learning OR simulation OR simulation-based learning OR virtual reality OR augmented reality OR mixed reality OR virtual lab OR serious games OR game-based learning OR educational games OR learning games OR computer-based simulation*). Figure 3 outlined the search, inclusion, and exclusion procedures. Ultimately, 22 papers yielded 49 studies that met the inclusion and exclusion criteria for the effect size calculation.

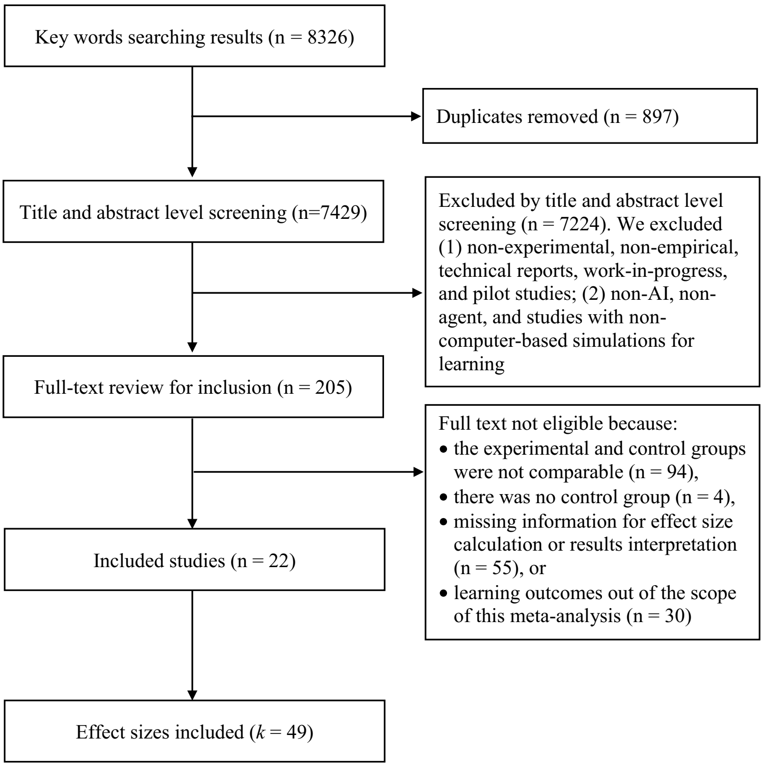

Fig. 3 Flow chart for the inclusion of studies in this meta-analysis

Inclusion and Exclusion Criteria

AI, Virtual Agents, and Computer-Based Simulations for Learning

To be included, the study must feature all three components—AI, virtual agents, and computer-based simulations for learning. For AI, the included studies should adopt at least one type of AI technologies defined in the literature (Dai & Ke, 2022, i.e., scripted AI, rule-based AI, module-based AI, and NLP/ML) to be coded as a moderator in this meta-analysis. Further, the included studies must feature at least a virtual agent to interact with the learners. To be included, the virtual agents in the study should demonstrate a simulated character, and they should initiate or maintain dialogues and/or use instructional strategies to facilitate learning (Gibbons, 2020; Sinatra et al., 2021) (for the definition of virtual agents, please see "Definition of Virtual Agents" section). The included studies should also adopt a form of computer-based simulations for learning. The form of the simulation can be either web-based simulation or VR/simulation games (Merchant et al., 2014). Both simulations enabled learner interactions via internet technologies mediated by graphical interfaces on a learner-side browser (Byrne et al., 2010). VR is an immersive and dynamic 3D environment, whereas simulation games constitute decision-making scenarios where learners learn from the consequences of the decisions they made (Dai & Ke, 2022; Sitzmann, 2011).

Learning Outcomes

In this meta-analysis, the dependent variable is the learning outcomes. According to Richey et al. (2011), learning outcomes can be behavioral, cognitive, and social. Learning outcomes can be assessed in a simulation, externally with tests, or via task performance. An example of behavioral outcomes would be computer skills development demonstrated by task completion and performance (e.g., van der Meij, 2013). For cognitive learning outcomes, knowledge retention, knowledge acquisition, or knowledge transfer were all considered. Social learning outcomes can be negotiation (e.g., Johnson, 2021) or communication skills (e.g., Kron et al., 2017). Due to our focus on performance-oriented learning outcomes, self-report beliefs, attitudes, or motivation were excluded in this meta-analysis.

Research Designs

Given the purpose of generalization with the meta-analytic method, only studies with experimental designs, including true/quasi-experimental designs, or randomized controlled trial, were included. The only difference between the experimental and control groups must be the presence/absence of AI-powered virtual agents in computer-based simulations for learning.

To be more specific, the experimental groups must adopt AI-powered virtual agents as an intervention *and* the control group must contain *no* AI-powered virtual agents, both in computer-based simulation environments for learning. In other words, we included studies that used AI-powered virtual agents in experimental conditions but used computer-based simulations for learning *without* AI-powered virtual agents in the control condition. We excluded studies that used AI-powered virtual agents in different learning environments or conditions, such as studies that examined AI-powered virtual agents in computer-based simulations for learning versus AI-powered virtual agents or human tutors in a classroom lecture setting (e.g., Daradoumis & Arguedas, 2020; Elkot, 2019).

Further, if learners in both groups have experienced AI-powered virtual agents in the experiments, the study will be excluded. For example, we excluded some studies that examined the presence/absence of scaffolding in AI-virtual-agent-integrated learning environments, as well as studies that examined the presence/absence of different features of AI-powered virtual agents (e.g., voice, appearance, or gestures). We also excluded studies that used AI-powered virtual agents in both the experimental and control conditions.

Reporting

Providing sufficient information for the calculation of effect sizes is an important inclusion criterion. To synthesize the effects, we calculated the mean difference. The information on means, standard deviations, sample size in both experimental and control groups must be presented to be included. If the studies reported standard errors, we calculated the standard deviations (Higgins et al., 2019). In addition to mean scores comparison, studies reported change scores from baseline were also included (see Deeks et al., 2019).

Publication Characteristics

We included studies in both peer- or non-peer reviewed outlets to avoid the file-drawer effect (Rosenthal, 1979). Journals, book chapters, conference proceedings, dissertations were all considered. However, only journal papers, conference proceedings, and dissertations met the final inclusion criteria. We only included studies published in the language used in this journal (i.e., English). We did not limit the year of publication.

Coding Procedures

Coding of each study was done by at least two trained coders. Twenty-five percent of the studies were first coded. After debriefing sessions, consensus was reached. After discussions for the disagreements, the two coders reached 100% consensus on the included studies. We identified 22 studies that met our inclusion and exclusion criteria. We then assigned shares of studies to be coded by at least two trained coders. We calculated inter-rater reliability between the coders using Cohen's kappa (Cohen, 1960). We used the *irr* package (Gamer et al., 2012) in R studio (R Core Team, 2019) to calculate the Cohen's kappa, which adjusts the overall agreement percentage to accommodate the expected level of agreement that might arise due to random assignment. The results of the kappa scores calculated for inter-rater reliability were as follows: Publication year (κ=1.00, z=20.8), Intervention length (κ=0.76, z=10.1), Education level (κ=1.00, z = 8.64), Domain (κ=0.80, z=9.71), Publication type (κ=1.00, z=8.92), the role of AI-powered virtual agents (κ=0.77, z=9.05), AI technologies in AI-powered virtual agents (κ=0.71, z=8.59), the modality of AI-powered virtual agents (κ=0.88, z=8.76), the representation of AI-powered virtual agents (κ=0.73, z=7.51), and the learning environment (κ=0.91, z=6.39).

Based on the literature and open analyses of the included studies, we categorized the codes in each moderator. Each moderator was dummy coded. The codes of *intervention length* were obtained openly from the included studies. We categorized it into "≤30 min," "between 31 and 60 min," "between 61 and 119 min," and "≥120 min." *Education level* was coded as "elementary," "middle/high school," and "college or adults." From the literature and with an open analysis from the included studies, we coded *domain* as "math," "science," "medical studies," "soft skills," and "language learning" (see "Domain" and "Learning Outcomes" sections for more details). We included both published and unpublished studies for an unbiased analysis. Because no book chapter met the inclusion criteria, the publication type was coded as "journal," "conference proceedings," and "dissertation."

For the AI-powered virtual agents, we coded four moderators (see the "Potential Moderators" section for the description of each code). First, *the role of AI-powered virtual agents* was coded as "guidance/cues," "feedback," "hints," "social companion," "instruction," and "mix." Second, four levels were coded for *AI technologies in AI-powered virtual agents*: "scripted AI," "rule-based AI," "module-based AI," and "NLP/ML." Third, "the modality of AI-powered virtual agents" was coded as "verbal," "text," and "multi-modal." Fourth, the *representation of AI-powered virtual agents* included the codes of "humanlike agent," "fictional (cartoon or wizard) agent," "mixed humanlike/fictional agents," "humanlike agent with text-only communication." Ultimately, the classification of the *learning environment*, reflecting the simulations used for learning, was consisted of "web-based simulation" and "VR/Simulation game."

Statistical Methods

Statistics for the Overall Effects

The effect sizes estimate mean differences. We first calculated Cohen's d, where $\overline{Y_T}$ is the posttest mean of treatment group, $\overline{Y_C}$ is the posttest mean of the control group, and $S_{Y\text{pooled}}$ is the pooled standard deviation:

$$\frac{\overline{Y}_T - \overline{Y}_C}{S_{Y\text{pooled}}}.$$

Subsequently, we used *Hedge's g* to adjust the effect sizes for unbiased, small-sample correction (Hedges, 1981). The formula was displayed as follows, where n_T is the sample size of the treatment group, n_C is the sample size of the control group:

$$g = \left[1 - \frac{3}{4\left(n_T + n_C - 2\right) - 1}\right] * d.$$

The sample variance of the estimate was calculated as the follows, where all definitions of term have been provided above:

$$v = \frac{n_T + n_C}{n_T n_C} + \frac{d^2}{2(n_T + n_C)}.$$

To determine the degree of homogeneity of the effects in the included studies, we first conducted Q statistics as demonstrated in the following formula,

$$Q = \sum\nolimits_{i=1}^{k} \frac{1}{V(T_i)}(T_i - T)^2,$$

where $\frac{1}{V(T_i)}$ is an inverse variance weight from the fixed-effect. Next, we used I^2 *tests* (Higgins et al., 2003) with the following formula:

$$I^2 = 100\% * \frac{[Q - (k - 1)]}{Q}.$$

We adopted random-effects models (Hedges & Vevea, 1998) for the overall effects in our meta-analysis for two reasons. First, there was evidence that the population effects estimated among the included studies were heterogeneous (as reported in the Results section). Second, we intended to generalize the findings of the overall effects to a broader population. As such, random-effects models were used for unconditional inferences (Hedges & Vevea, 1998). Further, we accounted for the dependence of effect sizes from the same study using robust variance estimation (RVE; Hedges et al., 2010).

Only papers provided enough information for effect size calculation can be included. However, some papers were included with missing data on the coded study characteristics. We thus opted to use listwise deletion (Roth, 1994) to handle study characteristics with missing data for moderator analysis. As a result, the total effect sizes for "intervention length" is 44 because five papers did not report the length of the intervention (involving five effect sizes). We conducted all statistical analyses in R studio (R Core Team, 2019) using the *metafor* package (Viechtbauer, 2010).

Statistics for the Moderator Analysis

We used ANOVA-like models and meta-regression for the moderator analyses to explain methodical heterogeneity of the effect sizes, in addition to the overall effects in this meta-analysis. Since these analyses only considered moderator-/within-group level of variance, we only expected within-group sampling error as the exclusive source of uncertainty within the chosen moderators. Essentially, the moderator is explaining why the effect sizes within the meta-analysis might differ, and the fixed-effects model adjusts for this variability by attributing it to the chosen moderator.

Hence, we opted for the adoption of a more stringent fixed-effects model in the context of conditional inference for the ANOVA-like moderator analysis (Hedges & Vevea, 1998). This decision takes into consideration the nuanced interpretability of each moderator. ANOVA-like models were applied to the following moderators (categorical independent variable): intervention length, education level, domain, publication type, the role of AI-powered virtual agents, AI technologies in AI-powered virtual agents, the modality of AI-powered virtual agents, the representation of AI-powered virtual agents, and learning environment. Meta-regression was applied to publication year (continuous independent variable), where the first appeared year in the included studies was coded as 0 and so forth (i.e., year 2002 was coded as “0,” year 2005 was coded as “3”).

Testing for publication bias is a standard practice in a sensible meta-analysis research. Publication bias can be a threat to the conclusion of a meta-analysis because the positive overall results of the included studies can be due to the file-drawer problem, of which negative results tended to be unpublished (Duval & Tweedie, 2000; Rosenthal, 1979; see also Scheel et al., 2021). We used several strategies to address the issues of publication bias in this meta-analysis: examining symmetry of the funnel plot, trim and fill analysis (Duval & Tweedie, 2000), and Egger’s regression test (Egger et al., 1997).

Results

RQ1: Overall Effects

From 22 included studies, we calculated 49 effects sizes, involving 4911 participants. The year of publication ranges from 2002 to 2021. The number of publications in the most recent years are 2021 ($n=7$), 2019 ($n=4$), 2018 ($n=9$). Meta-regression analysis suggested that publication year was not a statistically significant moderator for study heterogeneity ($Q_m(1)=0.79$, $\beta = 0.02$, $p=0.37$) (see Table 1 for the summary of results). Although all published in English, the included studies represented a diverse range of locations worldwide, including Australia ($n=1$), Belgium ($n=1$), Brazil ($n=3$), China ($n=2$), Germany ($n=10$), Greece ($n=2$), The Netherlands ($n=3$), Sweden ($n=1$), Taiwan ($n=2$), Turkey ($n=2$), and the USA ($n=2$).

Since there were studies that contributed multiple effect sizes with the same sample, we assumed dependency among effect sizes. We used RVE (Hedges et al., 2010) for the effect estimate by grouping effect sizes from the same authors in one cluster among the overall 22 clusters. We examined the effects of infusing AI technologies in virtual agents within computer-based simulations for learning. The study comparisons included in this meta-analysis had to demonstrate the effects of AI-powered virtual agents between groups. The results of the random effects model analysis suggested a medium overall effect size, $\overline{g} = 0.43$, $df=48$, $SE=0.08$, 95% C.I. [0.27, 0.59], $p<0.001$ (see Fig. 4. for the Forest plot).

There were three studies with the highest effect sizes. The three effect sizes were “Le and Wartschinski (2018) V” ($g=2.48$), “Le and Wartschinski (2018) VII”

Table 1 Moderator analysis

Type of analysis and the moderator	K	Mean (SE)	95% CI	Q_w
Overall	49	0.43 (0.08)	[0.26, 0.59]	
Intervention length (Q_B (3) = 11.44, p = 0.01)**				
≤ 30 min	13	0.49 (0.16)	[0.18, 0.80]**	27.03
Between 31 and 60 min	5	0.23 (0.26)	[−0.28, 0.74]	1.17
Between 61 and 119 min	8	0.12 (0.20)	[−0.28, 0.51]	17.14
≥ 120 min	18	0.64 (0.14)	[0.37, 0.91]***	142.48
Education level (Q_B (2) = 2.42, p = 0.30)				
Elementary	6	0.41 (0.24)	[−0.06, 0.88]	17.33
Middle/high school	5	0.51 (0.27)	[−0.02, 1.03]	27.28
College or adults	38	0.43 (0.10)	[0.25, 0.62]***	171.41
Domain (Q_B (4) = 1.92, p = 0.75)				
Math	5	0.51 (0.27)	[−0.02, 1.04]	29.14
Science	17	0.50 (0.15)	[0.21, 0.78]***	140.78
Medical studies	5	0.33 (0.26)	[−0.17, 0.84]	13.53
Soft skills	18	0.43 (0.14)	[0.15, 00.71]**	32.49
Language learning	4	0.25 (0.32)	[−0.37, 0.87]	0.57
Publication type (Q_B (2) = 7.20, p = 0.03)*				
Journals	41	0.40 (0.09)	[0.22, 0.57]***	175.97
Proceedings	3	0.68 (0.34)	[0.03, 1.34]*	23.74
Dissertation	5	0.64 (0.26)	[0.12, 1.16]*	4.58
The role of AI-powered virtual agents (Q_B(5) = 5.05, p = 0.41)				
Guidance/cues	6	0.32 (0.24)	[−0.16, 0.79]	3.35
Feedback	19	0.34 (0.14)	[0.08, 0.61]**	35.98
Hints	6	0.48 (0.24)	[0.01, 0.95]*	21.60
Social companion	1	0.14 (0.62)	[−1.07, 1.35]	0.00
Instruction	1	0.33 (0.63)	[−0.91, 1.57]	0.00
Mix	16	0.61 (0.15)	[0.31, 0.91]***	152.45
AI technologies (Q_B (3) = 9.54, p = 0.02)*				
Scripted AI	7	0.33 (0.11)	[0.13, 0.54]**	14.08
Rule-based AI	11	0.23 (0.06)	[0.12, 0.34]***	19.15
Module-based AI	10	0.50 (0.08)	[0.34, 0.66]***	5.42
NLP/ML	21	0.42 (0.05)	[0.32, 0.52]***	169.71
The modality of AI-powered virtual agents (Q_B (2) = 4.14, p = 0.13)				
Verbal	14	0.29 (0.15)	[−0.01, 0.59]	19.58
Text	17	0.59 (0.14)	[0.32, 0.86]***	150.24
Multi-modal	18	0.41 (0.14)	[0.14, 0.68]**	44.47
Representation (Q_B (3) = 12.67, p = 0.005)**				
Humanlike agent	35	0.35 (0.10)	[0.15, 0.54]***	86.47
Fictional agent	6	0.55 (0.23)	[0.10, 1.01]*	14.43
Mixed of humanlike and fictional agents	1	0.55 (0.54)	[−0.50, 1.61]	0.00
Humanlike agent with text-based interactions	7	0.78 (0.22)	[0.35, 1.21]***	104.87
Learning environment (Q_B (1) = 1.06, p = 0.30)				
Computer-based system	30	0.45 (0.11)	[0.24, 0.66]***	155.39
Immersive environment (e.g., VR/Simulation game)	19	0.41 (0.14)	[0.15, 0.68]**	62.00
Meta-regression	K	Coefficient (SE)	95% CI	Q_e
Publication year (Q_m(1) = 0.79, p = 0.37	49	0.02 (0.02)	[−0.03, 0.02]	216.32

***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$

($g = 1.98$), and "Gulz et al. (2011)" ($g = 1.81$). We checked each of those effect sizes. For "Le and Wartschinski (2018) V" and "Le and Wartschinski (2018) VII," these were the two outcome variables (i.e., testing the knowledge of students on "Regression to the Mean" and "Selection Task") in their study with the highest effect sizes, in comparison to other outcome variables. In "Gulz et al. (2011)," the authors presented between group (the treatment and the control group) difference (diff) between "low achieving" (diff = 0.003) "medium achieving" (diff = 0.22) and "high achieving" groups (diff = 0.77). It is possible that the high achieving students in their study contributed to the high effect size.

Nonetheless, through this process of checking the outliers, no anomalous findings were identified among these high effect sizes regarding the research design or measurement procedures. Importantly, with standard statistics procedures, we assessed publication bias among the effect sizes with a funnel plot, Egger's regression test (Egger et al., 1997), and trim and fill analysis (Duval & Tweedie, 2000). Egger's regression test ($z = 1.65$, $p = 0.10$) suggested no publication bias. Trim and fill analysis revealed that adding six effect sizes to the left can achieve symmetry which advised that it is unlikely that the effect size synthesized in this meta-analysis suffered from publication bias (for the funnel plot, see Fig. 5).

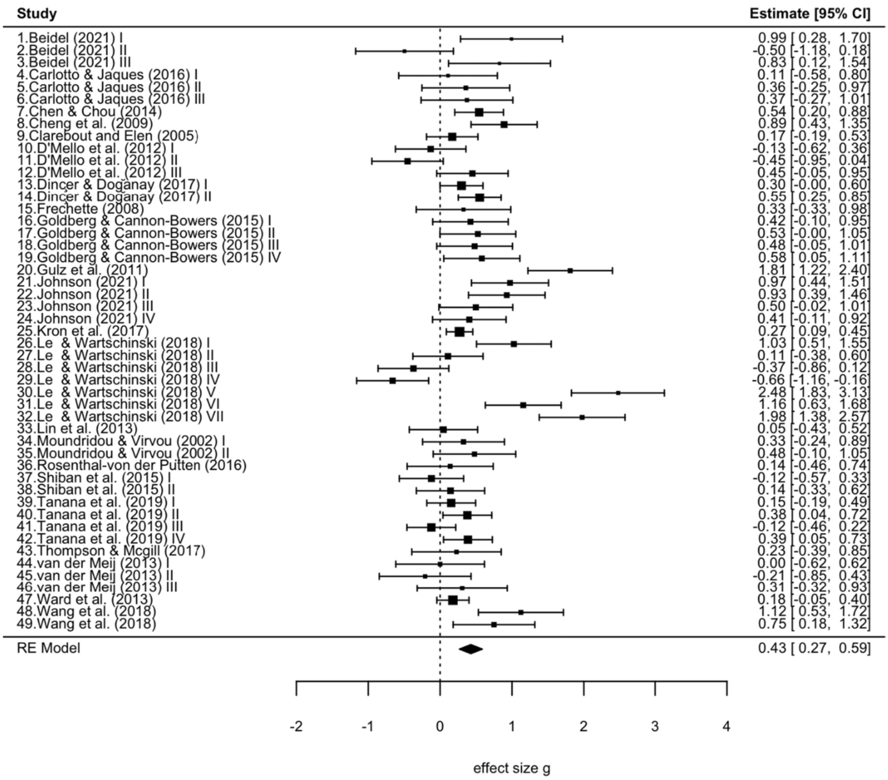

Fig. 4 Forest plot of the overall effect size g

The statistical homogeneity tests revealed that the included studies demonstrated heterogeneity as evidenced by I^2 test = 82.97% (according to Higgins et al. (2003), high heterogeneity is above 75%), $Q(48) = 211.39$, $p < 0.001$, and between-study variance of effect estimate $\tau^2 = 0.25(SE = 0.07)$.

RQ2: Moderator Analysis

We examined nine categorical moderators with ANOVA-like approach in this meta-analysis. A summary of the moderator analysis was shown in Table 1. We found that *four of nine categorical moderators* in the ANOVA-like analysis played a statistically significant role in explaining the heterogeneity observed in the overall effects: the intervention length, publication type, AI technologies, and the representation of the AI-powered virtual agents. *Publication year* in meta-regression analysis was not found to be statistically significant in explaining the heterogenous overall effects.

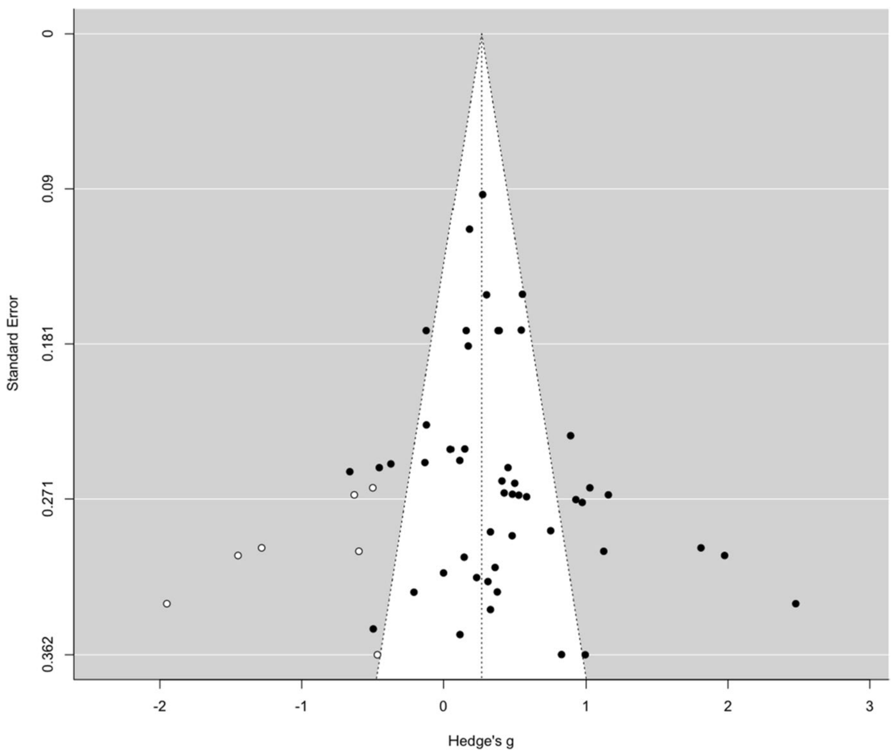

Fig. 5 Funnel plot

Intervention Length

We found that *intervention length* significantly influenced the variability of the effect sizes. Q_B $(3)=11.44$, $p=0.01$. Within-group variability was observed in two groups: *less or equal to 30 min*, $Q_W(12)=27.03$, $p<0.01$ and *equal to or over 120 min*, $Q_W(17)=142.48$, $p<0.001$. The within-group variability was not observed in the other two groups (i.e., *between 31 and 60 min* and *between 61 and 119 min*).

Education Level

Education level did not explain the variability of the effect size. Within-group variability was found for *College or adults group*, $Q_W(37)=171.41$, $p<0.001$, but not for Elementary, $Q_w(5)=17.33$, $p=0.06$ and Middle/high school, $Q_w(4)=27.28$, $p=0.08$. Overall, the explanatory power is equal for all education level groups.

Domain

Domain did not explain the variability of the effect size statistically, but within-group variability was found for *Science*, $Q_W(16)=140.78$, $p<0.001$ and *Soft skills* (e.g., negotiation, counseling, domain-generic problem solving), $Qw(17)=32.49$, $p=0.01$.

Publication Type

Effect-size heterogeneity was statistically significant for Publication type. With all three groups yielded significant within-group variability as well: *Journal papers*, $Q_W(40)=175.97$, $p<0.001$; *conference proceedings*, Q_W $(2)=23.74$, $p=0.04$; and *dissertation*, $Qw(4)=4.58$, $p=0.02$.

The Role of AI-Powered Virtual Agents

Different roles of the AI-powered virtual agents did not significantly vary in explaining the effect sizes in this meta-analysis. Nevertheless, Feedback ($Q_W(18)=35.98$, $p=0.01$), Hints ($Q_W(5)=21.60$, $p=0.05$), and Mix roles (e.g., guidance, inquiry-based cues, and feedback within one intervention) of AI-powered virtual agents ($Q_W(15)=152.45$, $p<0.001$) have significant within-group variability.

AI Technologies

AI technologies used in virtual agents in computer-based simulations for learning accounted for study heterogeneity (Q_B $(3)=9.54$, $p=0.02$). All four categories, Scripted AI ($Q_W(6)=14.08$, $p=0.002$), Rule-based AI ($Q_W(10)=19.15$, $p<0.001$), Module-based AI ($Q_W(9)=5.42$, $p<0.001$), and NLP/ML ($Q_W(20)=169.71$, $p<0.001$), showed a within-group variability. Module-based AI group demonstrated the highest effect size ($\overline{g} = 0.50$, SE = 0.08), followed by NLP/ML ($\overline{g} = 0.42$, SE = 0.05), Scripted AI ($\overline{g} = 0.33$, SE = 0.11), and Rule-based AI ($\overline{g} = 0.23$, SE = 0.06).

The Modality of AI-Powered Virtual Agents

We found that the heterogeneity to be explained by the modality of AI-powered virtual agents was non-significant (Q_B (2) = 4.14, $p = 0.13$). Significant with-in group effect-size heterogeneity was found in the text ($Q_W(16) = 150.24$, $p < 0.001$) and multimodal groups ($Q_W(17) = 44.47$, $p = 0.004$), but not in the verbal group ($Q_W(13) = 19.58$, $p = 0.06$).

The Representation of AI-Powered Virtual Agents

Representation is heterogeneous among the coded studies (Q_B (3) = 12.67, $p = 0.005$). Specifically, *Humanlike agent* ($Q_W(34) = 86.47$, $p < 0.001$) and *Humanlike agent (w/ text only)* ($Q_W(6) = 104.87$, $p < 0.001$) were significant at alpha level of 0.001, whereas *Fictional agent* ($Q_W(5) = 14.43$, $p = 0.02$) was significant at alpha level of 0.05.

Learning Environment

The amount of heterogeneity was not found to be significant in explaining the effect sizes by the *Learning environment* (Q_B (1) = 1.06, $p = 0.30$). Both web-based simulation ($Q_W(29) = 155.39$, $p < 0.001$) and VR/Simulation game ($Q_W(18) = 62.00$, $p = 0.002$) demonstrated significant within group heterogeneity.

Discussion

The meta-analysis results indicated that, in general, studies that used AI-powered virtual agents were associated with higher learning outcomes for learners ($\overline{g} = 0.43$, $p < 0.001$), in comparison to those that did not use such agents in computer-based simulations for learning. Our findings are consistent with the report of prior meta-analyses of virtual agents on their positive overall effects on learning (e.g., Castro-Alonso et al., 2021; Schroeder et al., 2013), while extending the contexts and providing an enhanced understanding of AI-powered virtual agent interventions in computer-based simulations for learning. AI and computer-based simulations for learning have great potential to transform education as they can provide personalized and adaptive learning experience for problem solving, knowledge construction and skills development, and deep learning (Dai & Ke, 2022). The current study findings address the imperative needs to understand the effects of using AI-powered virtual agents in computer-based simulations for learning.

In addition to the consistency on the directions of the effects, the magnitude of the overall effect size in our meta-analysis is larger than the overall effect sizes reported in the previous related meta-analyses (e.g., Castro-Alonso et al., 2021, $\overline{g} = 0.20$;

Schroeder et al., 2013, $\overline{g} = 0.19$), indicating a growing positive trend in the effectiveness of virtual agents, especially when the agents are driven by more advanced AI technologies that improve adaptivity and personalization. Nonetheless, there is still a scarcity of rigorous experimental studies that compare AI-powered virtual agents with non-AI-powered virtual agent conditions. We had to exclude multiple studies due to ineligible research designs, interventions, or research variables. Researchers should continue empirical investigations on virtual agents powered by AI in computer-based simulations for learning, especially given the rapid exponential advancements in AI technologies (Chen et al., 2020; Russell & Norvig, 2021). While the current evidence suggests the promising benefits of AI-powered virtual agents in facilitating learning, the connections between AI and human learning, and the broader impacts in education are largely under-explored, necessitating an urgent investment in further research (Dai & Ke, 2022; Dai et al., 2024; Williamson & Eynon, 2020).

Apart from providing empirical evidence on the positive effects of AI-powered virtual agents in computer-based simulations for learning, the current meta-analysis results also shed light on how the effects are moderated by the agent design features, AI technologies, and study characteristics. We discuss the moderator results in the following sections.

Intervention Length

We found that intervention length significantly explains the effect-size heterogeneity. Specifically, contrary to conventional belief of "the more, the better" regarding the dosage of educational technology, AI-powered virtual agents may demonstrate greater efficacy when used for interventions lasting 1) less or equal to 30 min, and 2) equal or larger than 2 h in computer-based simulations for learning. Interventions falling outside these specified duration categories may exhibit comparatively less benefit. This observation can be explained from two perspectives. *First*, memory span from a cognitive information processing thesis posits that an individual has a limited working memory capacity for new information or knowledge processing (Miller, 1956). Learners with a low-working memory span may suffer when the learning duration is longer than their capacity for information processing, as predicted by cognitive load theory (see Sweller et al., 2019). The content learned beyond this capacity can inhibit their knowledge or long-term memory retrieval (Kane & Engle, 2000) and hence diminish learning effectiveness. Nye et al. (2014) also reported that learning gains seem to be higher with short and organized interventions. On the other hand, if the learning duration is prolonged and learners have enough time and sufficient attention for information encoding and rehearsal (Atkinson & Shiffrin, 1968) for memory storage, better learning outcomes can be expected. *Second*, isomorphic to the time argument for information processing, we theorize that the time needed to achieve deep learning from a constructive learning stance serves as another contributing factor. Similar to guided discovery, learning constructively with AI-powered virtual agents may take learners more time to internalize the knowledge (see Mayer, 2004; Gorbunova et al., 2023).

In relation to the intervention length and learning theories, we conjecture that different roles of AI-powered virtual agents may also contribute to this result. For the "less or equal to thirty minutes" group, the agents were mainly used to provide hints or feedback; hence, the learners are able to adjust or revise their current practice or understanding to achieve satisfactory learning outcomes within a short time span. In comparison, the agents used in the "equal or larger than 2 h" group may facilitate inquiry-based or knowledge-constructive learning so that it enables the learners to learn deeply. We encourage future research to explore intervention length and learning characteristics by manipulating variables to deliver more robust results.

Education Level

We did not find significant explaining power by different education levels. Our results were consistent with a recent meta-analysis (Castro-Alonso et al., 2021) but in contrast to Schroeder et al.'s (2013). Similar to Castro-Alonso et al. (2021), we found that AI-powered virtual agents are equally beneficial across education levels, regardless of learners' stage of education and development. This finding is inconsistent with a previous argument that "in formal education, pedagogical agents seem to be more effective for younger learners than for older learners" (Johnson & Lester, 2016, p. 30). These inconsistent results suggest that further research is needed to explore the impact of education level and learners' stage of development on the outcomes of computer-based simulations for learning with AI-powered virtual agents. Specifically, it was unclear what design characteristics and contextual factors contribute to the differences in the use of AI-powered virtual agents between young and adult learners. Aside from the perspective of education level, some researchers approached this issue by stating that it is when the learner-agent interactions occurred within learners' zone of proximal development (Vygotsky, 1978) that it is beneficial (Dai & Ke, 2022; Graesser et al., 2008). Although we provided preliminary evidence suggesting the universal benefits of using AI-powered virtual agents in computer-based simulations for learning *for all developmental stages* in formal education programs, the mixed results in the literature point to the importance for future investigation on education level as a moderator of AI-powered virtual agents in computer-based simulations for learning.

Domain

Domain is another disputable moderator in the literature. While Schroeder et al. (2013) suggested that virtual agents are more effective for math and science learning than humanities, Johnson and Lester (2016) documented a divergent case that foreign language learning and intercultural skills training can be particularly effective when using virtual agents (Johnson & Lester, 2016). Our findings support Castro-Alonso et al.'s (2021) meta-analysis findings, in which the authors classified domain into STEM and non-STEM studies. The authors (Castro-Alonso et al., 2021) found that domain does not moderate effects of virtual agents. However, at a more granular level, discipline was a significant moderating factor. In particular, *biology*, *computing*, and *English* positively moderate the effects, whereas *history* negatively moderates the effects. The moderating effects of math and language were not significant. Overall, we found that AI-powered virtual agents can be implemented for all domains. However, this finding should be interpreted with caution. First, the classifications of our domains were rather constrained by the included studies; we did not classify science into more specific disciplines, such as physics or biology. We also did not include other domains of humanities, such as literature or history. Second, some domains have a paucity of studies on AI-powered virtual agents in computer-based simulations for learning, thus only limited studies of certain domains were included.

Moderators of AI-Powered Virtual Agent

The Role-Specific Functionality of AI-Powered Virtual Agents

The role-specific functionality of AI-powered virtual agents for learning has been a keen issue for researchers and practitioners (e.g., Heidig & Clarebout, 2011; Kim & Baylor, 2016). Earlier studies emphasized the role of virtual agents for cognitive information processing (Kim & Baylor, 2016) and guidance (Johnson, 2003), whereas there are growing interests in social roles played by virtual agents (Sinatra et al., 2021). The current meta-analysis findings revealed that there was no significant difference between different roles played by the AI-powered virtual agents in explaining the overall effects. That is, regardless of what roles the AI-powered virtual agents played, they are equally effective. Although narrative review has discussed the role of virtual agents for learning (Kim & Baylor, 2016), empirical evidence on the effects of various roles of AI-powered virtual agents in computer-based simulations for learning has been lacking. We promote inclusiveness and diversity for designing the role-specific functionality of AI-powered virtual agents. Specifically, AI-powered virtual agents can be designed as information providers (Schroeder & Gotch, 2015), experiential learning and discourse-rich facilitators (Woolf et al., 2013), and collaborative problem-solving companions (Graesser, 2016). The classification of roles has been disparate and hence making meta-analytic comparisons difficult. Our findings suggest the universal benefits of the role-specific functionality of AI-powered virtual agents. We note, however, that the limitations of a small number of studies in some categories may restrict the generalizability. More empirical studies are needed for the community to better understand different roles of the AI-powered virtual agents in computer-based simulations for learning.

AI Technologies in AI-Powered Virtual Agents

The findings on AI technologies were aligned with prior studies suggesting that AI in education should focus on facilitating learning in addition to pursuing machine learning accuracy (Dai & Ke, 2022; Dai et al., 2023), thus module-based AI revealed the highest explanatory power because module-based AI features student or knowledge modeling with predictive statistics for learner-machine interactions that can promote learning. Peng and Wang (2022) suggested that a higher degree of AI is better for personalization and adaptivity, whereas Rosé et al. (2019) explicated that interpretability, explainability, and actionability are crucial along with AI accuracy.

Our findings suggested that AI technologies overall explain the effects of computer-based simulations for learning, with module-based AI ($\overline{g} = 0.50$, $p < 0.001$) achieving the highest average effect sizes, followed by NLP/ML ($\overline{g} = 0.42$, $p < 0.001$). This implies that the current AI-powered virtual agents can benefit more from the design that incorporates the modeling of learners' knowledge, affective states, or other learner and contextual characteristics. Module-based AI is powerful because it emphasizes both learners and AI accuracy, which is essential for adaptive and personalized learning. In other words, the models in module-based AI that are human-centered and technology-supported have the capacity to respond to learners with heightened accuracy and effectiveness, fostering a pedagogically sound learning experience.

Module-based AI is also less complicated and more transparent to the designers, researchers, and practitioners. Notably, other AI technologies, such as large language models and generative AI, are almost equally effective. However, one may argue that the "black box" nature of these large language models is of concerns. We also maintain that the exploration of transparency of the "black box" is necessary. Taking the benefits of module-based AI, applications using NLP/ML can address the "black box" issues by focusing on the control of local input training data and testing of the applications with a particular interest—to enhance learning. Given our results, we suggest that when designing and developing AI-powered virtual agents for computer-based simulations for learning, it is imperative to consider learners and other learning-related models. We echo Johnson and Lester (2016) and express that AI-powered virtual agents are one of the integral components contributing to the design of effective learning environments. The key is to design an effective learning environment in the service of learners. With this goal in mind, future research should explore how large language models and generative AI can be effectively trained and implemented to deliver personalized learning with greater granularity, while considering the conditions under which these processes and algorithms are the most effective.

The Modality of AI-Powered Virtual Agents

The results of this meta-analysis suggested no significant difference between different modalities of AI-powered virtual agents in computer-based simulations for learning in explaining the heterogeneity of the overall effects. Specifically, text, voice, or multi-modal interactions can result in comparable benefits. This finding is similar to Kim's (2005). In contrast, the modality effect in multimedia design for learning from a cognitive perspective argued that information exhibition with audiovisual dual channels (voice and graphic presentations) can be easier to understand than the ones presented with illustrated text (i.e., text and graphic presentations) (Castro-Alonso & Sweller, 2022; Noetel et al., 2022; Reinwein, 2012). In Johnson and Lester's (2016) narrative review, the authors also believe that virtual agents can facilitate more robust outcomes when they "speak rather than communicate with text" (p. 31). In computer-based simulations for learning, learners can also neglect the contextual cues when focusing too much on text-interactions (Dai, 2023, 2024). However, the modality effect was unsupported by the findings in this meta-analysis focusing on AI-powered virtual agents in computer-based simulations for learning.

To elaborate, we note that the modality effect in AI-powered virtual agents is inherently different from the one in multimedia design for learning. For example, in multimedia design for learning, using text can be redundant in explaining self-explanatory graphics that contain learning materials (Sweller, 1994). But in computer-based simulations for learning with AI-powered virtual agents, using text can be beneficial when the conversation-oriented materials for learning are recorded in the intervention. When learners explore and experiment in the intervention, their attention can be split, their cognitive resources can be consumed. The recorded text-based conversational materials can be revisited in the intervention so that the learners can re-encode and rehearse the content in their cognitive system as well as supporting their decision making in a later scenario—especially when the conversations and discourses get richer over the course of the learning stages. Further, using text in conversational virtual agents can possibly result in similar effects as in captioning the interactions. Noetel et al. (2022) found captioning to be effective in learning with second-language videos. These discussions can find congruence with Nye et al.'s (2014) viewpoint stating that "whether the learning content is being consumed by the learners" is more important than the modality for learning with virtual agents in computer-based simulations for learning.

The Representation of AI-Powered Virtual Agents

The representation of AI-powered virtual agents concerns the perceptions of the learners toward the agents, that is, whether the agents are humanlike or character-figured can influence the social fidelity and trust of the learning interactions (Kim & Baylor, 2016). Given limited studies available for inclusion, the moderating effects of representation of AI-powered virtual agents have been ambiguous in prior meta-analysis (e.g., Heidig & Clarebout, 2011); more recent meta-analysis on virtual agents has not examined this moderator (Castro-Alonso et al., 2021). Our meta-analysis found a significant difference of representation in explaining the overall effects of AI-powered virtual agents in computer-based simulations for learning—with *humanlike agent with text-based interactions* yielding the highest effect size ($\overline{g} = 0.78$), followed by *fictional agent* ($\overline{g} = 0.55$), and *humanlike agent* representation ($\overline{g} = 0.35$). We conjectured that *humanlike agent with text-based interactions* shown advantages over *fictional agent* due to its social fidelity (Kim & Baylor, 2016; Sinatra et al., 2021); but our findings align, in part, with the explanations in Castro-Alonso et al.'s (2021). They argued that more complex agents may lead to cognitive overload thus "2D, cartoonish, or simpler appearance would be more effective" (p. 1007). Our study adds to the literature suggesting that text-based interactions without excessive embellishment or information (e.g., embodied human characters) can be beneficial. We also want to elucidate that it is possible that limited authenticity in humanlike gesture or representation due to the current technological capabilities may reduce the plausibility and naturalistic interactions in the *humanlike agent* conditions. As a result, it may make the interventions less effective. With improved technologies in the future, humanlike conditions have the potentials to be effective, if designed well. Nonetheless, generally speaking, all treatment conditions in the representation of AI-powered virtual agents contribute to small to medium positive effect sizes for learning.

Limitations

There were several limitations in this meta-analysis. First, due to reporting styles (i.e., the literature provided insufficient information for effect size calculation) and research designs in the literature, we were unable to identify and include a large number of eligible studies. As a result, our categories in the moderators were constrained by this factor. Some within group categories have small study samples. The robustness of the findings from the moderator analyses was thus influenced. Cautions are needed when using the findings in this meta-analysis.

Second, motivation, engagement, and self-efficacy have been the frequently examined variables that suggest possible beneficial effects from the use of AI-powered virtual agents (e.g., Heidig & Clarebout, 2011; Lane et al., 2013; Roll & Wylie, 2016). However, our strong focus on learning and the difficulties of study identification have convinced us to leave out motivation, engagement, and self-efficacy in this meta-analysis. Future research could include motivation, engagement, and self-efficacy as an outcome variable to determine AI-powered virtual agents' effects on motivation, engagement, and self-efficacy in computer-based simulations for learning.

Third, due to challenges in identifying related studies, we followed Merchant et al.'s (2014) classification of computer-based simulation and included simulation games in the "learning environment" moderator. However, with the advancement of technology, more studies may have been conducted in the setting of virtual reality alone, or more broadly, extended reality. Future research can distinguish between virtual reality and simulation games for more nuanced results.

Fourth, despite the affirmation that our findings in this meta-analysis were unlikely to observe publication bias (i.e., we included non-peer-reviewed studies; our statistical analyses also shown no evidence of publication bias), we are still aware that studies with significant results were more likely to be published and therefore can potentially impact the findings of our included studies. Further, given our research purposes (i.e., investigating the effects of the AI-powered virtual agent in computer-based simulations for learning with a robust experimental design) and the available literature to date, the number of the studies included was limited. While there is no consensus on the number of studies to be included for sufficient statistical power, the number of the studies included should be considered when interpreting the results in this meta-analysis. Finally, when the results of p values were equal to an alpha level of 0.05, we did not assume a more conservative stance, thus the decision of Type I error level (e.g., $\alpha = 0.05$) should also be considered when interpreting the significant results.

Conclusion

In this meta-analysis, we add to and expand the current literature on several aspects. First, we examined the effects of virtual agents focusing on the ones driven by AI. We reinforced the notion that AI-powered virtual agents are effective in computer-based simulations for learning. But a lack of *eligible* studies on this topic is evident. Researchers and practitioners may increase investments in this area. Given that virtual agents are the most effective with module-based AI being integrated, the design and development of transdisciplinary models and technologies for AI-powered virtual agents in computer-based simulations for learning is suggested. Indeed, to achieve cohesive learning outcomes for students, cognitive, psychological, pedagogical, and machine learning models must work coherently as a single architecture (e.g., Rickel, 2001). Although the advancement in NLP/ML can overcome the limitations of inauthentic interactions with AI virtual agents, the core focus should be on learning design with an integrative approach.

Second, our study contributed to the mixed results in the literature by exploring alternative categories for the moderators. To elaborate, education researchers and practitioners should consider adopting AI-powered virtual agents in computer-based simulations for learning with their purposes and guiding principles and considerations for learning design due to the fact that the intervention can be effective on the two ends of the spectrum—≤ 30 min and ≥ 120 min. We suggest future research to empirically validate whether the role-specific functionality interacts with the effects of intervention length with pedagogically sound learning designs. In addition, we echo a recent meta-analysis (Castro-Alonso et al., 2021) in that simpler representations of the AI-powered virtual agents can be more effective, but we suggest that adding humanlike social fidelity can increase their effectiveness. That is, it is beneficial for the learners to interact with text-based virtual agents with humanlike names, persona, and character design. We maintain that the possible benefits can be due to, first, discourse-rich, meaning-making processes; and second, the advantages that text-based interactions were easier to be recorded and revisited by the learners for decision making. However, we also observe that voice, multimodal and affective computing technologies are still growing. Unnatural gestures, behaviors, as well as unapparent gazes and lip movements, may impose limitations on the findings within our included studies. We propose continued research on the design and development of humanlike AI virtual agents for naturalistic interactions.

Third, our findings imply that AI-powered virtual agents are equally effective considering education level, domain, the role-specific functionality, and the modality. We take an inclusive and diverse stance on these variables, suggesting that AI-powered virtual agents can be used in multiple settings to induce effective learning. For instance, AI-powered virtual agents can be used for K-12 and postsecondary students as well as adult learners. Genuinely, as AI is permeating and reshaping our everyday lives, it should be beneficial for learners across ages for professional development and lifelong learning (Woolf et al., 2013). We also contribute to the conflicting viewpoints in the literature by promoting the multi-role functionality of AI virtual agents. Again, one should consider their guiding principles and purposes for learning design before they can decide the role-specific functionality of AI-powered virtual agents. Regardless, the effectiveness should be universal with carefully designed interventions. Finally, for modality, our results suggested that education researchers and practitioners should focus on learning experience design rather than the modality of the AI-powered virtual agents. Depending on the individual contexts and settings, one can find it is more useful to design agents with voice interactions, while others believe that text-based or multimodal agents can better serve their learners. Our insignificant finding suggests that designers should work on context-dependent and needs-based AI-powered virtual agents in terms of modality without worrying which modality is more effective. Overall, this meta-analysis provides systematic evidence and existing promises for the adoption of AI-powered virtual agents in computer-based simulations for learning. Education researchers, practitioners, and designers can refer to our findings in this meta-analysis (i.e., the overall effects and the moderators analyzed) carefully to make evidence-based, informed design decisions that are best-suited to their respective contexts.

Acknowledgements The work is most closely related to the National Science Foundation grant 2110777. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Declarations

Conflict of Interest The authors declare no conflict of interest.

References

Studies included in the meta-analysis are denoted by an asterisk mark in the references.

Al-Elq, A. H. (2010). Simulation-based medical teaching and learning. *Journal of family and Community Medicine, 17*(1), 35–40. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3195067/pdf/JFCM-17-35.pdf

Alinier, G. (2007). A typology of educationally focused medical simulation tools. *Medical Teacher, 29*(8), e243–e250. https://doi.org/10.1080/01421590701551185

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25*(2), 450–457. https://doi.org/10.1016/j.chb.2008.10.008

*Beidel, D. C., Tuerk, P. W., Spitalnick, J., Bowers, C. A., & Morrison, K. (2021). Treating childhood social anxiety disorder with virtual environments and serious games: a randomized trial. *Behavior Therapy, 52*(6), 1351-1363. https://doi.org/10.1016/j.beth.2021.03.003

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on your-self, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Byrne, J., Heavey, C., & Byrne, P. J. (2010). A review of Web-based simulation and supporting tools. *Simulation Modelling Practice and Theory, 18*(3), 253–276. https://doi.org/10.1016/j.simpat.2009.09.013

*Carlotto, T., & Jaques, P. A. (2016). The effects of animated pedagogical agents in an English-as-a-foreign-language learning environment. *International Journal of Human-Computer Studies, 95*, 15-26. https://doi.org/10.1016/j.ijhcs.2016.06.001

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*(8), 723–733. https://doi.org/10.1177/016146816306400801

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine, 22*(4), 67–83. https://doi.org/10.1609/aimag.v22i4.1593

Castro-Alonso, J. C., & Sweller, J. (2022). The modality principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 261–267). Cambridge University Press. https://doi.org/10.1017/9781108894333.026

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review, 33*(3), 989–1015. https://doi.org/10.1007/s10648-020-09587-1

*Chen, C.-H., & Chou, M.-H. (2015). Enhancing middle school students' scientific learning and motivation through agent-based learning. *Journal of Computer Assisted Learning, 31*(5), 481-492. https://doi.org/10.1111/jcal.12094

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

*Cheng, Y. M., Chen, L. S., Huang, H. C., Weng, S. F., Chen, Y. G., & Lin, C. H. (2009). Building a general purpose pedagogical agent in a web-based multimedia clinical simulation system for medical education. *IEEE Transactions on Learning Technologies*, *2*(3), 216-225. https://doi.org/10.1109/TLT.2009.18

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499–541. https://doi.org/10.3102/0034654320933544

*Clarebout, G., & Elen, J. (2006). Open learning environments and the impact of a pedagogical agent. *Journal of Educational Computing Research*, 35(3), 211-226. https://doi.org/10.2190/3UL1-4756-H837-2704

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

* D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-computer studies, 70*(5), 377-398. https://doi.org/10.1016/j.ijhcs.2012.01.004

Dai, C.-P. (2023). *Enhancing learning achievements and self-efficacy for preservice teachers using model-based support in simulation-based learning with artificial intelligence-powered virtual agents*. Doctoral dissertation. Florida State University.

Dai, C.-P. (2024). Applying machine learning to augment the design and assessment of immersive learning experience. In M. S. Khine (Ed.), *Machine learning in educational sciences: Approaches, applications and advances.* Springer. https://doi.org/10.1007/978-981-99-9379-6_12

Dai, C.-P., & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers & Education: Artificial Intelligence, 3*, 100087. https://doi.org/10.1016/j.caeai.2022.100087

Dai, Z., Ke, F., Dai, C.-P., Pachman, M., & Yuan, X. (2021). Role-play in virtual reality: A teaching training design case using OpenSimulator. In G. Akcayir & C. D. Epp (Eds.), *Designing, deploying, and evaluating virtual and augmented reality in education* (pp. 143–163). IGI Global. https://doi.org/10.4018/978-1-7998-5043-4.ch007

Dai, C.-P., Ke, F., Dai, Z., & Pachman, M. (2023). Improving teaching practices via virtual reality-supported simulation-based learning: Scenario design and the duration of implementation. *British Journal of Educational Technology, 54*(4), 836–856. https://doi.org/10.1111/bjet.13296

Dai, C-P., Ke, F., Zhang, N., Barrett, A., West, L., Bhowmik, S., Southerland, S. A., & Yuan, X. (2024). Designing conversational agents to support student teacher learning in virtual reality simulation: A case study. In *Proceedings of The ACM CHI conference on Human Factors in Computing Systems Extended Abstracts (ACM CHI EA '24), Honolulu*. https://doi.org/10.1145/3613905.3637145

Daradoumis, T., & Arguedas, M. (2020). Cultivating students' reflective learning in metacognitive activities through an affective pedagogical agent. *Educational Technology & Society*, 23(2), 19–31. https://www.jstor.org/stable/26921131

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review, 24*, 193–209. https://doi.org/10.1016/j.edurev.2018.05.002

de Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179–201. https://doi.org/10.3102/00346543068002179

Decker, S., Sportsman, S., Puetz, I., & Billings, I. (2008). The evolution of simulation and its contribution to competency. *The Journal of Continuing Education in Nursing, 39*(2), 74–80. https://doi.org/10.3928/00220124-20080201-06

Deeks, J. J., Higgins, J. P., Altman, D. G., Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. In J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 241–284). John Wiley & Sons. https://doi.org/10.1002/9781119536604.ch10

*Dinçer, S., & Doğanay, A. (2017). The effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load. *Computers & Education*, *111*, 74-100. https://doi.org/10.1016/j.compedu.2017.04.005

Domagk, S. (2010). Do pedagogical agents facilitate learner motivation and learning outcomes?: The role of the appeal of agent's appearance and voice. *Journal of Media Psychology: Theories, Methods, and Applications, 22*(2), 84–97. https://doi.org/10.1027/1864-1105/a000011

Duval, S., & Tweedie, R. (2000). A nonparametric 'trim and fill' method of assessing publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867–872. https://doi.org/10.1016/0140-6736(91)90201-Y

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

Elkot, M. A. (2019). Embedding adaptation levels within intelligent tutoring systems for developing programming skills and improving learning efficiency. *International Journal of Advanced Computer Science and Applications, 10*(12), 82–87. https://doi.org/10.14569/IJACSA.2019.0101211

*Frechette, M. C. (2008). Animated pedagogical agents: How the presence and nonverbal communication of a virtual instructor affect perceptions and learning outcomes in a computer-based environment about basic physics concepts [Unpublished doctoral dissertation]. The University of New Mexico.

Gamer, M., Lemon, J., & Fellows, I. (2012). irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. Retrieved from https://cran.r-project.org/web/packages/irr/index.html

Gibbons, A. S. (2020). What is instructional strategy? Seeking hidden dimensions. *Educational Technology Research and Development, 68*(6), 2799–2815. https://doi.org/10.1007/s11423-020-09820-2

*Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior, 52*, 1-11. https://doi.org/10.1016/j.chb.2015.05.008

Gorbunova, A., van Merrienboer, J.J.G. & Costley, J. (2023). Are inductive teaching methods compatible with cognitive load theory?. *Educational Psychology Review, 35,* Article 111. https://doi.org/10.1007/s10648-023-09828-z

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., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes, 45*(4–5), 298–322. https://doi.org/10.1080/01638530802145395

*Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a teachable agent with a social conversation module–effects on student experiences and learning. In *International conference on artificial intelligence in education* (pp. 106–114). Springer, Berlin, Heidelberg. https://dl.acm.org/doi/10.5555/2026506.2026524

Guo, Y. R., & Goh, D. H. L. (2015). Affect in embodied pedagogical agents: Meta-analytic review. *Journal of Educational Computing Research, 53*(1), 124–149. https://doi.org/10.1177/0735633115588774

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

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486–504.

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

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27–54. https://doi.org/10.1016/j.edurev.2010.07.004

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

Higgins, J. P., Li, T., & Deeks, J. J. (2019). Choosing effect measures and computing estimates of effect. In J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 143–176). John Wiley & Sons. https://doi.org/10.1002/9781119536604.ch6

Hwang, G. J., Xie, H., Wah, B. W., & Gašević. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*(4), 1036–1049. https://doi.org/10.1037/a0032580

Johnson, W. L. (2003). Interaction tactics for socially intelligent pedagogical agents. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (pp. 251–253). https://doi.org/10.1145/604045.604090

*Johnson, E. (2021). An intelligent tutoring system's approach for negotiation training. [Unpublished doctoral dissertation]. University of Southern California.

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

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 336–358. https://doi.org/10.1037/0278-7393.26.2.336

Ke, F., Dai, Z., Dai, C.-P., Pachman, M., Chaulagain, R., & Yuan, X. (2020). Designing virtual agents for simulation-based learning in virtual reality. In R. Zheng (Ed.), *Cognitive and affective perspectives on immersive technology in education* (pp. 151–170). IGI Global. https://doi.org/10.4018/978-1-7998-3250-8.ch008

Kim, S. (2005). The effect of multimedia design elements on learning outcomes in pedagogical agent research: a meta-analysis. In M. Grandbastien (Ed.), *Young Research Track Proceedings of 12th International Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 69–76).

Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Kolb, D. A. (1984). *Experiential learning experience as a source of learning and development*. Prentice Hall.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3–34. https://doi.org/10.1007/BF00155578

*Kron, F. W., Fetters, M. D., Scerbo, M. W., White, C. B., Lypson, M. L., Padilla, M. A., ... & Becker, D. M. (2017). Using a computer simulation for teaching communication skills: A blinded multisite mixed methods randomized controlled trial. *Patient education and counseling, 100*(4), 748-759. https://doi.org/10.1016/j.pec.2016.10.024

Kulik, J. A., & Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lane, H. C., Cahill, C., Foutz, S., Auerbach, D., Noren, D., Lussenhop, C., & Swartout, W. (2013). The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In *Proceedings of International conference on artificial intelligence in education* (pp. 309–318). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_32

*Le, N. T., & Wartschinski, L. (2018). A cognitive assistant for improving human reasoning skills. *International Journal of Human-Computer Studies, 117*, 45-54. https://doi.org/10.1016/j.ijhcs.2018.02.005

*Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education, 67*, 239–249. https://doi.org/10.1016/j.compedu.2013.04.017

Lippert, A., Shubeck, K., Morgan, B., Hampton, A., & Graesser, A. (2020). Multiple agent designs in conversational intelligent tutoring systems. *Technology, Knowledge and Learning, 25*, 443–463. https://doi.org/10.1007/s10758-019-09431-8

Luck, M., & Aylett, R. (2000). Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied Artificial Intelligence, 14*(1), 3–32. https://doi.org/10.1080/088395100117142

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International Conference on Information and Education Technology* (pp. 111–119). https://doi.org/10.1145/3323771.3323824

Mascarenhas, S., Guimarães, M., Prada, R., Dias, J., Santos, P. A., Star, K., Hirsh, B., Spice, E., & Kommeren, R. (2018). A virtual agent toolkit for serious games developers. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)* (pp. 1–7). IEEE. https://doi.org/10.1109/CIG.2018.8490399

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine, 27*(4), 12–14. https://doi.org/10.1609/aimag.v27i4.1904

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158

*Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*(3), 253–261. https://doi.org/10.1046/j.0266-4909.2001.00237.x

Noetel, M., Griffith, S., Delaney, O., Harris, N. R., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2022). Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Review of Educational Research, 92*(3), 413–454. https://doi.org/10.3102/00346543211052329

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5

Pavlik Jr., P. I., Brawner, K., Olney, A., & Mitrovic, A. (2013). A review of student models used in intelligent tutoring systems. In R. A. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems: Learner modeling* (Vol. 1, pp. 39–67). U.S. Army Research Laboratory. Retrieved from: https://gifttutoring.org/documents/42

Peng, T. H., & Wang, T. H. (2022). Developing an analysis framework for studies on pedagogical agent in an e-Learning environment. *Journal of Educational Computing Research, 6*(3). https://doi.org/10.1177/07356331211041701

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming, 41*(1), 72–93. https://doi.org/10.1177/1046878109355684
R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/
Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research, 41*(1), 1–32. https://doi.org/10.1007/s10936-011-9180-4
Richey, R. C., Klein, J. D., & Tracey, M. W. (2011). *The instructional design knowledge base: Theory, research, and practice*. Routledge.
Rickel, J. (2001). Intelligent virtual agents for education and training: Opportunities and challenges. In de Antonio, A., Aylett, R., Ballin, D. (Eds). *Intelligent Virtual Agents. IVA 2001. Lecture Notes in Computer Science, vol 2190* (pp. 15–22). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44812-8_2
Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3
Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology, 50*(6), 2943–2958. https://doi.org/10.1111/bjet.12858
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin, 86*(3).
*Rosenthal-von der Pütten, A. M., Straßmann, C., & Krämer, N. C. (2016, September). Robots or agents–neither helps you more or less during second language acquisition. In Traum, D., Swartout, W., Khooshabeh, P., Kopp, S., Scherer, S., Leuski, A. (Eds). *Proceedings of International conference on intelligent virtual agents* (pp. 256–268). Springer, Cham. https://doi.org/10.1007/978-3-319-47665-0_23
Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560. https://doi.org/10.1111/j.1744-6570.1994.tb01736.x
Russell, S. J., & Norvig, P. (Ed.). (2021). Artificial intelligence: A modern approach (4th edition). Pearson.
Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science, 4*(2), 1–12. https://doi.org/10.1177/25152459211007467
Schroeder, N. L., & Gotch, C. M. (2015). Persisting issues in pedagogical agent research. *Journal of Educational Computing Research, 53*(2), 183–204. https://doi.org/10.1177/0735633115597625
Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1–39. https://doi.org/10.2190/EC.49.1.a
Schweppe, J., & Rummer, R. (2014). Attention, working memory, and long-term memory in multimedia learning: An integrated perspective based on process models of working memory. *Educational Psychology Review, 26*(2), 285–306. https://doi.org/10.1007/s10648-013-9242-2
Seshia, S. A., Sadigh, D., & Sastry, S. S. (2022). Toward verified artificial intelligence. *Communications of the ACM, 65*(7), 46–55. https://doi.org/10.1145/3503914
*Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., & Mühlberger, A. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior, 49*, 5–11. https://doi.org/10.1016/j.chb.2015.01.077
Sinatra, A. M., Pollard, K. A., Files, B. T., Oiknine, A. H., Ericson, M., & Khooshabeh, P. (2021). Social fidelity in virtual agents: Impacts on presence and learning. *Computers in Human Behavior, 114*, 106562. https://doi.org/10.1016/j.chb.2020.106562
Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x\
Spronck, P., Ponsen, M., Sprinkhuizen-Kuyper, I., & Postma, E. (2006). Adaptive game AI with dynamic scripting. *Machine Learning, 63*(3), 217–248. https://doi.org/10.1007/s10994-006-6205-6
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

*Tanana, M. J., Soma, C. S., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2019). Development and evaluation of ClientBot: Patient-like conversational agent to train basic counseling skills. *Journal of Medical Internet Research, 21*(7), e12529. https://doi.org/10.2196/12529

*Thompson, N., & McGill, T. J. (2017). Genetics with Jean: The design, development and evaluation of an affective tutoring system. *Educational Technology Research and Development, 65*(2), 279–299. https://doi.org/10.1007/s11423-016-9470-5

Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning, 35*(3), 407–420. https://doi.org/10.1111/jcal.12347

Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining simulation fidelity for healthcare education. *Simulation & Gaming, 46*(2), 159–174. https://doi.org/10.1177/1046878115576103

*van der Meij, H. (2013). Do pedagogical agents enhance software training?. *Human–Computer Interaction, 28*(6), 518–547. https://doi.org/10.1080/07370024.2013.789348

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

Veletsianos, G., & Russell, G. S. (2013). What do learners and pedagogical agents discuss when given opportunities for open-ended dialogue? *Journal of Educational Computing Research, 48*(3), 381–401. https://doi.org/10.2190/EC.48.3.e

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

Vygotsky, L.S. (1978). Mind in society: Development of higher psychological processes. Harvard university press.

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies, 66*(2), 98–112. https://doi.org/10.1016/j.ijhcs.2007.09.003

*Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. Journal of Educational Psychology, *110*(2), 250-268. https://doi.org/10.1037/edu0000221

*Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., & Weston, T. (2013). My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology, 105*(4), 1115–1125. https://doi.org/10.1037/a0031589

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI Magazine, 34*(4), 66–84. https://doi.org/10.1609/aimag.v34i4.2490

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

Authors and Affiliations

Chih-Pu Dai[1] · **Fengfeng Ke**[2] · **Yanjun Pan**[3] · **Jewoong Moon**[4] · **Zhichun Liu**[5]

✉ Chih-Pu Dai
cdai@hawaii.edu

1 Department of Learning Design and Technology, College of Education, University of Hawai‘i at Mānoa, 1776 University Ave, Honolulu, HI 96822, USA

2 Florida State University, Tallahassee, FL, USA

3 Southern Methodist University, Dallas, TX, USA

4 The University of Alabama, Tuscaloosa, AL, USA

5 The University of Hong Kong, Hong Kong SAR, China