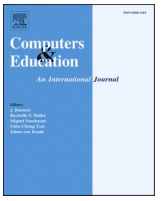
Computers & Education 223 (2024) 105165

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu



Pedagogical agent design for K-12 education: A systematic review

Shan Zhang [a], Chris Davis Jaldi [b], Noah L. Schroeder [a,*], Alexis A. López [c], Jessica R. Gladstone [d], Steffi Heidig [e]

[a] University of Florida, Gainesville, FL, 32611, USA
[b] Wright State University, Dayton, OH, 45435, USA
[c] Educational Testing Service, Princeton, NJ, 08541, USA
[d] University of Illinois Urbana-Champaign, Champaign, IL, 61820, USA
[e] University of Applied Sciences Zittau/Görlitz, Görlitz, Germany

ARTICLE INFO

Keywords:
Pedagogical agent
K-12 education
Human-computer interface
Virtual human
Virtual character

ABSTRACT

Pedagogical agents (PAs) are increasingly being integrated into educational technologies. Although previous reviews have examined the impact of PAs on learning and learning-related outcomes, it still remains unclear what specific design features, social cues, and other contextual elements of PA implementation can optimize the learning process. These questions are even more prevalent with regards to the K-12 population, as most reviews to date have largely focused on post-secondary learners. To address this gap in the literature, we systematically review empirical studies around the design of PAs for K-12 learners. After reviewing 1374 studies for potential inclusion, we analyzed 44 studies that met our inclusion criteria using Heidig and Clarebout's (2011) frameworks. Our findings showed that learners had preferences for specific types of PAs. While these preferences were not always associated with increased learning outcomes, there is a lack of research specifically investigating the intersection of perceptions and learning. Our results also showed that pedagogical strategies that are effective for human teachers were effective when used by PAs. We highlight what specific design features instructional designers can use to design PAs for K-12 learners and discuss promising research directions based on the extant work in the field.

1. Introduction

From early computer-assisted instruction to the current innovations in artificial intelligence (AI), advancements in educational technology have continually transformed traditional teaching and learning methods, providing students with a richer and more diverse range of resources and interactive experiences. One popular approach to enhancing educational technologies is to embed pedagogical agents (PAs)—embodied virtual characters on the screen that interact with learners—into systems to engage learners and facilitate learning. While PAs have been extensively studied for decades (Siegle et al., 2023), having been shown to positively impact student learning (Kim, 2009; Kizilkaya & Askar, 2008) and improve their engagement and motivation (Arguedas & Daradoumis, 2021; Chen & Chen, 2014) in K-12 settings, recent advancements in AI have once again drawn attention to their design and use, highlighting their potential to enhance educational experiences.

* Corresponding author. University of Florida, Department of Computer and Information Science and Engineering, Gainesville, FL, 32611-6120, USA.
E-mail address: schroedern@ufl.edu (N.L. Schroeder).

https://doi.org/10.1016/j.compedu.2024.105165
Received 26 June 2024; Received in revised form 10 September 2024; Accepted 13 September 2024
Available online 14 September 2024

Previous reviews have found mixed effects of PAs for facilitating learning among K-12 students, although sample sizes were small which makes generalization difficult. For example, Schroeder et al. (2013) found that PAs were more advantageous to K-12 students than to students in post-secondary education, whereas Castro-Alonso et al. (2021) and Peng and Wang (2022) did not find any significant advantages for K-12 students compared to other age groups. More recently however, Schroeder et al.'s (under review) three-level meta-analysis of PAs' impact on K-12 learners had a larger sample size than previous studies and found that PAs do help K-12 students learn and improve their motivation.

Although there are several reviews in the field of PAs examining how effective they are (Castro-Alonso et al., 2021; Dai et al., 2022; Davis, 2018; Heidig & Clarebout, 2011; Peng & Wang, 2022; Schroeder et al., 2013; Wang et al., 2017), these reviews generally focused on comparing PAs to non-agent systems, rather than comparing the design of one PA to a PA of a different design. Consequently, we have little information about how to design effective PAs (Siegle et al., 2023), especially for K-12 learners as opposed to post-secondary learners. In order to bridge this gap and build our understanding of how to design effective PAs for K-12 learners in various learning situations, we systematically review the literature comparing PAs of one design style to PAs of another design style through the lens of published frameworks for PA design (Heidig & Clarebout, 2011). To the best of our knowledge, this study represents the first systematic, large-scale synthesis effort to address the question: How should we design PAs to optimize their effectiveness in facilitating learning for K-12 learners?

2. Theoretical framing and prior work

Until recently, it was easy to believe that there were few studies focused on investigating PAs with K-12 learners because previous reviews found few studies in K-12 contexts (Castro-Alonso et al., 2021; Dai et al., 2022; Schroeder et al., 2013; Wang et al., 2023). However, Zhang et al.'s (2024) scoping review located more than 100 studies of PAs taking place with K-12 learners. There is now meta-analytic evidence that PAs can facilitate learning for K-12 learners, as Schroeder et al. (under review) conducted three-level meta-analyses and found that PAs help K-12 students learn ($g = 0.42$, $p < 0.001$, $k = 70$) and improve their motivation ($g = 0.48$, $p < 0.001$, $k = 47$) compared to non-PA conditions. However, due to the nature of the analysis, Schroeder et al. were unable to suggest how to design PAs to facilitate learning for the K-12 population. In this study, we therefore examine *how* to design PAs most effectively for this population.

2.1. Why pedagogical agent design matters

Why might PA design influence K-12 students' learning? The cognitive-affective-social theory of learning in digital environments (CASTLE) (Schneider et al., 2022) offers an explanation that can apply to learners of many ages. This theoretical perspective posits that the presence of social cues in digital environments not only activates learners' social schemata, with activation increasing with the number and strength of the social cues, but also significantly influences the other cognitive processes involved in learning. CASTLE highlights how the social processes between learners and digital learning materials are influenced by affective, motivational, and metacognitive factors.

While CASTLE highlights the importance of social cues in digital learning environments, such as communication, appearance, PA gestures, eye contact, and facial expressions, it does not explicitly inform how we should design agents that could positively impact learning. It remains unclear what specific design features are more effective, which social cues more significantly influence learning, and how these elements should be implemented to optimize the learning process. Yet, this theoretical grounding remains the most specific we have in regards to PA design and why certain design features may provide benefits compared to others, and it pairs well with the design framework proposed by Heidig and Clarebout (2011) as discussed below.

2.2. The current state of pedagogical agent design for K-12 learners

Various frameworks have guided agent design. We use Heidig and Clarebout's (2011) interconnected Pedagogical Agents-Conditions of Use Model (PACU) and Pedagogical Agents-Levels of Design (PALD) framework as our guides because they can be broadly applied across many different types of PA implementations.

2.2.1. Pedagogical Agents-Conditions of Use Model

Informed by previous work, Heidig and Clarebout (2011) proposed a multi-level framework, the PACU model, to systematically design PAs and to guide systematic comparisons of existing studies on PAs. The PACU framework comprises four conditions for the use of PAs. These include the learning environment where the PA is used, learner characteristics, PA function, and PA design. We break down each condition and provide a brief description with examples from studies in a K-12 context below.

2.2.1.1. Learning environment. PAs are used across various learning environments such as tutorial programs (e.g., Domagk & Niegemann, 2005), intelligent tutoring systems (e.g., Arroyo et al., 2010; Beal et al., 2010) or cognitive tutors (Pane et al., 2010), microworlds (Yalçin, Lalle, & Conati, 2022) or virtual reality (Dai et al., 2024).

2.2.1.2. Characteristics of the learner. In Heidig and Clarebout's (2011) model, learner characteristics encompass cognitive factors such as prior knowledge, academic competency like GPA scores (e.g., Johnson et al., 2013), emotional factors such as boredom, pride,

pleasure, and shame, motivational factors such as interest, achievement motivation, self-efficacy (e.g., Arguedas & Daradoumis, 2021; Pérez-Marín & Pascual-Nieto, 2013), and metacognitive factors like self-regulation (e.g., Daradoumis & Arguedas, 2020)). These learner characteristics may moderate the effectiveness of the PA and thus should be considered when designing a PA.

2.2.1.3. Functions of the pedagogical agent. Heidig and Clarebout (2011) suggested that PA functions and roles can be categorized as motivation (e.g., arouse interest), information (e.g., drawing the learner's attention to the learning content, activate prior knowledge), information processing (e.g., providing explicit information about prerequisites), storing and retrieving (e.g., comparing and integrating new information into the existing structure), transfer of information (e.g., applying and transferring new knowledge to other new problems), or monitoring and directing learner's activities.

2.2.1.4. Design of the pedagogical agent. Clearly, PAs can play multiple roles in the K-12 learning environment, but what should they look like? Heidig and Clarebout (2011) noted that PA design is the final element of the PALD model. However, they also proposed an in-depth design framework to facilitate understanding the levels of PA design, which we describe next.

2.2.2. Pedagogical agents - Levels of Design Model

Recognizing PA design is complex, Heidig and Clarebout (2011) proposed the PALD framework to organize design features into three different levels for systematically describing the design of the PA.

2.2.2.1. Global design level. This level of design involves deciding to use humanoid characters, such as cartoon humans (Sahimi et al., 2010), or non-humanoid characters, such as a bug (Stelling, 2002).

2.2.2.2. Medium design level. This level involves two categories. The first category (1) is about technical decisions concerning the PA's visual and auditory presence. The choices range from (1a) the level of realism and lifelikeness varying from simple cartoons to lifelike human representation (Kautzmann & Jaques, 2019), to (1b) the presence and the complexity of animations (Bringula et al., 2018), from basic movements to intricate motions. Further, this category includes (1c) decisions on how the PA communicates, whether via spoken or printed text, in which spoken text may be presented by human vs. computer-simulated voice, and speech style, whether it addresses the learner in a personalized or formal style. The second category (2) is the choice of character, and it can be guided by: (2a) determining characteristics of the PA such as its likeability or competence (e.g., high, low), (2b) defining PA's role (e.g., teachers, experts, or peers), or referring to real role models (e.g., lectures, trainers). The decisions on the medium level of PA design can subsequently inform the design/choice of the PA's appearance, actions, and statements on the detail design level (Heidig & Clarebout, 2011).

2.2.2.3. Detail design level. After decisions are made at the global and medium levels, decisions at the detailed level must also be made specifically around features related to the visual and auditory presence of the PA. Visual features include agent age, gender, clothing, weight, and ethnicity whereas auditory features of the voice involve intonation, accentuation, and speech rate (Heidig & Clarebout, 2011).

Although empirical studies have investigated the impact of various design components on K-12 students' learning (Bringula et al., 2018; Dinçer & Doğanay, 2017) and motivation outcomes (Riedmann et al., 2022), at the time of Heidig and Clarebout's (2011) review of PA design, there was insufficient evidence to make many specific claims outside of more research being needed (Heidig & Clarebout, 2011).

2.3. The present study

As mentioned, previous reviews have highlighted the benefits of PAs in facilitating learning among K-12 students (Peng & Wang, 2022; Schroeder et al., 2013; Schroeder et al., under review). Given these impacts and the increasing number of studies examining different design components of PAs, there is a crucial need to systematically understand how PA design influences learning and learning-related outcomes. Guided by Heidig and Clarebout's (2011) framework, in this systematic review we sought to investigate how we can most effectively design PAs for K-12 learners. Our research questions are as follows:

Pedagogical Agents - Conditions of Use Model (PACU):

RQ1. How does the learning environment affect the effectiveness of pedagogical agents?

RQ2. In what ways do learner characteristics influence the effectiveness of pedagogical agents?

RQ3. How do various functions of pedagogical agents influence K-12 learners?

Pedagogical Agents - Levels of Design Model (PALD):

RQ4. What is the influence of agent design at the global level?

RQ5. What is the influence of agent design at the medium level?

RQ6. What is the influence of agent design at the detail level?

3. Methods

We report our systematic review following the PRISMA guidelines (Page et al., 2021). All data used in the analyses are available as supplementary materials.

3.1. Literature search

This systematic review is part of a large-scale, multifaceted review examining the impact of PAs on K-12 students' learning. Accordingly, the studies identified for this analysis were based on the literature search from Zhang et al.'s (2024) scoping review. Below, we provide a high-level overview of the literature search process.

In September 2023, we conducted a comprehensive search across nine major databases (with no restriction on publication dates) covering various fields such as education, psychology, social sciences, computing, and health professions using the following search string, ("virtual human"* OR "embodied agent"* OR "virtual character"* OR "pedagogical agent"* OR "conversational agent"* OR "motivational agent"*) AND (k-12 OR elementary OR primary OR secondary OR middle OR high) AND (learn* OR motivat* OR self-

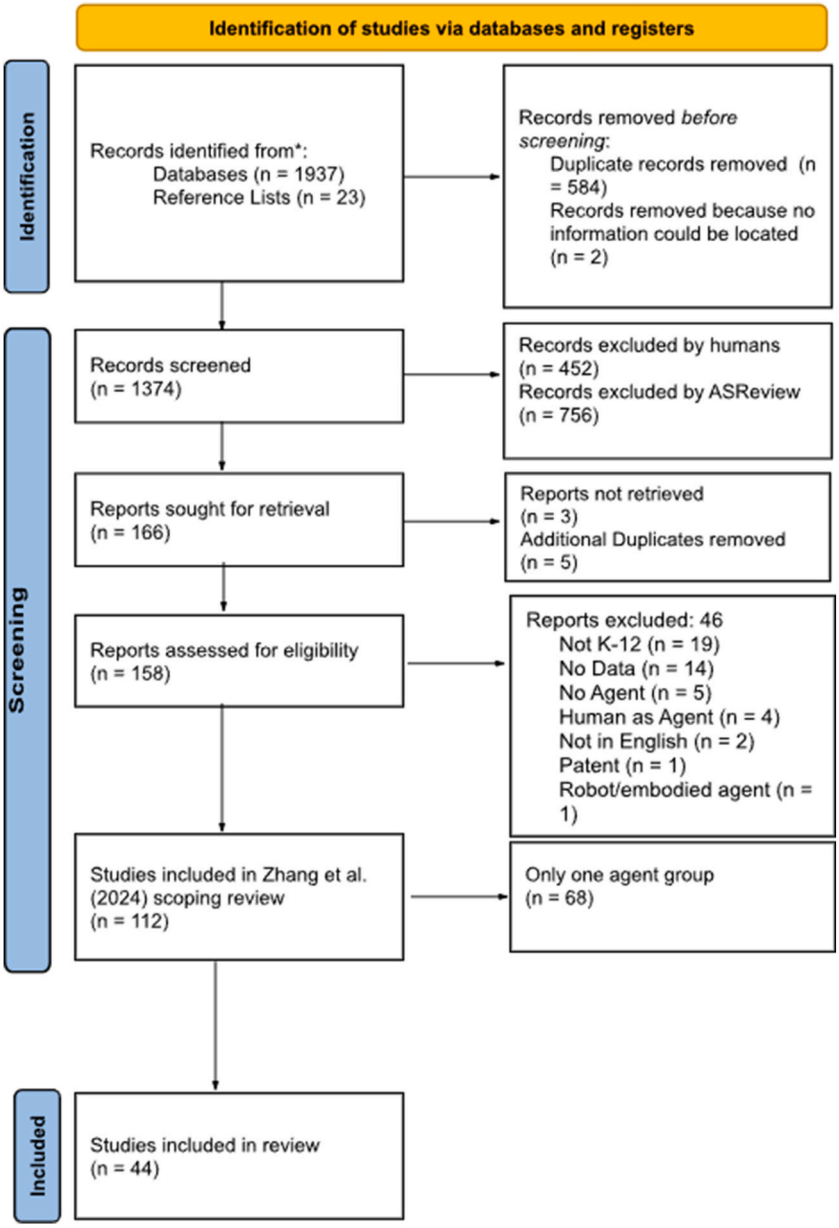


Fig. 1. PRISMA flowchart adapted from Page et al. (2021).

efficacy OR self-confidence OR ability belief* OR self-concept OR interest* OR engag* OR value* OR util* OR "sense of belonging" OR belong* OR achiev* OR develop*). Additionally, we included studies that met inclusion criteria from three prior pedagogical agent meta-analyses (Castro-Alonso et al., 2021; Schroeder et al., 2013; Wang et al., 2023). After excluding duplicates, we identified 1374 abstracts for initial review. After applying the inclusion (The study must include at least one condition with a visible pedagogical agent, which is a virtual character and not a video, rendering, or image of an actual human, involve K-12 learners as participants, and collect either quantitative or qualitative data) and exclusion criteria during abstract screening, 158 full texts were selected for full-text screening. Of these full texts, 112 studies were included in the scoping review. To identify studies for the systematic review, the full text of the 112 studies in Zhang et al.'s (2024) scoping review were examined to see if they met the inclusion and exclusion criteria below.

3.2. Inclusion and exclusion criteria

To be eligible to be included in the systematic review, studies had to.

1. Compare at least two conditions with a visible PA, which must be a virtual character and not a video, rendering, or image of an actual human.
2. Be a study with K-12 learners as the participants.
3. Collect either quantitative or qualitative data.

Studies were excluded if.

1. The PA was a physically embodied robot or other physically embodied object.
2. The study examined the use of avatars (or other representations of self).
3. The study only examined health-relevant outcomes (i.e., body mass index).
4. The study was not published in English.
5. The study did not contain primary data (e.g., conceptual work or no agents present).
6. The study was not publicly available.

3.3. Study screening

We examined the 112 full texts of articles included in Zhang et al.'s (2024) scoping review, and following the inclusion and exclusion criteria above, 44 studies were included in the analyses (Fig. 1).

3.4. Data extraction

One author extracted data from all of the studies.

3.4.1. Inter-rater agreement

To calculate inter-rater agreement, a second author coded 22% of all studies ($n = 10$) independently, resulting in an inter-rater agreement of 90%. Any discrepancies were resolved through discussion and re-examining the primary studies.

3.4.2. Variables extracted

To understand how to design agents most effectively, we extracted a variety of variables from each study focused on the specific design components experimentally manipulated, as well as the study outcomes. For example, we coded each study in relation to the PACU and PALD frameworks, recording how each study aligned with each level within each framework. Full details of our coding scheme are available in Appendix A.

3.5. Data availability

The full coding forms with the data extracted from each study are included as Supplementary Materials as Table S1.

4. Results

We located 44 studies that compared a group learning with a PA to another group learning with a different PA. Some of these studies also included a no-PA control group. The studies are categorized according to the PACU model in Table A2, and the PALD framework in Table A3.

4.1. RQ1: How does the learning environment affect the effectiveness of pedagogical agents?

Among the 44 studies, only two manipulated the learning environment in which the PAs were implemented. Riedmann et al. (2022) developed a theory-based educational mobile application that included two games with one PA to help improve primary students'

reading skills. They investigated two learning environments with the same basic structure. One followed the ARCS (Attention, Relevance, Confidence, and Satisfaction) model with adaptable intelligent behavior. The other provided only auditory content. Results showed no significant differences in motivation between the environments. However, one of the two games in the ARCS model group had a positive impact on students' performance. Meanwhile, Yalçin et al. (2022) investigated how an intelligent PA (IPA) deployed in a free-form game-design learning environment can support primary students' computational thinking by providing personalized interventions. Compared to no intervention and a one-time intervention, they found that repeated interventions by the IPA led to fewer errors and more correct behaviors in students. Overall, Riedmann et al. (2022) suggests that specific, theory-based design elements within the learning environment could enhance students' learning and Yalçin et al. (2022) highlights the importance of continuous and adaptive support provided by PAs to enhance learning. However, these studies are quite different in scope which makes generalizations difficult.

4.2. RQ2: In what ways do learner characteristics influence the effectiveness of pedagogical agents?

Seven studies examined how learner variables such as mental load, prior knowledge, and demographics influence the efficacy of pedagogical agents. For instance, Beege et al. (2020) found that learners under a high mental load performed better on multiple-choice questions when assisted by neutral PAs, whereas those under a lower mental load showed improved performance with enthusiastic PAs. Researchers have also explored how learners with autism learn from PAs. Grynszpan et al. (2008) compared the performance of participants with and without autism using a simple or rich multimedia interface. The findings indicated that participants with autism performed better in simple interfaces compared to rich interfaces. The influence of learners' prior knowledge has also been explored. Johnson et al. (2013, 2015) found through two studies that learners with low prior knowledge learned more from animated PAs that provided signaling as compared to no signaling, and this learning gain was not found for high prior knowledge learners. Moreover, research into the effects of gender and ethnicity on agent interaction reveals interesting insights. Notably, multiple studies found that learners' gender and/or ethnicity influenced their perceptions and attitudes towards the PA (Kim & Lim, 2013; Kim & Wei, 2011; Ozogul et al., 2013), but this did not always translate to impacts on learning outcomes.

Overall, studies showed that learning from PAs was most effective when learners had low prior knowledge and the interventions were designed to fit within the cognitive capacity of learners (e.g., Beege et al., 2020; Grynszpan et al., 2008; Johnson et al., 2013, 2015). Meanwhile, several studies highlighted the impact of learner gender and ethnicity on their experiences in PA-based learning (Kim & Lim, 2013; Kim & Wei, 2011; Ozogul et al., 2013). However, these studies did not consistently find significant differences in learning outcomes based on learner demographics.

4.3. RQ3: How do various functions of pedagogical agents influence K-12 learners?

Eighteen out of the 44 studies examined in this review focused on agent function. Among these studies, nine investigated teaching strategies. For example, Abdelghani et al. (2022) and Alaimi et al. (2020) investigated the impact of PAs asking convergent or divergent questions on subject efficacy, performance, and motivation towards interaction with the system. The results indicated significant differences between conditions, with the divergent questions leading to better outcomes in efficacy ($p < 0.001$), performance ($p < 0.001$), and motivation as reflected by increased time on task ($p = 0.04$). Meanwhile, Jaques et al. (2009) and Lester et al. (1997) examined the effects of PAs with different behaviors, focusing on agent presence and communication strategies. Others manipulated assistive behavior in teaching through dynamic metacognitive instruction scaffolding strategies (Kautzmann & Jaques, 2018, 2019; Molenaar et al., 2012), communication aimed at reducing subject anxiety (Wei, 2010), and personalized intelligent corrective feedback systems (Xu, 2009). Critically examining these studies, we generally found that agents with more pedagogical and humanlike features, such as providing hints and displaying emotions, supported learning more effectively and were preferred by learners more than those with fewer of these features (Jaques et al., 2009; Lester et al., 1997), and providing metacognitive instruction and scaffolds was also effective (Molenaar et al., 2012).

Six studies manipulated the PA's communication strategies. These studies examined factors such as who initiated the conversation first and the modes of communication (Holmes, 2007; Wiggins, 2021). Researchers also examined the role of the agent (Carpenter, 2013; Girard & Johnson, 2010; Murray & Tenenbaum, 2010), such as instructor versus peer, with instructor-style communication yielding better learning but findings also suggesting that learner characteristics, such as prior knowledge level and role preference might have impacted the findings. Additionally, researchers have investigated altering the behavior of the agent, such as presenting information through being informative, testing, encouraging, or grumpy (Pérez-Marín & Pascual-Nieto, 2013). All of these studies focused on measuring learning, motivation or perception of the agent system in terms of retention, self-efficacy and engagement, with better results observed for agents that asked questions followed by elaborative explanations (Holmes, 2007; Pérez-Marín & Pascual-Nieto, 2013; Wiggins, 2021). Researchers have also examined game-like elements in PA-based systems. Two studies (Chen & Chan, 2008; Chen & Chen, 2014) involved the introduction of game-based competition, with learners engaging in competitive activities against opponents where each learner's PA was pitted against each other. Results indicate that competitive conditions demonstrated greater interaction and performance results ($p = 0.048$) and significant differences in relevance, challenge, and enjoyment measures compared to conditions that didn't contain competition.

In summary, effective teaching strategies utilized by human teachers, such as metacognitive instructions and scaffolding, significantly improved performance and motivation when learning with PAs, and PAs initiating conversations and providing explanations increased engagement. Additionally, game-based competition boosted learner interaction and performance.

4.4. RQ4: What is the influence of agent design at the global level?

At the global level of Heidig and Clarebout's (2011) framework, PA designers have to decide to either use humanoid or non-humanoid agents. We also included studies where learners were able to choose which PA or customize the PA they learn with into this category. Eighteen of the 44 studies in our sample experimentally manipulated variables at the global design level.

Several studies included not only different PA groups but also a non-PA control group. Sixteen of the studies included a humanoid agent. We consistently found a significant increase in learning when humanoid PAs were tested against non-agent conditions (Dincer and Doganay, 2015, 2017; Holmes, 2007; Jaques et al., 2009; Johnson et al., 2013, 2015; Kim et al., 2007; Murray & Tenenbaum, 2010). Meanwhile, five studies in our sample used a non-humanoid agent. Non-humanoid agents may take the form of animals like a dog (Chen & Chen, 2014), figures like "Smiley" (Girard & Johnson, 2010), or even inanimate objects such as a drone (Makransky et al., 2019). We consistently found that non-humanoid characters were associated with significantly better learning outcomes than no-agent conditions (Dincer and Doganay, 2015; Jing et al., 2022). While the studies above are largely focused on comparing various PAs to non-agent conditions, four studies have directly investigated the use of humanoid to non-humanoid agents. However, these studies showed mixed results. Specifically, we found that preferences may vary according to the agent's role type (Girard & Johnson, 2010) and sometimes the gender of the participants (Makransky et al., 2019).

Although six studies examined student preferences for choosing their own agents in STEM classroom settings, no consistent pattern emerged in terms of their choice of agent (Girard & Johnson, 2010; Haake & Gulz, 2009; Kim et al., 2007; Kim & Wei, 2011). We also located four studies focused on agent customization for each learner. In this context, agent customization was defined as the capacity to personalize and design one's preferred agent appearance, typically encompassing factors such as hairstyle, hair color, skin color, gender, name, clothing, and age, among others. These studies, all conducted by the same first author, demonstrated significant gains in learning and motivation outcomes (Mei, 2016; Mei et al., 2015a, 2015b, 2018). Although the customization of the PA that learners were able to choose in these particular studies were on detail design level (e.g., hairstyle, clothing, gender), we categorized these studies on the global design level as the decision to provide customization or not is a global design decision when implementing PAs.

To summarize, both humanoid and non-humanoid agents were consistently found to lead to better learning than non-PA conditions. When it comes to comparing humanoid and non-humanoid agents, the results were inconsistent and there were few studies. We found no consistent pattern around what PA students chose in the six studies which investigated this, and while we located four studies around learners being able to customize their agent and all led to positive effects from letting learners customize their PA, all four studies were conducted by one research group.

4.5. RQ5: What is the influence of agent design at the medium level?

For a more precise understanding of PA design, we conducted a detailed analysis at the medium level of Heidig and Clarebout's (2011) framework. At the medium level, two categories can be distinguished: (1) technical decisions concerning (1a) lifelikeness and realism of the PA, (1b) animation level and (1c) the way the PA communicates as well as (2) the choice of the character either guided by (2a) desired characteristics (e.g. competence, likeability) or (2b) the PA's role in the learning environment. Thirteen out of the 44 studies reviewed can be classified at the medium design level.

Focusing on (1) technical decisions, five studies in our sample examined (1a) the concept of lifelikeness or realism of the PA. Lifelikeness/Realism pertains to the degree of fidelity in both visual and auditory presentations, aiming to closely resemble real-life entities in behavior and appearance, emphasizing synchronization and overall presence. We found studies that manipulated features such as zoomorphism (Jing et al., 2022), PAs appearance ranging from a cartoonish appearance to a more realistic appearance (Girard & Johnson, 2010; Sahimi et al., 2010), and textual and facial expressions (Bringula et al., 2018). In short, while some studies found that the level of realism positively influenced engagement and learning outcomes, others found no significant differences or observed nuanced effects contingent on specific contexts or conditions. This indicates that while lifelikeness/realism may influence learners' preferences, the specific context and role of the agent are crucial in determining these preferences. Moreover, it is clear that lifelikeness/realism have various critical factors within them, from visual fidelity through the use of facial expressions, and therefore it is clear more work is needed to better differentiate and understand the factors that may, or may not, influence learning in relation to lifelikeness/realism of the PA.

Meanwhile, none of the studies in our sample specifically focused on (1b) animating the PA by comparing static to animated agents.

Two studies considered (1c) the PA's communication, encompassing various modalities such as visual, auditory, or their combination (Holmes, 2007; Nguyen, 2022). Their results were inconclusive regarding the impact of communication on learning outcomes.

Focusing on (2) the choice of the character, two studies determined desired characteristics of the PA (2a) and manipulated the agent's competence. They employed diverse metrics to evaluate the competence of the agents (Nguyen, 2022; Tärning and Silvervarg, 2019). We did not see consistent trends in outcomes across studies.

Nine studies focused on (2b) the role of the PA. This is a complex aspect of PA design that encompasses the agent's instructional function and approach to learning, which may influence communication dynamics (e.g. communication style of an instructor agent versus a peer agent; Haake & Gulz, 2009). Most studies focused on specifically comparing instructor versus peer PAs (Girard & Johnson, 2010; Haake & Gulz, 2009; Murray & Tenenbaum, 2010; Nguyen, 2022). However, there were often multiple design factors at play which may confound our interpretation. We found that varying the PA's role often resulted in learner preferences for a particular agent condition within the experiment, such as a preference for an instructor and peer PA condition over only an instructor PA condition (Murray & Tenenbaum, 2010), but the overall impact on learning remained inconclusive across studies (Girard & Johnson, 2010; Haake & Gulz, 2009; Nguyen, 2022; Stelling, 2002).

To summarize the medium design level within the context of PAs, we found that researchers investigated a wide variety of PA design manipulations in their studies, although there were few consistencies. Studies on lifelikeness/realism have revealed learner preferences predominantly based on the specificity of roles rather than their impact on learning. Studies on communication design yielded mixed insights. Results indicated deep explanations led to better learning, and the effect of tone was inclusive. Similarly, studies around the PA's competence also revealed mixed outcomes. As such, it seems as though the role of the PA may not be as important as the pedagogy the PA facilitates through that role. These findings highlight the complexity and interconnectedness of medium-level design decisions in shaping PA effectiveness, making it challenging to conclusively say specific features are the cause of observed effects.

4.6. RQ6: What is the influence of agent design at the detail level?

The design decision process also involves the fine-grained selection of attributes of PAs. Nine studies involved detailed design manipulation. They investigated the following characteristics of the PA: age, gender, clothing, ethnicity, and voice type.

Four studies examined the PAs age, all of which consistently showed learners' preference for near peer age agents (Alsharbi & Richards, 2017; Kim et al., 2007; Ozogul et al., 2013, pre-study, exp. 2). However, we did not see any consistent evidence that working with a similarly aged PA or an older PA led to increased learning.

Six studies analyzed the impacts of PA gender, with the majority showing learners' preferences for female agents, particularly among male learners (Alsharbi & Richards, 2017; Kim et al., 2007; Ozogul et al., 2013, pre-study, exp. 1). However, when it came to learning, only one study specifically examined the influence of agent gender on learning and better transfer scores ($p < 0.05$, $\eta_P^2 = 0.013$) were obtained when learning with an agent of the opposite gender of the learner (Ozogul et al., 2013, exp. 2).

Two studies manipulated the PA's clothing. They showed mixed results, with no specific preference or significant impact between a formally attired PA wearing a tie to an informally attired one with a cap in one study (Nguyen, 2022) but a strong preference among learners for fun and cool characters in the other study (Ozogul et al., 2013, pre-study).

Regarding the ethnicity of PAs, two studies showed consistent results for preference towards agents of the same ethnicity as the learner (Alsharbi & Richards, 2017; Kim & Wei, 2011). However, learning outcomes and self-efficacy remained unaffected regardless of the learners' preference.

Lastly, two studies examined the impact of voice type within interventions. Beege et al. (2020) investigated the effects of enthusiastic versus neutral voice types and high versus low mental load conditions on various outcomes, finding significant interactions between enthusiasm and mental load on video ratings and cognitive load. However, most main effects were nonsignificant. In contrast, Sedlacek et al. (2017) compared content-novice versus content-expert and professional versus amateur voice actors, discovering few significant differences, with some notable exceptions where content experts outperformed novices in lesson completion time, and content novices excelled in quiz accuracy. However, the two studies yielded mixed results when examined together.

Overall, while we can see that decisions at the detail level may influence K-12 learners' preferences, we do not see strong evidence that these preferences directly influence learning outcomes. However, we note that research in most areas is limited, as we discuss further in the discussion section.

5. Discussion

5.1. RQ1: How does the learning environment affect the effectiveness of pedagogical agents?

We located two studies that manipulated the learning environment itself and included PAs (Riedmann et al., 2022; Yalçin et al., 2022). These studies were quite different from one another, making it challenging to draw generalizable conclusions. Our results are consistent with those of Heidig and Clarebout (2011), who found that just adding additional features or employing a more complex design (e.g., interactivity) in learning environments can lead to mixed results and may even hinder learning.

5.2. RQ2: In what ways do learner characteristics influence the effectiveness of pedagogical agents?

Our results highlighted that K-12 learners' personal characteristics could notably influence their interactions with a PA. This suggests, although we note it should be seen as preliminary findings since it is derived from two studies from the same research group (Johnson et al., 2013, 2015), that PAs can help less knowledgeable students learn without impeding the knowledge gain of those who have high prior knowledge. This is in contrast to some instructional design effects that suffer from an expertise reversal effect (Kalyuga, 2007, 2009), where what is effective for low prior knowledge learners actually inhibits those with more prior knowledge.

Moreover, there is some evidence that learner gender and ethnicity can influence their interactions with or perceptions of agents. Specifically, researchers found evidence that learner demographics influenced students' attitudes or choice of agent (Kim & Wei, 2011; Ozogul et al., 2013). While this work reminds us of the similarity-attraction hypothesis (Moreno & Flowerday, 2006), we did not see consistent impacts on learning.

Together, although the work investigating the impact of learner characteristics on learning with PAs for K-12 students is limited, we see promise in this line of inquiry. For example, it is well known that cognitive characteristics can influence learning in various instructional contexts, and we see preliminary evidence of that here. Similarly, we see preliminary evidence that learners' demographic characteristics may influence how they perceive PAs. One critique of this finding could be if we did not see impacts on learning, why should we consider learner demographic characteristics? We argue that there are various reasons why PAs may be

implemented - while they are always designed to facilitate learning, perhaps this facilitation is driven, either directly or indirectly, by increasing learners' interest or self-efficacy. In short, we can foresee many situations in which considering learner demographics may be important to the learning situation even if the inclusion of the PA does not directly influence learning, but rather some other learning-related variable.

5.3. RQ3: How do various functions of pedagogical agents influence K-12 learners?

Unsurprising, we found that strategies human teachers use in the classroom, such as providing metacognitive instruction and scaffolding (Kautzmann & Jaques, 2018, 2019; Molenaar et al., 2012), and initiating conversations and providing explanations (Holmes, 2007; Pérez-Marín & Pascual-Nieto, 2013; Wiggins, 2021), fostered better outcomes than when PAs did not have these features. These findings support long-held claims that the instructional methods are of utmost importance, and the way the instruction is delivered (i.e., the medium of delivery) is of lesser importance (Choi & Clark, 2006). In a PA context, this implies that the pedagogical approaches the agent uses, rather than its physical appearance, is what may influence learning. However, we believe it is premature to fully support that conclusion with confidence. As mentioned when discussing RQ2, we see evidence that learner characteristics may interact with PA characteristics. Thus, while the pedagogy the PA engages in is important, its appearance may also be important in supporting the learning process even if learning outcomes are not directly impacted.

5.4. RQ4: What is the influence of agent design at the global level?

Our results reiterate meta-analytical findings (Schroeder et al., 2013, under review) showing that including a PA was generally found to be more effective for supporting K-12 student learning outcomes than conditions without a PA. Two more pertinent questions then come to mind: should students be able to choose or customize their PA, and what should a PA look like?

The customization of PAs by learners was investigated less than agent choice. We located only four studies by the same laboratory that investigated customization, but they found promising results (Mei, 2016; Mei et al., 2015a, 2015b, 2018). Due to the limited scope of work in this area, it does not seem appropriate to draw generalizable results at this time, however this is a potential area for future research. Specifically, we wonder if customizing the PA may help the learner develop a relationship with the PA and support their motivation and interest, which can subsequently support learning. Future research is needed to investigate this line of reasoning.

With regards to student choice of agent, we found largely mixed results regarding student preferences when choosing between a humanoid or non-humanoid PA. For example, our results showed that the learners' preference may depend on the role of the agent (Girard & Johnson, 2010) or the learners' demographic characteristics (Makransky et al., 2019).

So, with no clear preference for a humanoid or non-humanoid PA, how should we design PAs? Our remaining research questions investigated the visual design of PAs in more detail.

5.5. RQ5: What is the influence of agent design at the medium level?

Few studies investigated similar design concepts in similar ways. We observed some effects on learners' perceptions, although these were not always aligned with increases in learning outcomes (Girard & Johnson, 2010; Haake & Gulz, 2009; Nguyen, 2022; Stelling, 2002). As such, it is challenging to be able to draw generalizable conclusions such as if a specific design feature in the medium level is often leading to increases in learning depending on the specific context. We believe this may be due to the fact that we found significant overlaps between the function of the agent (a condition of use) and the PA design at the medium level. Specifically, within Heidig and Clarebout's (2011) framework, what the PA does (e.g., provide feedback) is conceptualized as the function of the PA - a condition of use (PACU). Meanwhile, what the PA looks like (e.g., appearing as an instructor or peer) is conceptualized on the medium design level (PALD). In design and research practice, however, the actions of the PA and its appearance are often intertwined by designing PAs that look and act like instructors or peers. Clearly, these PAs are going to function differently given that they are designed to play significantly different roles.

5.6. RQ6: What is the influence of agent design at the detail level?

Our results showed that K-12 learners preferred similarly aged PAs (Alsharbi & Richards, 2017; Kim et al., 2007; Ozogul et al., 2013, pre-study, exp. 2) and students generally preferred female agents (Alsharbi & Richards, 2017; Kim et al., 2007; Ozogul et al., 2013, pre-study, exp. 1). While this again, in part, aligns with the previously suggested similarity - attraction hypothesis (Moreno & Flowerday, 2006), learner preferences did not always align with better learning outcomes.

The studies conducted around the type of voice used by the PA were largely inconsistent with one another with regards to what they were testing. However, we know from studies with learners over the age of 18 that the type of voice used by the PA can influence learning. For example, we know modern text-to-speech engines can provide similar learning results to a human-recorded voice (Chiou et al., 2020; Craig & Schroeder, 2017). Thus, there is clearly opportunity for more work around voice, exploring features such as prosody (Davis et al., 2019), speech engines (Craig & Schroeder, 2017), or various other features (for a review, see Seaborn et al., 2022).

In closing, outside of a learners' preferences for PAs of a similar age to them, results at the detail level of design are generally mixed. We hypothesize this is largely due to a lack of consistent, systematic lines of research investigating the intersection of learner demographics and agent design at the detail level. There is clearly space for work in this research area.

6. Implications

6.1. Implications for theory

While this systematic review was not positioned to test one theoretical approach compared to another, we do see preliminary, partial support for the general premise driving CASTLE (Schneider et al., 2022). A general premise of CASTLE is that social cues in multimedia learning environments (e.g., a PA) can activate social processes in the learner, alongside metacognitive and motivational processes, and subsequently influence learning. It is clear that PAs are intended to act as social cues in the learning environment. We also know from our results and the results of previous meta-analyses that the inclusion of PAs can help K-12 students learn (Schroeder et al., 2013, under review). In this review, we found evidence that learner demographic characteristics influenced their interaction with the system or their preferences (Kim & Lim, 2013; Kim & Wei, 2011; Makransky et al., 2019; Ozogul et al., 2013), which implies that the social cue (the PA), as well as the learners' characteristics, are influencing learners' social processes. While we do not see concrete evidence that learners' preferences consistently lead to increases in learning, there are few studies that have investigated this in depth, and few studies that explored learners' demographic or situational characteristics in enough detail to really understand the complexities involved with social processes. Moreover, we found few studies that examined moderation or mediation of learning based on other factors. We view the intersections of learner characteristics, PA design, and mediation and moderation of learning as key areas for researchers to explore in the future. We note, as Schroeder and Craig (2021) did, that systematic lines of research rather than 'one-shot' studies are what are needed in the field.

6.2. Implications for practice

Based on the evidence in this review, we have a few conclusions that may help guide practitioners as they implement PAs for K-12 students. First, there seems to be no detriment to letting learners choose the appearance of their PA. Our results showed that the instructional strategies used by the PA influenced learning, while the appearance of the PA only influenced learners' *preferences*. So, we feel that in many cases it makes sense for instructional designers to design PAs to use pedagogical strategies they believe to be effective while letting the learners' select what that PA may look like. Using a strategy such as this may help learners feel engaged, interested, motivated, and efficacious, while the designer knows they have implemented pedagogically strong instructional materials.

To expand on this, our review highlighted the continued importance of designing PAs that implement effective pedagogical methods. Speaking broadly, PAs were effective when they were able to use pedagogical techniques that are also effective for human teachers. In other words, PAs should be designed to add pedagogical value to the learning environment rather than merely being present.

Finally, for those designing PAs, our review showed that there are notable overlaps in the design space between PA function and the medium level of design. For example, a PA that functions as a peer may look and behave differently than a PA that functions as an instructor. Consequently, we found Heidig and Clarebout's (2011) frameworks to be quite useful from a PA design perspective and we continue to believe that designers should follow them to ensure they have considered many aspects of their PAs implementation.

7. Limitations and directions for future research

The most notable limitation of this study is actually a reflection of the field: there have been few systematic lines of research investigating PA design specifically among K-12 learners. We hope that since PAs are once again capturing the attention of researchers and creating them is more accessible than it's ever been, future researchers will aim to conduct systematic lines of inquiry around the use of PAs in K-12 contexts.

Another limitation of our study was that it was at times challenging to draw generalizable conclusions about the intersection of agent function and the medium level of design due to these components being intertwined. For future research, we think it is necessary to distinguish between a PA's function as instructor or peer (What does the PA do? PACU: function of the PA) and a PA's role to appear as an instructor or peer (What does the PA look like? PALD: medium design level). Using the terms "function" and "appearance" rather than "role" can help clarify these ambiguities. We also note that throughout this review we added two new features to the global level of the PALD framework, adding agent choice and customization as global level decisions. We suspect that as technology advances, the PACU and PALD models will need to adapt as well.

One of the most promising areas of work we found in this review is the intersection of student characteristics and PA design. While there has been some work in this area, there has not been much. The increasing diversity in K-12 public schools in the United States, encompassing culture, ethnicity, and language (NCES, 2023), poses a significant challenge in designing PAs that can effectively engage with all learners. While PAs have been utilized to enhance learner interaction in virtual learning environments, there is a scarcity of information on their impact on the learning experience of diverse students (Do et al., 2023). Therefore, we propose that exploring how a learner's background, culture, and demographic characteristics can be leveraged to mediate how PAs might engage with diverse learners to support learning is essential. The incorporation of culture in the design of embodied PAs can draw from principles of culturally relevant (Ladson-Billings, 1995), responsive (Gay, 2018), and sustaining (Paris, 2012) pedagogies. These asset-based approaches aim to recognize, value, and strengthen diverse students' language, literacy, and cultural practices (Paris, 2012).

Building on this, we note that PAs can be culturally adaptive, which does not imply designing them to know everything about all different cultures. Instead, the PA could be designed to be aware of the learners' cultural, demographic, and cognitive background, respectful of cultural and linguistic differences, sensitive to all students' linguistic and cultural needs, and flexible and able to adjust

when needed. In other words, PAs must be interculturally competent. *Intercultural competence* is the speakers' ability to shift their "cultural perspective and appropriately adapt behavior to cultural differences and commonalities" (Hammer, 2015, p. 483). Intercultural communication includes both verbal (e.g., what they say, how they say it) and non-verbal (e.g., gestures, postures, facial expressions, eye contact, voice, and tone) elements (Le Roux, 2002). The conversational patterns of PAs play a pivotal role in how learners interact with them and their perceptions of the PAs (Endrass & André, 2014). Mismatches in verbal and non-verbal messages in intercultural communication often led to misunderstanding, confusion, and frustration (Le Roux, 2002). Hence, it is critical to design PAs sensitive to the potential pitfalls of intercultural communication to prevent communication problems or intercultural conflicts. These communication problems could lead diverse learners to disengage in PA-mediated learning environments (López et al., 2021).

8. Conclusion

This systematic review examined how to design PAs for K-12 learners in various contexts. Our results largely reiterated Heidig and Clarebout's (2011) findings that we cannot exclude context as an important factor of PA design. We found that the appearance of the PA may not *directly* influence learning outcomes to the same extent as the pedagogy the PA embodies, yet we noted that learner characteristics may influence their perceptions and preferences of PA appearance and argued that this could directly influence the learning *experience*. To this end, we identified that systematic lines of research are needed around the intersection of learner characteristics, PAs, and the relationships between various learning-related outcomes (e.g., learners' motivation and perceptions) and learning. Overall, we conclude that our results indicate designing PAs with sound pedagogy is undoubtedly important, yet how to design the appearance of an effective PA for K-12 learners is still an open question.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

This material is based upon work supported by the National Science Foundation and the Institute of Education Sciences under Grant #2229612.

CRediT authorship contribution statement

Shan Zhang: Writing – review & editing, Writing – original draft, Formal analysis. **Chris Davis Jaldi:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. **Noah L. Schroeder:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Alexis A. López:** Writing – original draft. **Jessica R. Gladstone:** Writing – review & editing, Writing – original draft, Conceptualization. **Steffi Heidig:** Writing – review & editing, Methodology.

Data availability

All data used in the analyses are available as supplementary materials.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2024.105165.

References

*indicates study included in the Systematic Review Analysis

* Abdelghani, R., Oudeyer, P.-Y., Law, E., de Vulpillières, C., & Sauzéon, H. (2022). Conversational agents for fostering curiosity-driven learning in children. *International Journal of Human-Computer Studies, 167*, 1–14. https://doi.org/10.1016/j.ijhcs.2022.102887.

* Alaimi, M., Law, E., Pantasdo, K. D., Oudeyer, P.-Y., & Sauzeon, H. (2020). Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13). https://doi.org/10.1145/3313831.3376776.

* Alsharbi, B., & Richards, D. (2017). Using virtual reality technology to improve reality for young people with chronic health conditions. In *Proceedings of the 9th international conference on computer and automation engineering* (pp. 11–15). https://doi.org/10.1145/3057039.3057080.

Arguedas, M., & Daradoumis, T. (2021). Analysing the role of a pedagogical agent in psychological and cognitive preparatory activities. *Journal of Computer Assisted Learning, 37*(4), 1167–1180. https://doi.org/10.1111/jcal.12556.

Arroyo, I., Woolf, B. P., Royer, J. M., Tai, M., & English, S. (2010). In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (pp. 423–432). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13388-6_46, 6094.

Beal, C. R., Arroyo, I. M., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning, 9*(1), 64–77.

* Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2020). Does the effect of enthusiasm in a pedagogical agent's voice depend on mental load in the learner's working memory?. *Computers in Human Behavior, 112*, Article 106483. https://doi.org/10.1016/j.chb.2020.106483.

* Bringula, R. P., Fosgate, I. C. O., Garcia, N. P. R., & Yorobe, J. L. M. (2018). Effects of pedagogical agents on students' mathematics performance: A comparison between two versions. *Journal of Educational Computing Research, 56*(5), 701–722. https://doi.org/10.1177/0735633117722494.

* Carpenter, K. K. (2013). *Strategy instruction in early childhood math software: Detecting and teaching single-digit addition strategies*. Columbia University https://www.proquest.com/docview/1353399719/abstract/7047DA8172434478PQ/1.

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, Z.-H., & Chan, T.-W. (2008). Learning by substitutive competition: Nurturing my-pet for game competition based on open learner model. In *2008 second IEEE international conference on digital game and intelligent toy enhanced learning* (pp. 124–131). https://doi.org/10.1109/DIGITEL.2008.36.

* Chen, Z.-H., & Chen, S. Y. (2014). When educational agents meet surrogate competition: Impacts of competitive educational agents on students' motivation and performance. *Computers & Education, 75*, 274–281. https://doi.org/10.1016/j.compedu.2014.02.014.

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education, 146*, Article 103756. https://doi.org/10.1016/j.compedu.2019.103756

Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an animated pedagogical agent for learning English as a second language. *Journal of Educational Computing Research, 34*(4), 441–466. https://doi.org/10.2190/A064-U776-4208-N145

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education, 114*, 193–205. https://doi.org/10.1016/j.compedu.2017.07.003

Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers & Education, 190*, Article 104607. https://doi.org/10.1016/j.compedu.2022.104607

Dai, L., Kritskaia, V., Van Der Velden, E., Vervoort, R., Blankendaal, M., Jung, M. M., … Louwerse, M. M. (2024). Text-to-speech and virtual reality agents in primary school classroom environments. *Journal of Computer Assisted Learning, 13046*. https://doi.org/10.1111/jcal.13046.

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

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

Davis, R. O., Vincent, J., & Park, T. (2019). Reconsidering the voice principle with non-native language speakers. *Computers & Education, 140*, Article 103605. https://doi.org/10.1016/j.compedu.2019.103605

* Dincer, S., & Doganay, A. (2015). The impact of pedagogical agent on learners' motivation and academic success. *Practice and Theory in Systems of Education, 10*(4), 329–348. https://doi.org/10.1515/ptse-2015-0032.

* 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.

Do, T. D., Zelenty, S., Gonzalez-Franco, M., & McMahan, R. P. (2023). Valid: A perceptually validated virtual avatar library for inclusion and diversity. *Frontiers in Virtual Reality, 4*, Article 1248915. https://doi.org/10.3389/frvir.2023.1248915

Domagk, S., & Niegemann, H. M. (2005). Pedagogical agents in multimedia learning environments: Do they facilitate or hinder learning?. *Proceedings of the 2005 conference on towards sustainable and scalable educational innovations informed by the learning sciences*. Sharing Good Practices of Research, Experimentation and Innovation.

Endrass, B., & André, E. (2014). *Integration of cultural factors into the behavioural models of virtual characters*.

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. teachers college press.

* Girard, S., & Johnson, H. (2010). What do children favor as embodied pedagogical agents? In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 307–316). Springer. https://doi.org/10.1007/978-3-642-13388-6_35.

* Grynszpan, O., Martin, J.-C., & Nadel, J. (2008). Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human-Computer Studies, 66*(8), 628–639. https://doi.org/10.1016/j.ijhcs.2008.04.001.

* Haake, M., & Gulz, A. (2009). A look at the roles of look & roles in embodied pedagogical agents—a user preference perspective. *International Journal of Artificial Intelligence in Education, 19*(1), 39–71.

Hammer, M. R. (2015). Intercultural competence development. *The SAGE Encyclopedia of Intercultural Competence, 2*, 483–486.

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

* Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education, 48*(4), 523–547. https://doi.org/10.1016/j.compedu.2005.02.007.

* Jaques, P. A., Lehmann, M., & Pesty, S. (2009). Evaluating the affective tactics of an emotional pedagogical agent. In *Proceedings of the 2009 ACM symposium on applied computing* (pp. 104–109). https://doi.org/10.1145/1529282.1529304.

* Jing, B., Liu, J., Gong, X., Zhang, Y., Wang, H., & Wu, C. (2022). Pedagogical agents in learning videos: Which one is best for children?. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2022.2141787.

* Johnson, A. M., Ozogul, G., Moreno, R., & Reisslein, M. (2013). Pedagogical agent signaling of multiple visual engineering representations: The case of the young female agent. *Journal of Engineering Education, 102*(2), 319–337. https://doi.org/10.1002/jee.20009.

* Johnson, A. M., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*(2), 97–115.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S. (2009). The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58–80). IGI Global.

* Kautzmann, T. R., & Jaques, P. A. (2018). Improving the metacognitive ability of knowledge monitoring in computer learning systems. In A. Cristea, I. Bittencourt, & F. Lima (Eds.), *Higher education for all. From challenges to novel technology-enhanced solutions* (pp. 124–140). Springer International Publishing. https://doi.org/10.1007/978-3-319-97934-2_8.

* Kautzmann, T. R., & Jaques, P. A. (2019). Effects of adaptive training on metacognitive knowledge monitoring ability in computer-based learning. *Computers & Education, 129*, 92–105. https://doi.org/10.1016/j.compedu.2018.10.017.

Kim, Y. (2009). The role of learner attributes and affect determining the impact of agent presence. *International Journal of Learning Technology, 4*(3–4), 234–249. https://doi.org/10.1504/IJLT.2009.028808.

* Kim, Y., & Lim, J. H. (2013). Gendered socialization with an embodied agent: Creating a social and affable mathematics learning environment for middle-grade females. *Journal of Educational Psychology, 105*(4), 1164–1174. https://doi.org/10.1037/a0031027.

* Kim, Y., & Wei, Q. (2011). The impact of learner attributes and learner choice in an agent-based environment. *Computers & Education, 56*(2), 505–514. https://doi.org/10.1016/j.compedu.2010.09.016.

* Kim, Y., Wei, Q., Xu, B., Ko, Y., & Ilieva, V. (2007). Mathgirls: Toward developing girls' positive attitude and self-efficacy through pedagogical agents. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 119–126). IOS Press.

Kizilkaya, G., & Askar, P. (2008). The effect of an embedded pedagogical agent on the students' science achievement. *Interactive Technology and Smart Education, 5*(4), 208–216. https://doi.org/10.1108/17415650810930893.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491.

Le Roux, J. (2002). Effective educators are culturally competent communicators. *Intercultural Education, 13*(1), 37–48.

* Lester, J., Converse, S., Stone, B., Kahler, S., & Barlow, S. (1997). Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. In B. duBoulay, & R. Mizoguchi (Eds.), *Artificial intelligence in education* (pp. 23–30).

López, A. A., Guzman-Orth, D., Zapata-Rivera, D., Forsyth, C. M., & Luce, C. (2021). Examining the accuracy of a conversation-based assessment in interpreting English learners' written responses. *ETS Research Report Series, 2021*(1), 1–15. https://doi.org/10.1002/ets2.12315

* Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning, 35*(3), 349–358. https://doi.org/10.1111/jcal.12335.
Mei, C. (2016). *Improving virtual reality asd intervention games with 3dui and 3d virtual humans*. The University of Texas at San Antonio.
* Mei, C., Mason, L., & Quarles, J. (2015a). How 3d virtual humans built by adolescents with asd affect their 3d interactions. *Proceedings of the 17th international ACM SIGACCESS Conference on computers & accessibility* (pp. 155–162). https://doi.org/10.1145/2700648.2809863.
* Mei, C., Mason, L., & Quarles, J. (2015b). "I built it!"—exploring the effects of customizable virtual humans on adolescents with ASD. *2015 IEEE virtual reality (VR)* (pp. 235–236). https://doi.org/10.1109/VR.2015.7223382.
* Mei, C., Zahed, B. T., Mason, L., & Quarles, J. (2018). Towards joint attention training for children with asd—a vr game approach and eye gaze exploration. *2018 IEEE Conference on virtual Reality and 3D user interfaces (VR)* https://doi.org/10.1109/VR.2018.8446242.
* Molenaar, I., Roda, C., van Boxtel, C., & Sleegers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning environment. *Computers & Education, 59*(2), 515–523. https://doi.org/10.1016/j.compedu.2011.12.006.
Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*(2), 186–207.
* Murray, M., & Tenenbaum, G. (2010). Computerized pedagogical agents as an educational means for developing physical self-efficacy and encouraging activity in youth. *Journal of Educational Computing Research, 42*(3), 267–283. https://doi.org/10.2190/EC.42.3.b.
NCES. (2023). Racial/Ethnic enrollment in public schools. *National Center for Education Statistics*. https://nces.ed.gov/programs/coe/pdf/2023/cge_508.pdf.
* Nguyen, H. (2022). Let's teach kibot: Discovering discussion patterns between student groups and two conversational agent designs. *British Journal of Educational Technology, 53*(6), 1864–1884. https://doi.org/10.1111/bjet.13219.
* Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education, 67*, 36–50. https://doi.org/10.1016/j.compedu.2013.02.006.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71
Pane, J., Mccaffrey, D., Slaughter, M., Steele, J., & Ikemoto, G. (2010). An Experiment to Evaluate the Efficacy of Cognitive Tutor Geometry. *Journal of Research on Educational Effectiveness, 3*, 254–281. https://doi.org/10.1080/19345741003681189.
Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97. https://doi.org/10.3102/0013189X12441244
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, 60*(3), 547–578. https://doi.org/10.1177/07356331211041701
* Pérez-Marín, D., & Pascual-Nieto, I. (2013). An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology, 32*(9), 955–964. https://doi.org/10.1080/0144929X.2012.687774.
* Riedmann, A., Schaper, P., Jakob, B., & Lugrin, B. (2022). A theory based adaptive pedagogical agent in a reading app for primary students—a user study. In S. Crossley, & E. Popescu (Eds.), *Intelligent tutoring systems* (pp. 276–292). Springer International Publishing. https://doi.org/10.1007/978-3-031-09680-8_26.
Sahimi, S. M., Zain, F. M., Kamar, N. A. N., Samar, N., Rahman, Z. A., Majid, O., … Luan, W. S. (2010). The pedagogical agent in online learning: Effects of the degree of realism on achievement in terms of gender. *Contemporary Educational Technology, 1*(2), 175–185. https://doi.org/10.30935/cedtech/5973
Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (castle). *Educational Psychology Review, 34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5
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
Schroeder, N. L., & Craig, S. D. (2021). Learning with virtual humans: Introduction to the special issue. *Journal of Research on Technology in Education, 53*(1), 1–7. https://doi.org/10.1080/15391523.2020.1863114
Schroeder, N. L., Zhang, S., Jaldi, C., Gladstone, J. R., López, A. A., & Dorley, E. (under review). Anonymized for blind review. Virtual characters help k-12 students learn and improve motivation: A meta-analysis.
Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2022). Voice in human–agent interaction: A survey. *ACM Computing Surveys, 54*(4), 1–43. https://doi.org/10.1145/3386867
* Sedlacek, L., Kostyuk, V., Labrum, M., Mulqueeny, K., Petronella, G., & Wiltshire-Gordon, M. (2017). Pedagogical voice in an e-learning system: Content expert versus content novice. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies. Novel learning ecosystems* (pp. 262–272). Springer International Publishing. https://doi.org/10.1007/978-3-319-58509-3_21.
Siegle, R. F., Schroeder, N. L., Lane, H. C., & Craig, S. D. (2023). Twenty-five years of learning with pedagogical agents: History, barriers, and opportunities. *TechTrends, 67*(5), 851–864. https://doi.org/10.1007/s11528-023-00869-3
* Stelling, G. D. (2002). *Affective behavior control for lifelike pedagogical agents*. North Carolina State University.
* Tärning, B., & Silvervarg, A. (2019). "I didn't understand, i'm really not very smart"—how design of a digital tutee's self-efficacy affects conversation and student behavior in a digital math game. *Education Sciences, 9*. https://doi.org/10.3390/educsci9030197.
Wang, Y., Gong, S., Cao, Y., Lang, Y., & Xu, X. (2023). The effects of affective pedagogical agent in multimedia learning environments: A meta-analysis. *Educational Research Review, 38*, Article 100506. https://doi.org/10.1016/j.edurev.2022.100506
Wang, F., Li, W., Xie, H., & Liu, H. (2017). Is pedagogical agent in multimedia learning good for learning? A meta-analysis. *Advances in Psychological Science, 25*(1), 12. https://doi.org/10.3724/SP.J.1042.2017.00012
* Wei, Q. (2010). *The effects of pedagogical agents on mathematics anxiety and mathematics learning*. Utah State University.
* Wiggins, J. B. (2021). *Examining dialogue initiative policies for conversational pedagogical agents in game-based learning environments*. University of Florida.
* Xu, M. (2009). *An investigation of the effectiveness of intelligent elaborative feedback afforded by pedagogical agents on improving young Chinese language learners' vocabulary acquisition*. Northern Illinois University.
* Yalçin, Ö. N., Lalle, S., Conati, C., & Assoc Comp Machinery. (2022). An intelligent pedagogical agent to foster computational thinking in open-ended game design activities. *Intelligent user interfaces*, 633–645. https://doi.org/10.1145/3490099.3511124.
Zhang, S., Jaldi, C., Schroeder, N. L., & Gladstone, J. R. (2024). Pedagogical agents in K-12 education: A scoping review. *Journal of Research on Technology in Education*, 1–28. https://doi.org/10.1080/15391523.2024.2381229.

Further reading

* Maizura Sahimi, S., Zain, F. M., Kamar, N. A. N., Samar, N., Rahman, Z. A., Majid, O., Atan, H., Fook, F. S., & Luan, W. S. (2010). The pedagogical agent in online learning: Effects of the degree of realism on achievement in terms of gender. *Contemporary Educational Technology, 1*(2), 175–185.
* Sao Pedro, M. A., Gobert, J. D., & Baker, R. S. (2014). The impacts of automatic scaffolding on students' acquisition of data collection inquiry skills. *Grantee Submission*.