Education and Information Technologies
https://doi.org/10.1007/s10639-025-13328-8

Can AI-generated pedagogical agents (AIPA) replace human teacher in picture book videos? The effects of appearance and voice of AIPA on children's learning

Jie Bai[1] · Xiulan Cheng[1] · Hui Zhang[1] · Yihang Qin[2] · Tao Xu[2] · Yun Zhou[1]

Received: 13 June 2024 / Accepted: 2 January 2025

Abstract
The advancement of AI-generated content (AIGC) has made creating pedagogical agents (PAs) for multimedia learning increasingly realistic, simpler, and more efficient. However, little is known about the acceptance of AI-generated pedagogical agents (AIPAs) in picture book videos among young children aged 3–6. To address this, the present study examines the effects of AIPA appearance and voice on children's reading performance, using a 2 × 2 between-subjects design with varied combinations of real and AI-generated voices and appearances. Children learned from one of the following picture book videos: (1) a real teacher's voice and appearance, (2) a real teacher's voice and AI-generated appearance, (3) an AI-synthesized voice with a real appearance, or (4) both AI-synthesized voice and appearance. The results revealed no significant difference in reading performance between the AI teacher and the real teacher. Eye-tracking data indicated that AIPA appearance and voice did not increase cognitive load, and children expressed a comparable preference for AIPAs and human teachers. These findings support the equivalence principle in children's multimedia learning. While AIPAs may lack human micro-expressions and intonation nuances, they hold promise as complementary tools in early education.

Keywords AI-generated content (AIGC) · Pedagogical agents (PA) · Appearance · Voice · Picture book videos

Extended author information available on the last page of the article

Published online: 09 January 2025

1 Introduction

Children aged 3 to 6 years undergo rapid language and cognitive development (Weinert, 2022), and picture books are crucial in supporting this growth (Crawford et al., 2024). Recent advancements in multimedia have transformed picture books from physical to digital formats. As digital natives, young children increasingly engage with electronic devices, and show a growing preference for digital picture books (Bus et al., 2020). Digital picture books and videos integrate visuals and text, supporting children's comprehension in diverse ways (Mayer, 2009). Studies indicate that digital picture books improve reading skills and engage children more with visual and auditory elements (Korat et al., 2021; Masataka, 2014). For example, young readers spend 77.2% of their time on images, engaging nearly 3.79 times more with visuals than text (Liao et al., 2020), and five times more with visuals than print (Arslan-Ari & Ari, 2021). Combining text, images, and sound supports story comprehension and vocabulary acquisition (Bus et al., 2019; Korat et al., 2021; Li & Bus, 2023). Dynamic images with music and sound improve comprehension, even outperforming paper books enhanced with storyline-supporting technology (Furenes et al., 2021).

Pedagogical Agents (PAs) are virtual characters in multimedia learning environments, providing cognitive and emotional support to facilitate learning (Veletsianos & Russell, 2014). Studies show that PAs, combining text, speech, and images, improve learning (Castro-Alonso et al., 2021; Martha & Santoso, 2019). Animated pedagogical agents (APAs) have been found to reduce cognitive load, boost motivation, and make learning more enjoyable (Johnson et al., 2015). However, the impact of PAs on young children aged 3–6 years in preschool education remains unclear (Dai et al., 2022).

The rapid advancement of generative AI and AIGC has become increasingly integrated into daily life and work (Pataranutaporn et al., 2021), impacting even young children's activities. Platforms like YouTube and TikTok now include AIGC, like digital picture books and language learning videos for children (Netland et al., 2025; Xu et al., 2024). These videos use generated PAs with realistic facial expressions and behaviours to simulate human teachers. The social agency theory (Martha & Santoso, 2019) explains the effectiveness of PAs, as they create a sense of interpersonal connection, boosting learner engagement, trust, and credibility (Kim et al., 2022). The equivalence principle (Horovitz & Mayer, 2021) suggests virtual instructors can perform similar to humans, and research has shown that AIPAs can improve learner engagement, motivation, and performance (Pi et al., 2022), particularly when designed with a likable appearance. In language learning, they can improve retention and knowledge transfer (Deng et al., 2022). According to Piaget's theory of the Preoperational Stage (Piaget, 1968), children aged 3–6 learn through symbolic play, imitation, and direct experiences, with increasing needs for social interaction and language development. Multimodal AI technologies can help meet these needs, such as helping children understand how to engage with interactive elements (Djonov et al., 2021). AI robots can support embodied learning in young children (Yang et al., 2023). However, research on AI applications for children aged 3–6 and the interaction of young children with AI tools remains limited (e.g., Dai et al., 2022; Lawson & Mayer, 2022). While AI tools like chatbots engage children, build trust (Xu, 2023),

and stimulate curiosity (Abdelghani et al., 2022), challenges include ensuring content is effective and age-appropriate. AI in early childhood education must address concerns such as the uncanny valley effect (Mori et al., 2012) and ensure expert review to avoid negative impacts. Further research is essential to understand AIGC's impact on young children's development.

This study aims to investigate whether the synthesized voice and human-like appearances of AIPA in digital picture books enhance children's reading. First, we examine the impact of synthesized versus human voices on reading performance. Second, we compare the effects of dynamic appearances of AI teachers with that of human teachers. The findings could influence education policy and technology development. For developers, understanding AIPAs' impact can guide the design of adaptive and age-appropriate educational software, promoting innovative AI solutions for early education.

2 Relevant research

2.1 The impact of digital picture books on young children's reading

Reading is a complex cognitive process involving print decoding and language comprehension, as outlined in the Simple View of Reading (SVR) model (Hoover & Gough, 1990). This model divides reading ability into reading comprehension and retelling, with two levels of cognitive processing: decoding symbols (outside-to-inside) and encoding language comprehension (inside-to-outside). According to multimodal cognitive learning theory (Mayer, 2003) and dual coding theory (Paivio, 2007), digital picture books can enhance young children's reading abilities. Digital picture books have been found to enhance young children's story comprehension compared to traditional ones (Takacs et al., 2015). Scanned static picture books have been shown to improve print awareness and vocabulary acquisition (Fathi, 2014). Interactive digital picture books further promote story comprehension and vocabulary expression (Takacs & Bus, 2016). Virtual reality picture books have also been demonstrated to improve reading comprehension (Danaei et al., 2020). Digital picture books are particularly beneficial for children from low socioeconomic backgrounds, those with special needs, or at risk (Shamir & Korat, 2015). However, digital books have been criticized for potentially reducing interaction between caregivers and young children (Munzer et al., 2019) and causing distractions (Takacs et al., 2015). To facilitate language learning, the design of interactive features should be instructional. Digital picture books should complement rather than replace quality interaction between adults and children (Kelley & Kinney, 2017).

Taken together, digital picture books and picture book videos with voice and image features offer an immersive experience, thereby enhancing children's interest and motivation to read. Our study is grounded on the context of picture book videos, given their importance in children's reading.

2.2 The effects of PA's appearance on learning

PAs plays a crucial role in learning by providing effective instruction and fostering an active learning experience (Castro-Alonso et al., 2021; Dai et al., 2022; Martha & Santoso, 2019). Studies indicate that a PA's appearance positively impacts on learners' behaviour, attitude and motivation (Pataranutaporn et al., 2021). PAs can be human-like (Deng et al., 2022), anthropomorphic (Sikorski et al., 2019), or animated (Li et al., 2022). Adult learners tend to prefer human-like appearances (Sikorski et al., 2019). Kuang et al. (2021) found that human teachers provide a better learning experience than animated ones in instructional videos. Likable human teachers (Pi et al., 2022) can also lead to better outcomes. When the learning content is challenging, the instructor's facial expressions significantly impact the learning (Wang et al., 2019). Shiban et al. (2015) found that PAs' appearances influence motivation and performance, subsequently shaping student interactions with PAs.

Children aged 3 to 6 experience rapid language and cognitive development (Weinert, 2022), with picture books playing a key role in fostering cognitive growth (Crawford et al., 2024). PAs, often designed as cartoon characters, are increasingly integrated into digital picture books as substitutes for human teachers. These agents engage children by redirecting their attention, providing guidance, and facilitating interaction (Jing et al., 2022; Xu et al., 2021a, b). Advances in AI have improved the realism and social interactivity of PAs, making them comparable instructors in effectiveness (e.g., Pi et al., 2022). However, the impact of PA appearance on young children's learning remains unclear, limiting optimal design of digital picture books. Therefore, further investigation is required to determine whether PA appearance enhances reading and learning for children aged 3 to 6.

2.3 The effects of PAs' voice on learning

The impact of voice on learning may be moderated by factors like naturalness and human-likeness of the voice (Seaborn et al., 2022). According to the voice effect, human voices are more effective than machine-synthesized voices in facilitating deep learning (Davis, 2018; Mayer, 2014; Mayer et al., 2003). The human voice also enhances the social cues of PAs, stimulating deep learning and improving near and far transfer performance. Learners in human-voice groups rate PAs more positively in social interactions (Atkinson et al., 2005; Mayer & DaPra, 2012).

Advances in AI text-to-speech technology has made PAs' voices nearly comparable to human voices. Chiou et al. (2020) found PAs' voices as effective as human voices in learning, with high-quality voices improving trust. Enthusiastic voices, characterized as "friendly," "energetic," and "exciting," improve social evaluations, knowledge transfer, and engagement compared to calm voices (Liew et al., 2020). They also boost recall performance, intrinsic motivation, and time estimation (Moè, 2016).

Voice in digital picture books supports children by directing attention to pictures over text (Skibbe et al., 2018). Reich et al. (2019) found no significant differences in attention and emotional engagement between digital and human reading, though children prefer human readers. Story comprehension remains similar whether books

were read by adults or on tablets (Zipke, 2017). However, Krcmar and Cingel (2014) found adult-read paper books better support comprehension than digital ones. While digital picture books can aid emergent literacy, skeptical attitude persists about their integration into learning (Bai et al., 2022).

Thus, studies show mixed conclusions on the impact of voices in digital picture books on children's reading. Empirical research is required to address this issue and evaluate the effectiveness of AIPA voices in enhancing picture book videos.

2.4 Using eye-tracking to understand cognitive load and attention in learning

A considerable number of studies have consistently used eye movement to analyze learning processes (Lai et al., 2013). Eye movement patterns are influenced by the cognitive resources required and the attention demands of the task (Holmqvist et al., 2011). Pupil diameter is used to measure cognitive load (Krejtz et al., 2018), which is defined as the average pupil diameter of the left and right pupil (Haro et al., 2022). Research on young children's reading highlights fixation duration as a visual attention indicator (Liao et al., 2020; Takacs & Bus, 2016), and pupil diameter as a cognitive load indicator (Ozeri-Rotstain et al., 2020). To evaluate children's visual attention in the reading process, the Time-to-First-Fixation (TFF) was used, along with the Ratio of Total Fixation Duration (RTFD), reflecting the immediate impact of visual attention on reading (Liao et al., 2020).

2.5 The present study

This study aims to investigate the influence of the appearance and voice of AIPAs on children's reading. Based on the equivalence principle, this study assumes that the voice and appearance of AIPAs have an equivalent impact on young children's reading compared to a human teacher. The hypotheses are as follows: The AI-generated instructor's appearance and voice will have the same impact on children's performance, including reading comprehension and retelling, as the human teacher (Hypothesis 1). Additionally, the AI-generated instructor's appearance and voice will not increase children's cognitive load (Hypothesis 2), and their effect on children's attention will be equivalent (Hypothesis 3). It is expected that children will prefer the appearance and voice of the human teacher over the AI-generated instructor (Hypothesis 4), and the interaction effect in both voice and appearance will be equivalent (Hypothesis 5).

The research questions are as follows:

RQ1 What is the impact of the AI-generated instructor's appearance and voice on children's learning performance? (H1)
RQ2 What is the impact of the AI-generated instructor's appearance and voice on children's cognitive load? (H2)
RQ3 What is the impact of the AI-generated instructor's appearance and voice on children's attention? (H3)
RQ4 What appearances and voices of PAs do the children prefer? (H4)
RQ5 Do appearance and voice interact in their impact on learning? (H5)

3 Method

3.1 Participants

The experiment included 80 senior kindergarten children (43 girls, 53.7%; 37 boys, 46.3%), aged 5.5 to 6.5 years ($M = 5.84$, $SD = 0.30$), with one child declining participation. All participants were not with vision or hearing issues reported by parents. Ethical approval was obtained from our institution, and informed consent was provided by parents. Kindergarten teachers confirmed normal development of all participants. Using G*power3.1 (Faul et al., 2009), a required sample size of 73 was calculated for this 2×2 between-subjects design (effect size $f > 4$, power = 0.8, alpha = 0.05) (Peng & Wang, 2022). The final sample size of 80 met this requirement.

3.2 Design

We adopted the Wav2Lip model, incorporating Text-to-Speech (TTS) technology, to generate lip-syncing videos that seamlessly convert text into human voices. As shown in Fig. 1, automated virtual teacher generation is divided into two steps, the first step involves converting text to speech, and the second utilizes lip synthesis technology to create the video (Xu et al., 2021a, b). In this study, the virtual teachers were generated using the same photo of the human teacher.

As shown in Table 1, all participants were randomly assigned to one of the following four conditions:

(1) Human teacher with human voice (HT): A recorded video of the teacher reading the picture book;
(2) Virtual teacher with AI voice (VTA): Generated appearance based on the human teacher's photo, and the synthesized voice;

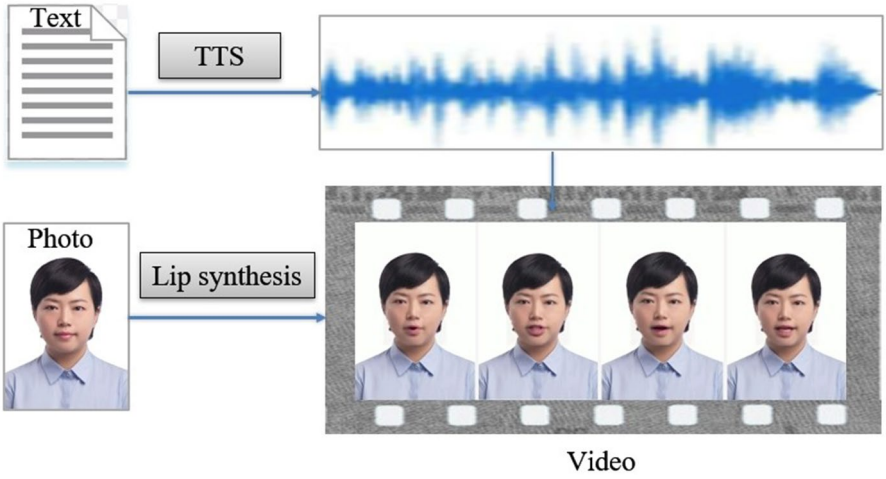

Fig. 1 The work flow of virtual teacher generation

Table 1 The differences in the teacher's appearance and voice across each experimental condition

Condition	Appearance	Voice
Group 1: Human teacher with human voice (HT)	Real teacher	Real human voice
Group 2: Virtual teacher with AI voice (VTA)	AI-generated teacher	AI voice
Group 3: Human teacher with AI voice (HTA)	Real teacher	AI voice
Group 4: Virtual teacher with human voice (VTH)	AI-generated teacher	Real human voice

Table 2 Ranking of children's familiarity with picture books

Picture book title	Familiarity	Rank
A Crooked Tree	0%	1
Maybe look	4%	2
I'm a rainbow fish	9%	3
Madeline Finn and the Library Dog	9%	4
The kiss that missed	11%	5
When a dragon lives in a sandcastle	11%	6
Nibbles the Book Monster	16%	7
I don't know who I am	17%	8
The giving tree	23%	9
The Kissing Hand	31%	10

(3) Human appearance with AI voice (HTA): A recorded video of the teacher with the synthesized voice;
(4) Virtual teacher with human voice (VTH): Generated appearance based on the human teacher's photo, with the human voice.

3.3 Learning materials

First, four kindergarten teachers screened 10 picture books suitable for 5-6-year-old children to ensure the appropriateness of the learning materials. Then, a familiarity assessment followed, where children indicated prior exposure by raising hands. As shown in Table 2, 0% familiarity indicates no prior exposure. Finally, *A Crooked Tree* (Lasa, 2022) was chosen as the experimental picture book, with all children confirming they had not read it.

3.4 Measures

3.4.1 Pretest

All participants took the Peabody Picture Vocabulary Test Revised (PPVT-R) (Dunn & Dunn, 1981) before the experiment. This norm-referenced test, consisting of 125 items, measures children's receptive language skills. For each item, children select the correct picture corresponding to a spoken word (e.g., "which one is a bus?"). Correct answers score 1 point, and incorrect answers score 0. The test stops after eight

consecutive errors. The PPVT-R showed high reliability for 3-6-year-olds, with a Cronbach's coefficient alpha of 0.94 calculated in this study.

3.4.2 Learning performance

The questionnaire to measure children's story comprehension consisted 14 items. It was divided into (1) the recall of story (e.g., characters, causes) and (2) the prediction (guessing what happens next in the story). Nine items tested recall and five items tested prediction. For each item, children chose between two options (Strouse et al., 2022) (e.g., "Who saved the tree in the story?"), with one point for a correct answer and zero for an incorrect one. The total score was the sum of correct responses, with a maximum of 14 points. Reliability analysis using the Spearman-Brown formula (Eisinga et al., 2013) showed moderate internal consistency $r_{kk} = 0.45$ (LeBreton & Senter, 2008).

Ten items were adopted to assess young children's macro-narrative ability using the Story Grammar (SG) model, based on the Edmonton Narrative Specification Tool. These items covered context, initiating event, internal response, plan, attempt, outcome, and response, collecting language information from children aged 4 to 9 through story retelling (Schneider et al., 2005). Core elements (initiating event, attempt and outcome) were scored 0–2 points; others were scored 0–1. Children retold the story, and two evaluators assessed their retelling video. Disagreements were resolved by a third evaluator. The evaluators' internal consistency was high, with a Pearson correlation coefficient $r = 0.69$ and Cronbach's $\alpha = 0.81$.

3.4.3 Correlations between reading comprehension, retelling and PPVT

Table 3 shows a significant positive correlation between reading comprehension, PPVT and story retelling, suggesting that both reading comprehension and story retelling effectively measured children's receptive language skills. While PPVT correlated with reading comprehension, no significant correlation was found between story retelling and PPVT, indicating that story retelling reflects expressive language skills in reading.

Table 3 Correlations among the tests

	Reading Comprehension	RCR	RCP	Story Retelling	PPVT	M	SD
Reading Comprehension	1	0.887**	0.757**	0.286*	0.522**	10.16	1.831
RCR (Recall)		1	0.369**	0.290**	0.430**	6.84	1.287
RCP (Prediction)			1	0.165	0.442**	3.33	0.911
Story Retelling				1	0.157	5.91	3.101
PPVT					1	85.64	18.268

Note Pearson correlation coefficients and significance levels reported

$p < 0.05$ denoted as *, $p < 0.01$ denoted as **

3.4.4 Eye movement analysis tools and indicators

The study employed the Tobii Nano, a screen-based eye tracker that captures gaze data at 60 Hz, paired with a Lenovo computer monitor. The recording monitor was a Lenovo T2224rF 21.5-inch (31.26×51.29 cm) widescreen LCD with a resolution ratio of 1920×1080, and the distance from the children to the screen was maintained at 70–80 cm. Cognitive load was measured using pupil diameter, calculated as the average diameter of the left and right pupils, with larger diameters indicating higher load. Attention was measured by two indicators: Time-to-First-Fixation (TFF) and the Ratio of Total Fixation Duration (RTFD). TFF measures the time it takes for children to fixate on either the picture book area or the PA area, with shorter TFF indicating greater attractiveness. RTFD measures the proportion of time spent fixating on either the PA or picture book area, with higher RTFD indicating more attention allocation.

3.4.5 Children's preference

We asked the children to express their preference by giving the number of stars after watching the video, which was on a five-level scale.

3.5 Procedure

Data collection was conducted in a kindergarten and consisted of two phases (see Fig. 2). In the first phase, a baseline test of the PPVT was administered individually in a quiet classroom. The second phase involved an eye-tracking experiment where the children watched the picture book video one by one in a quiet classroom. Afterwards, the children immediately answered questions orally about their preference and reading comprehension of the picture book, and then retold the story in another classroom. The entire process was video-recorded and lasted for approximately 30 min.

3.6 Data-analysis

First, we examined whether there were individual differences across the four experimental groups by conducting a one-way ANOVA to compare the means of children's learning performance since the data were normally distributed. We employed η^2 to measure the effect sizes in ANOVAs (Small effect = 0.01; medium effect = 0.06; large effect = 0.14). Then, a UNIANOVA analysis was employed to investigate the main effects and interaction effects related to learning performance, cognitive load, and attention during the experimental process. In the study, p-values less than 0.05 were considered statistically significant.

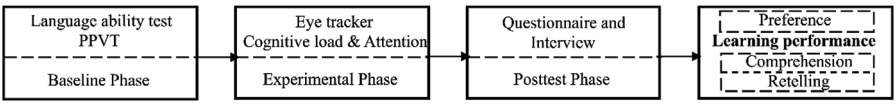

Fig. 2 The procedure of date collection

4 Results

The descriptive statistics for the dependent variables are presented in Table 4. The results indicated that participants exhibited a medium level of receptive language ability, with no significant differences observed across the four experimental groups. Upon comparing the reading performance, picture book preference, and eye movement characteristics of the children, the results showed no significant differences among the four groups.

4.1 Learning performance

The results indicated no significant differences in learning performance of reading comprehension and story retelling across the four groups (see Fig. 3). A UNIANOVA analysis of the interaction effects of voice and appearance on reading ability showed no significant interaction for reading comprehension ($F = 1.315$, $p = 0.255$, $\eta^2 = 0.017$) or story retelling ($F = 0.159$, $p = 0.691$, $\eta^2 = 0.002$).

4.1.1 Reading comprehension

The ANOVA test showed no significant difference in reading comprehension across experimental conditions ($F = 1.028$, $p = 0.385$, $\eta^2 = 0.039$). The results indicated that AIPAs with generated appearance and voice had the same influence on children's recall and prediction as the human teachers, although the HT and VTA groups had

Table 4 Means (*M*) and standard deviations (*SD*) of dependent variables across the four groups

Measures		Groups M (SD)			
		HT ($n=19$)	HTA ($n=21$)	VTA ($n=21$)	VTH ($n=19$)
Baseline test	PPVT	85.74(15.34)	82.48(18.43)	89.05(17.93)	80.79(29.00)
Learning performance	Reading Comprehension	10.32(2.21)	10.19(1.5)	10.24(1.89)	9.26(2.83)
	Recall	6.89(1.24)	6.86(1.15)	7.05(1.36)	6.11(2.02)
	Prediction	3.42(1.17)	3.33(0.66)	3.19(0.81)	3.16(1.26)
	Story retelling	5.16(3.72)	6.5(2.77)	6.57(2.82)	4.68(2.85)
Cognitive Load	Pupil diameter	3.11(0.34)	3.13(0.35)	3.03(0.32)	3.08(0.44)
Attention	TFF of picture book	1330.42 (1314.64)	1186.57 (1978.96)	880.81 (1473.66)	1033.11 (1776.11)
	TFF of PA	2459.58 (5316.42)	11,172 (26649.05)	14070.81 (43118.12)	12226.89 (43597.23)
	RTFD of picture book	0.67(0.15)	0.64(0.14)	0.66(0.15)	0.68(0.18)
	RTFD of PA	0.05(0.04)	0.07(0.05)	0.08(0.07)	0.06(0.05)
Preference		4.42(0.83)	4.52(0.81)	4.14(0.79)	4.11(1.29)

Note: RTDF is the Ratio of Total Fixation Duration, TFF is Time-to-First-Fixation. M: The values outside parentheses are means. SD: The values inside parentheses are standard deviations. HT, HTA, VTA, VTH, respectively, represent the real human teacher with human voice, human teacher with AI voice, virtual teacher with AI voice, virtual teacher with human voice

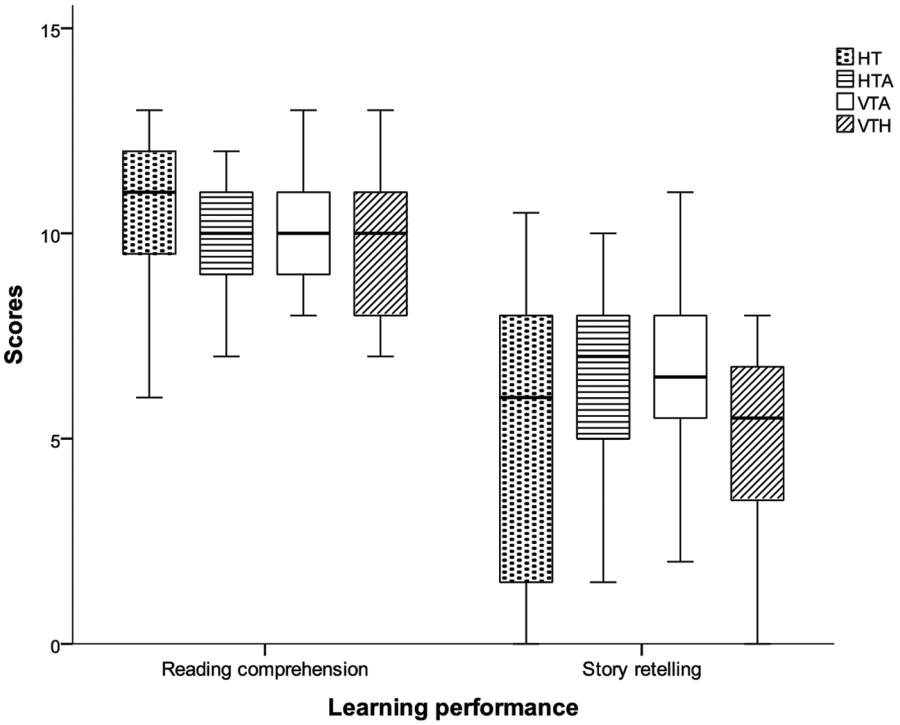

Fig. 3 The boxplot of learning performance

higher reading comprehension scores, especially for recall, with the VTA group scoring the highest.

4.1.2 Story retelling

The ANOVA test indicated no significant difference in story retelling across groups ($F=1.94$, $p=0.13$, $\eta^2 = 0.071$). The results suggested that AIPAs with generated appearance and voice had the same influence on the children's story retelling as the human teachers, although the VTA and HTA groups had higher scores, with the VTA group scoring highest.

4.2 Cognitive load and attention measured by eye tracking

A UNIANOVA analysis of cognitive load and attention showed no significant differences across the groups (see Table 5). These findings indicated that AIPAs with generated appearance and voice raised a similar cognitive load on the children's reading as the human teachers. However, the VTA and VTH groups experienced a lower cognitive load.

The main and interaction effects of voice and appearance were also not significant ($F=0.582$, $p=0.849$, $\eta^2 = 0.097$). The VTA group had the smallest pupil diam-

Table 5 Differences of attention and cognitive load across the four groups

	Groups M(SD)						
Indicators	HT (n=19)	HTA (n=21)	VTA (n=21)	VTH (n=19)	F	p	η^2
COGNITIVE LOAD							
Pupil diameter	3.11(0.34)	3.13(0.35)	3.03(0.32)	3.08(0.44)	0.288	0.834	0.011
ATTENTION							
TFF of picture book	1330.42 (1314.64)	1186.57 (1978.96)	880.81 (1473.66)	1033.11 (1776.11)	0.273	0.845	0.011
TFF of PA	2459.58 (5316.42)	11,172 (26649.05)	14070.81 (43118.12)	12226.89 (43597.23)	0.456	0.714	0.018
RTFD of picture book	0.67(0.15)	0.64(0.14)	0.66(0.15)	0.68(0.18)	0.210	0.889	0.008
RTFD of PA	0.05(0.04)	0.07(0.05)	0.08(0.07)	0.06(0.05)	1.102	0.353	0.042

Note: RTDF is the Ratio of Total Fixation Duration, TFF is Time-to-First-Fixation

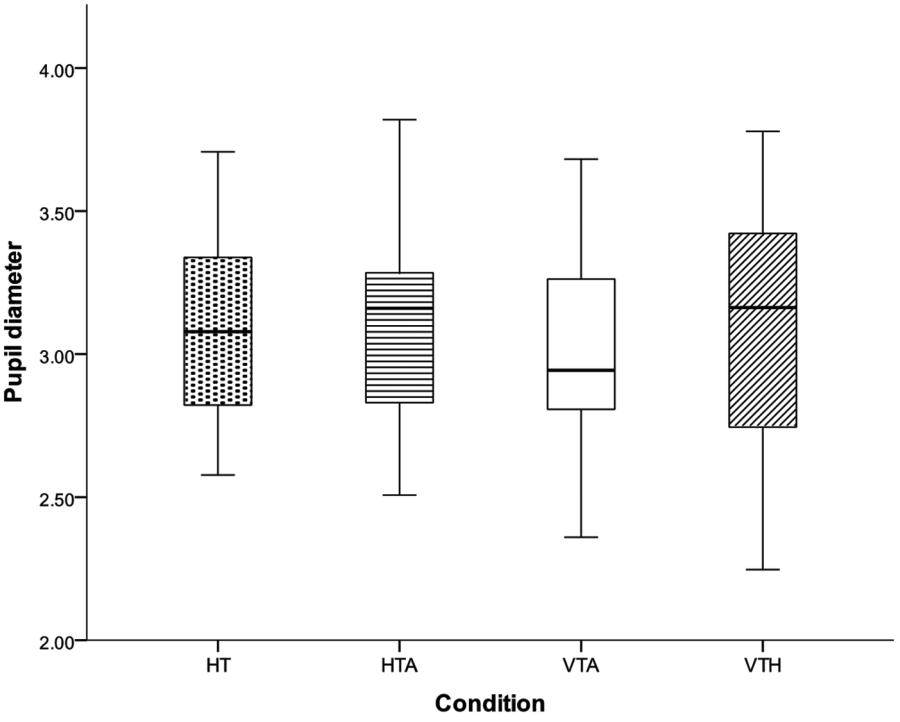

Fig. 4 The boxplot of cognitive load indicated by pupil diameter

eter (M=3.03, SD=0.32), indicating the least cognitive load, while the HTA group showed the most cognitive load (see Fig. 4).

Figure 5 shows attention measured by TFF (Fig. 5: left) and RTFD (Fig. 5: right). The results indicated that children in the VTA group focused more on the PA area, while those in the HT and HTA groups focused more on the picture book area. The VTA group had the shortest TFF, showing the most interest in the picture book.

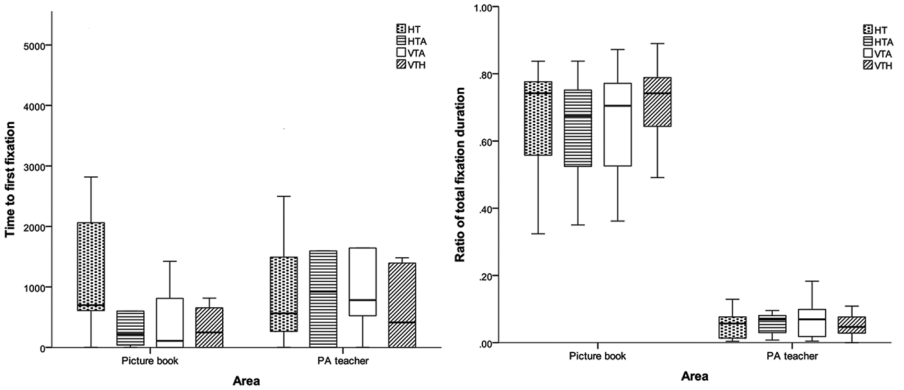

Fig. 5 The boxplot of attention indicated by (left: TTF, right: RTFD)

Regarding RTFD, the VTA and VTH group spent the longest time on the picture book, especially the VTA group had the longest time in the PA area of the picture book.

4.3 Children's preferences

A one-way ANOVA compared children's preferences for PA picture books across four groups. The results revealed that 44 children (55%) gave 5-star ratings, while only 2 (2.5%) gave 2-star ratings. Overall, the children generally preferred all types of PA picture books, especially those with a human teacher and AI voice. The preference order was: HTA ($M=4.52$, $SD=0.81$)>HT ($M=4.42$, $SD=0.83$)>VTA ($M=4.14$, $SD=0.79$)>VTH ($M=4.11$, $SD=1.29$), however, the results indicated no significant differences between groups ($F=0.954$, $p>0.05$, $\eta^2 = 0.036$; see Table 5). This suggests that children had the same preferences for the different voices and appearances of PAs in the picture books.

5 Discussion

This study investigated the impact of PAs' appearances and voices on young children's performance, cognitive load, attention and preferences in digital picture books. Our results showed that AIPAs facilitated young children's learning in the same way as human teachers, further supporting the equivalence principle of multimedia learning (Horovitz & Mayer, 2021). To our knowledge, this is one of the first studies to investigate the effects of AIPAs' appearances and voices on young children's reading. The results are important for guiding the design of integrating AI into digital learning resources for children.

RQ1 What is the impact of the AI-generated instructor's appearance and voice on children's learning performance?

The results showed no significant differences in learning performance between the appearance and voices of human and AI teachers. Both AIPAs and human teachers had similar effects on young children's learning in digital picture books, supporting the equivalence principle. This extends the principle's application to early childhood education, suggesting AIPAs can be as effective as human teachers. While prior research has validated this principle on college students (e.g., Pi et al., 2022; Xu et al., 2024), our study advances the scope by demonstrating its validity for children aged 3–6. From a technological perspective, current quality of AI technology is advanced enough to effectively mimic human appearances and voices, creating an engaging experience for young learners.

The effectiveness of AIPAs can be explained by the authentic impact of PA (Sikorski et al., 2019), and by social presence theory. With advanced AI technology, the AI teacher's human-like appearance and realistic micro-expressions convey social cues (Jing et al., 2022), fostering a feeling of engaging in a human-like dialogue for learners. When learners perceive AI teachers as a social presence, they engage more deeply, resulting in deeper cognitive processing (Mayer, 2009).

Additionally, AI voices were as effective as human voices in facilitating reading comprehension and even surpassed them in improving story retelling. Although AI voices are not yet as nuanced as human voices, their clear pronunciation is capable of achieving the same effect in facilitating children's comprehension as human speech. This finding is consistent with previous studies showing no difference between device and adult voices (Reich et al., 2019; Zipke, 2017). However, this contrasts with Mayer's voice effect theory, suggesting that human voices are better for deep learning (Mayer, 2014). This discrepancy may be due to advances in synthetic voice quality. We believe that, with technological advancements, future AIPAs will accurately mimic voice tones of teachers and parents, becoming powerful tools in children's multimedia learning.

RQ2 What is the impact of the AI-generated instructor's appearance and voice on children's cognitive load?

This study also examines cognitive states in early childhood reading by analysing eye movements. There were no statistically significant differences in cognitive load across the four groups. However, the VTA and VTH groups experienced lower cognitive load, likely due to the high standardized generated voice. Modern synthetic voices, closely mimicking human speech, provide clear articulation and consistent accents, aiding comprehension (Kaur & Singh, 2023). Although human instructors may receive training to enhance their vocal skills, they still exhibit variations in accent.

Additionally, unlike static PAs, our AI teacher is dynamic, displaying micro-expressions and social cues. These cues may help learners integrate images, sound, and text, boosting their attention and motivation, and cognitive resources allocation (Dinçer & Doğanay, 2017). The congruence between the voice and appearance in the VTA group also likely enhanced social presence, which may have led to a reduction in cognitive load compared to the HTA and VTH groups. These findings supported

Hypothesis 2 that AI-generated instructors' appearance and voice have the same impact on the children's cognitive load as the human teachers.

RQ3 What is the impact of the AI-generated instructor's appearance and voice on children's attention?

The results showed that children in the VTA groups had the shortest TFF first attention times, indicating that they were initially more attracted to the picture book with AI teacher's appearance and voice compared to the other groups. This may be due to the novelty effect of new technology associated with the VTA groups (Koch et al., 2018). However, this effect did not impact the learning performance. Both RTFD and TFF results showed no significant differences in attention allocation across the four conditions, suggesting AIPAs did not distract the children. This aligns with findings related to the embodiment effects (Mayer & DaPra, 2012), where a teacher's social cues improve engagement without distracting the students (Davis et al. 2023). Therefore, Hypothesis 3 is supported, indicating that the appearance and voice of an AI-generated instructor have the same impact on children's attention as a human teacher.

RQ4 What appearances and voices of PA do the children prefer?

The results revealed similar preferences across all four conditions, with the HTA group scoring the highest, indicating a preference for a human teacher with synthetic speech. This preference may be influenced by social presence (Schneider et al., 2022) and novelty. First, real teachers' micro-expressions and eye cues, likely caught children's attention. The combination of a real person with synthetic speech introduced a noticeable discord, more pronounced than pairing a synthetic character with a real voice. While this discord triggered curiosity, it did not lead to significant differences in preference.

RQ5 Do appearance and voice interact in their impact on learning?

We found no significant interaction between appearance and voice on learning performance, cognitive load, or preference. Although human appearance and voice are more realistic than AIPA, previous research (Oh et al., 2018) found that agents' visual realism and consistency in appearance and behaviour are essential in conveying social presence. Our hypothesis was that the realism consistency of PAs' appearance and voice would impact social presence. However, our results indicated no significant difference on learning outcomes, likely due to the close proximity in realism between human and AI teachers in this study.

6 Practical implications

This study explores the use of AIPAs children's multimedia learning, offering an illustrative example of their application. Results suggest that AI-generated appearances and voices can enhance learning outcomes comparable to human teachers in

digital picture book reading. We recommend that AIPAs could also be applied in interactive museum, math exercises, and language learning apps. These AIPAs can reduce production costs and allow human teachers to focus on instructional design. The trend towards integrating AIPA in multimedia learning is expected to increase significantly in the coming years. The use of AIPAs in early education is expected to grow, complementing rather than replacing human teachers.

However, challenges remain, such as improving AIPAs' social cues to replicate human interactions. While current AIPAs can display basic expressions, they lack subtle facial expressions and micro-expressions necessary for authentic engagement. Enhancing these aspects will make AIPAs more natural and effective. Additionally, adapting AIPAs to different cultural contexts is crucial to meet global educational expectations. Addressing these challenges will be key to making AIPAs a sustainable tool in children's education.

7 Conclusion, limitations and Future directions

This study aimed to bridge the digital divide in home literacy environments and alleviate the burden on early childhood educators by exploring the use of AIPAs in digital picture books. It provided empirical evidence on the use of AIPAs for young children and expanded the equivalence principle to early childhood education. The findings highlight the potential of AIPAs to enhance home-school cooperation in digital literacy.

However, this research has several limitations. First, the study's short duration calls for further research to confirm if the findings can be sustained and replicated. Second, individual differences, such as PPVT scores, were not considered, and future studies will examine how the levels of PPVT interact with PA characters to impact on learning. Finally, the study did not fully explore the role of social cues like tone, micro-expressions, and gestures in enhancing learning. Future research will evaluate how these social cues impact learning for students aged 3–6.

Author contributions Jie Bai, Xiulan Cheng, and Yun Zhou contributed to the study conception and design. Material preparation and software implementation were performed by Yihang Qin and Tao Xu, and data collection and analysis were performed by Jie Bai and Hui Zhang. The first draft of the manuscript was written by Jie Bai and Yun Zhou. All authors read and approved the final manuscript.

Funding This work was supported by the Shaanxi Provincial Social Science Fund Project [2022P036]; Research Projects of the National Natural Science Foundation of China under Grant [62077036] and [62377039]; National Key R&D Program of China under Grant [2022YFC3303600]; CAAI-CANN Open Fund, developed on OpenI Community; and Higher Education Research Fund of Northwestern Polytechnical University in 2025, No. GJJJM202504.

Data availability Data will be available from the corresponding author upon reasonable request.

Code Availability The application used in this study will be available from the corresponding author upon reasonable request.

Declarations

Ethical approval This research was approved by the Ethics Committee of the Faculty of Education at Shaanxi Normal University.

Consent to participate All participants were informed about this investigation, and their parents were also informed and signed an informed consent agreement.

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

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*, 102887. https://doi.org/10.1016/j.ijhcs.2022.102887

Arslan-Ari, I., & Ari, F. (2021). The effect of visual cues in e-books on pre-K children's visual attention, word recognition, and comprehension: An eye tracking study. *Journal of Research on Technology in Education*, *54*(5), 800–814. https://doi.org/10.1080/15391523.2021.1938763

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, *30*(1), 117–139. https://doi.org/10.1016/j.cedpsych.2004.07.001

Bai, J., Zhang, H., Chen, Q., Cheng, X., & Zhou, Y. (2022). Technical supports and emotional design in digital picture books a review. *Procedia Computer Science*, *201*, 174–180. https://doi.org/10.1016/j.procs.2022.03.025

Bus, A. G., Sarı, B., & Takacs, Z. K. (2019). The promise of multimedia enhancement in children's digital storybooks. *Reading in the Digital Age: Young Children's Experiences with E-books: International Studies with E-books in Diverse Contexts*, 45–57.

Bus, A. G., Neuman, S. B., & Roskos, K. (2020). Screens, apps, and digital books for young children: The promise of multimedia. *AERA Open*, *6*(1), 233285842090149. https://doi.org/10.1177/2332858420901494

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

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*, 103756. https://doi.org/10.1016/j.compedu.2019.103756

Crawford, P. A., Roberts, S. K., & Lacina, J. (2024). Picture books and young children: Potential, power, and practices. *Early Childhood Education Journal*, *52*, 1273–1279. https://doi.org/10.1007/s10643-024-01701-0

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*, 104607. https://doi.org/10.1016/j.compedu.2022.104607

Danaei, D., Jamali, H. R., Mansourian, Y., & Rastegarpour, H. (2020). Comparing reading comprehension between children reading augmented reality and print storybooks. *Computers & Education*, *153*, 103900. https://doi.org/10.1016/j.compedu.2020.103900

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., Park, T., & Vincent, J. (2023). A meta-analytic review on embodied pedagogical agent design and testing formats. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331221100556

Deng, L., Zhou, Y., Cheng, T., Liu, X., Xu, T., & Wang, X. (2022). My English teachers are not human but I like them: Research on virtual teacher self-study learning system in K12. In *International Conference on Human-Computer Interaction* (pp. 176–187). Cham: Springer International Publishing.

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

Djonov, E., Tseng, C. I., & Lim, F. V. (2021). Children's experiences with a transmedia narrative: Insights for promoting critical multimodal literacy in the digital age. *Discourse Context & Media*, *43*, 100493.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. American Guidance Service. Circle Pines, MN.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Fathi, M., I (2014). The effect of electronic books on enhancing emergent literacy skills of pre-school children. *Computers & Education*, *79*, 40–48. https://doi.org/10.1016/j.compedu.2014.07.008

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, *91*(4), 483–517. https://doi.org/10.3102/0034654321998074

Haro, S., Rao, H. M., Quatieri, T. F., & Smalt, C. J. (2022). EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. *European Journal of Neuroscience*, *55*(5), 1262–1277. https://doi.org/10.1111/ejn.15616

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Horovitz, T., & Mayer, R. E. (2021). Learning with human and virtual instructors who display happy or bored emotions in video lectures. *Computers in Human Behavior*, *119*, 106724. https://doi.org/10.1016/j.chb.2021.106724

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*, 1–17. https://doi.org/10.1080/10494820.2022.2141787

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. https://doi.org/10.1111/jcal.12078

Kaur, N., & Singh, P. (2023). Conventional and contemporary approaches used in text to speech synthesis: A review. *Artificial Intelligence Review*, *56*(7), 5837–5880. https://doi.org/10.1007/s10462-022-10315-0

Kelley, E. S., & Kinney, K. (2017). Word learning and story comprehension from digital storybooks: Does interaction make a difference? *Journal of Educational Computing Research*, *55*(3), 410–428. https://doi.org/10.1177/0735633116669811

Kim, J., MerrillJr., K., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383. https://doi.org/10.1016/j.chb.2022.107383

Koch, M., von Luck, K., Schwarzer, J., & Draheim, S. (2018). The novelty effect in large display deployments—Experiences and lessons-learned for evaluating prototypes. *ECSCW exploratory papers*. https://api.semanticscholar.org/CorpusID:46980445

Korat, O., Tourgeman, M., & Segal-Drori, O. (2021). E-book reading in kindergarten and story comprehension support. *Reading and Writing*, *35*(1), 155–175. https://doi.org/10.1007/s11145-021-10175-0

Krcmar, M., & Cingel, D. P. (2014). Parent–child joint reading in traditional and electronic formats. *Media Psychology*, *17*(3), 262–281. https://doi.org/10.1080/15213269.2013.840243

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Kuang, Z., Zhang, Y., Wang, F., Yang, X., & Hu, X. (2021). Can the presence of human teacher promote video learning? *Advances in Psychological Science*, *29*(12), 2184–2194. https://doi.org/10.3724/SP.J.1042.2021.02184

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Lasa (2022). *A crooked tree*. Shaanxi People's Education.

Lawson, A. P., & Mayer, R. E. (2022). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education*, *32*(4), 971–990. https://doi.org/10.1007/s40593-021-00282-y

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interevaluator reliability and interevaluator agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Li, X., & Bus, A. G. (2023). Efficacy of digital picture book enhancements grounded in multimedia learning principles: Dependent on age? *Learning and Instruction*, *85*, 101749. https://doi.org/10.1016/j.learninstruc.2023.101749

Li, W., Wang, F., Mayer, R. E., & Liu, T. (2022). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning*, *38*(3), 621–637. https://doi.org/10.1111/jcal.12634

Liao, C. N., Chang, K. E., Huang, Y. C., & Sung, Y. T. (2020). Electronic storybook design, kindergartners' visual attention, and print awareness: An eye-tracking investigation. *Computers & Education*, *144*, 103703. https://doi.org/10.1016/j.compedu.2019.103703

Liew, T. W., Tan, S. M., Tan, T. M., & Kew, S. N. (2020). Does speaker's voice enthusiasm affect social cue, cognitive load and transfer in multimedia learning? *Information and Learning Sciences*, *121*(3/4), 117–135. https://doi.org/10.1108/ILS-11-2019-0124

Martha, A. S. D., & Santoso, H. (2019). The design and impact of the pedagogical agent: A systematic literature review. *The Journal of Educators Online*, *16*(1). https://doi.org/10.9743/jeo.2019.16.1.8

Masataka, N. (2014). Development of reading ability is facilitated by intensive exposure to a digital children's picture book. *Frontiers in Psychology*, *5*, 396. https://doi.org/10.3389/fpsyg.2014.00396

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, *13*(2), 125–139. https://doi.org/10.1016/S0959-4752(02)00016-6

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 345–368). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.017

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied*, *18*(3), 239–252. https://doi.org/10.1037/a0028616

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*(2), 419–425. https://doi.org/10.1037/0022-0663.95.2.419

Moè, A. (2016). Does displayed enthusiasm favour recall, intrinsic motivation and time estimation? *Cognition and Emotion*, *30*(7), 1361–1369. https://doi.org/10.1080/02699931.2015.1061480

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine*, *19*, 98–100. https://doi.org/10.1109/MRA.2012.2192811

Munzer, T. G., Miller, A. L., Weeks, H. M., Kaciroti, N., & Radesky, J. (2019). Differences in parent-toddler interactions with electronic versus print books. *Pediatrics*, *143*(4), e20182012. https://doi.org/10.1542/peds.2018-2012

Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, *224*, 105164. https://doi.org/10.1016/j.compedu.2024.105164

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*, 114. https://doi.org/10.3389/frobt.2018.00114

Ozeri-Rotstain, A., Shachaf, I., Farah, R., et al. (2020). Relationship between eye-movement patterns, cognitive load, and reading ability in children with reading difficulties. *Journal of Psycholinguist Research*, *49*, 491–507. https://doi.org/10.1007/s10936-020-09705-8

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Lawrence Erlbaum Associates.

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, *3*(12), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9

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

Pi, Z., Deng, L., Wang, X., Guo, P., Xu, T., & Zhou, Y. (2022). The influences of a virtual instructor's voice and appearance on learning from video lectures. *Journal of Computer Assisted Learning*, *38*(6), 1703–1713. https://doi.org/10.1111/jcal.12704

Piaget, J. (1968). Quantification, conservation, and nativism. *Science*, *162*, 976–979.

Reich, S. M., Yau, J. C., Xu, Y., Muskat, T., Uvalle, J., & Cannata, D. (2019). Digital or print? A comparison of preschoolers' comprehension, vocabulary, and engagement from a print book and an e-book. *AERA Open*, *5*(3), 233285841987838. https://doi.org/10.1177/2332858419878389

Schneider, P., Dubé, R. V., & Hayward, D. (2005). The Edmonton Narrative Norms Instrument. Retrieved from University of Alberta Faculty of Rehabilitation Medicine website: http://www.rehabresearch.ualberta.ca/enni

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

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

Shamir, A., & Korat, O. (2015). Educational electronic books for supporting emergent literacy of kindergarteners at-risk for reading difficulties—what do we know so far? *Computers in the Schools*, *32*(2), 105–121. https://doi.org/10.1080/07380569.2015.1027868

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

Sikorski, E., Mulvey, S., & Wiese, E. (2019). Effect of anthropomorphic design on the effectiveness of motivational messages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), 1888–1892. https://doi.org/10.1177/1071181319631349

Skibbe, L. E., Thompson, J. L., & Plavnick, J. B. (2018). Preschoolers' visual attention during electronic storybook reading as related to different types of textual supports. *Early Childhood Education Journal*, *46*(4), 419–426. https://doi.org/10.1007/s10643-017-0876-4

Strouse, G. A., Chesnut, S. R., Newland, L. A., Mourlam, D. J., Hertel, D., & Nutting, B. (2022). Preschoolers' electrodermal activity and story comprehension during print and digital shared reading. *Computers & Education*, *183*, 104506. https://doi.org/10.1016/j.compedu.2022.104506

Takacs, Z. K., & Bus, A. G. (2016). Benefits of motion in animated storybooks for children's visual attention and story comprehension. An eye-tracking study. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01591. 7.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research*, *85*(4), 698–739. https://doi.org/10.3102/0034654314566989

Veletsianos, G., & Russell, G. S. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 759–769). Springer. https://doi.org/10.1007/978-1-4614-3185-5_61

Wang, Y., Liu, Q., Chen, W., Wang, Q., & Stein, D. (2019). Effects of instructor's facial expressions on students' learning with video lectures. *British Journal of Educational Technology*, *50*(3), 1381–1395. https://doi.org/10.1111/bjet.12633

Weinert, S. (2022). Language and Cognition. In J. Law, S. Reilly, & C. McKean (Eds.), *Language Development: Individual differences in a Social Context* (pp. 122–143). Cambridge University Press.

Xu, Y. (2023). Talking with machines: Can conversational technologies serve as children's social partners? *Child Development Perspectives*, *17*, 53–58. https://doi.org/10.1111/cdep.12475

Xu, T., Wang, X., Wang, J., & Zhou, Y. (2021a). From textbook to teacher: An adaptive intelligent tutoring system based on BCI. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 7621–7624. https://doi.org/10.1109/EMBC46164.2021.9629483

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021b). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*, 104059. https://doi.org/10.1016/j.compedu.2020.104059

Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2024). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13530

Yang, W., Hu, X., Yeter, I. H., Su, J., Yang, Y., & Lee, J. C. K. (2023). Artificial intelligence education for young children: A case study of technology-enhanced embodied learning. *Journal of Computer Assisted Learning*, *40*(2), 465–477.

Zipke, M. (2017). Preschoolers explore interactive storybook apps: The effect on word recognition and story comprehension. *Education and Information Technologies*, *22*(4), 1695–1712. https://doi.org/10.1007/s10639-016-9513-x

Authors and Affiliations

Jie Bai[1] · Xiulan Cheng[1] · Hui Zhang[1] · Yihang Qin[2] · Tao Xu[2] · Yun Zhou[1]

✉ Xiulan Cheng
chengxiulan@snnu.edu.cn

✉ Yun Zhou
zhouyun@snnu.edu.cn

Jie Bai
baijie@snnu.edu.cn

Hui Zhang
z0915h@snnu.edu.cn

Yihang Qin
qinyihang@mail.nwpu.edu.cn

Tao Xu
xutao@nwpu.edu.cn

[1] Faculty of Education, Shaanxi Normal University, South Chang'an Road 199, Yanta District, Xi'an, Shaanxi Province 710062, P.R. China

[2] School of Software, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an, Shaanxi Province 710072, P.R. China