

Maximizing effectiveness of AI-generated instructors through human-like behavior and dynamic visual cues in instructional videos: Evidence from an eye-tracking study[☆]

Tao Xu^a, Qian Chen^b, Zhiwei Zhang^a, Bingyu Dong^b, Hui Zhang^b, Jie Bai^b, Yun Zhou^{b,*}

^a School of Software, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an 710072, Shaanxi Province, PR China

^b Faculty of Education, Shaanxi Normal University, South Chang'an Road 199, Yanta District, Xi'an 710062, Shaanxi Province, PR China

ARTICLE INFO

Keywords:

Artificial intelligence generated content (AIGC)
Social presence
Gestures
Visual cues
Eye-tracking
Instructional videos

ABSTRACT

The rapid advancement of Artificial Intelligence Generated Content (AIGC) has revolutionized the production of instructional videos and lectures in MOOCs, opening a new chapter in their use for education. This technology enables the generation of talking instructors from text and images or clips of real people. However, the impact of behavioral fidelity and visual cues provided by these generated instructors on learning outcomes in instructional videos remains unknown. This study addresses this gap by employing a 2×2 within-subjects design (behavioral fidelity: high/low; visual cues: with/without) with 36 college student participants. We assessed learning performance, social presence, satisfaction, cognitive load, humanness, and motivation, and analyzed eye movements. Results revealed that high behavioral fidelity increased social presence and more effectively engages students, leading to better transfer performance. Visual cues were found to facilitate the organization and integration of information. The generated instructors were most effective when resembling human behavior and using dynamic visual cues. However, further refinement is needed for the voice and gestures of these instructors. The findings extend the social agency theory related to animated and human instructors to the context of AI-generated instructors. We recommend using instructors with high behavioral fidelity in instructional videos, featuring realistic beat gestures, body movements, facial expressions, and eye contact. Our study offers valuable insights into the implications of using AIGC in educational settings.

1. Introduction

Synthesized voices and virtual characters have proliferated on popular video platforms like YouTube and TikTok (Byun et al., 2023). A growing number of informal instructional videos, knowledge-sharing content, and even online courses (e.g., Netland et al., 2025) on these platforms are generated by artificial intelligence (AI) and incorporate such synthesized characters. These AI-driven contents, created through technologies like lip-syncing and text-to-speech synthesis, have raised concerns on potential privacy violations and ethical issues (Guo et al., 2023). However, such Artificial Intelligence Generated Content (AIGC) has shown great promise in enhancing efficiency and replacing manual labor (Ooi et al., 2023). Given these benefits, it is likely to extend to instructional video platforms soon, potentially leading to significant

transformations in the field of education (Baidoo-Anu & Ansah, 2023).

Instructional videos with on-screen instructors are a prevalent and effective medium for facilitating learning (Henderson & Schroeder, 2021). AIGC technology enables the automatic generation of lifelike virtual instructors for instructional videos, which can be achieved through text-to-speech (Kaur & Singh, 2023) and lip movement synthesis techniques (Prajwal et al., 2020). These technologies allow for the creation of virtual instructors who can replicate human-like facial expressions, gestures, blinking, and body movements to enhance their behavioral fidelity. Behavioral fidelity, refers to the degree of realism in “nonverbal social behaviors” (Frenkel et al., 2024, p. 1), involving aspects such as appearance, rhythmic gestures, body movements, facial expressions, and eye contact. These virtual roles hold great potential for use in instructional videos (Dao et al., 2021), as they can alleviate the

[☆] This article is part of a Special issue entitled: ‘Inclusive Education in the Age of AI’ published in The Internet and Higher Education.

* Corresponding author.

E-mail addresses: xutao@nwpu.edu.cn (T. Xu), qian.chen@snnu.edu.cn (Q. Chen), zhangzhiwei01@mail.nwpu.edu.cn (Z. Zhang), dongbingyu@snnu.edu.cn (B. Dong), z0915h@snnu.edu.cn (H. Zhang), baijie@snnu.edu.cn (J. Bai), zhouyun@snnu.edu.cn (Y. Zhou).

<https://doi.org/10.1016/j.iheduc.2025.101034>

Received 30 June 2024; Received in revised form 30 April 2025; Accepted 30 June 2025

Available online 1 July 2025

1096-7516/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

burden of video recording on teachers and allow them to devote more time and energy to instructional design. Human instructor-produced videos generally involve a crucial step: the recording process. This requires repeated practice by instructors to achieve fluent delivery, appropriate facial expressions, and effective gestures, often resulting in considerable time investment as they make adjustments and rehearse until they are satisfied with the outcome. In contrast, with the use of lip-syncing and text-to-speech synthesis technologies, instructional videos can be generated quickly and efficiently, maintaining high quality without the need for extensive rehearsal. As a result, there is a growing interest among researchers in exploring the educational applications of AIGC. They have begun to examine the impact of AI instructors on learning and explore their instructional design boundaries, to gain a deeper insight into how these virtual instructors can facilitate learning or complement traditional educational methods (e.g., Leiker et al., 2023; Pi et al., 2022). Several studies comparing the effectiveness of human and AI instructors in instructional videos (Netland et al., 2025; Pi et al., 2022; Xu et al., 2024) have shown that AI instructors can match human instructors in improving academic performance, in some cases, even outperform them, such as having a likable appearance or on retention. However, AI instructors still fall short in certain dimensions, such as social presence, when compared to human instructors. Despite these insights, the impact of integrating AI-generated instructors in videos is still not well understood. Existing research indicates that human teachers' gestures may enhance their social presence and promote student attention, thereby improving learning performance (Oh et al., 2018). Whether virtual instructors mimicking these gestures can similarly facilitate learning or inadvertently hinder it remains unclear, lacking empirical evidence.

Regarding redirecting and promoting student attention, another common method is providing visual cues. In an educational context, visual cues refer to visual stimuli used to direct a student's attention and facilitate information searching (Van Gog, 2014). In the form of text, arrows, or other visual representations, these cues provide additional information to help students understand a concept or process (Jamet, 2014). Although previous studies have reported how visual cues promote students' attention and learning in short instructional videos (e.g., Wang et al., 2020; Zhang et al., 2024; Zheng et al., 2023), there is a lack of research examining how both behavioral fidelity and visual cues provided by AI-generated virtual teachers collectively improve students' learning. It remains unclear whether visual cues can enhance the indicative effectiveness of virtual teachers.

The present study aims to better understand the role of AI-generated instructors in learning from instructional videos and explore its theoretical and practical implications. In this study, "AI" refers to the use of the Wav2Lip model (Prajwal et al., 2020) for generating synchronized lip movements in combination with text-to-speech technique, allowing the creation of realistic, talking virtual instructors. The main objectives are: 1) to assess whether an AI-generated instructor can be more effective with the addition of high human-like movement behaviors and visual cues in instructional videos, and 2) to determine if there is an interaction effect between the behavioral fidelity of AI-generated instructors and visual cues.

1.1. The impact of AI-generated characters as virtual instructors

AIGC refers to a method for automatically generating diverse forms of media, including text, images, and video content, by leveraging AI technologies (Baidoo-Anu & Ansah, 2023; Ooi et al., 2023). This method employs AI algorithms to generate content with minimal human intervention, thereby reducing the need for repetitive human labor (Al Naqbi et al., 2024). Recent advancements in AIGC have significantly transformed virtual characters, allowing them to evolve from their previously awkward, puppet-like appearance and ones that now display smooth body movements and human-like conversational gestures (Demirel et al., 2022; Pataranutaporn et al., 2021). The generated videos with

these virtual characters have the potential to be used in education, potentially replacing recorded videos. This shift could alleviate the burden of video recording on teachers, allowing them to devote more time to instructional design. Therefore, researchers have shown an increased interest in exploring the impact of applying AIGC in video lectures and boundaries conditions, to better understand how these virtual instructors can enhance the learning (e.g., Dao et al., 2021; Leiker et al., 2023; Pi et al., 2022). For example, Dao et al. (2021) synthesized video lectures utilizing text-to-speech and lip-sync technologies and conducted a preliminary user study. The findings indicated that learners accepted these AI-generated videos; however, they noted that the quality of the face and voice required improvement and recommended incorporating head movements. Furthermore, Deng et al. (2022) compared the learning outcomes and experience between synthesized virtual instructors and recorded human instructors. Their findings revealed that synthesized instructors could even enhance retention and transfer scores more effectively than human instructors in teaching English vocabulary. Likewise, Leiker et al. (2023) examined the impact of using GAI to create instructional videos with virtual instructors and found no significant differences in learning performance or perceived experience between AI-generated and human-produced videos on the clean energy economy. Craig and Schroeder (2017, 2019) found that machine voices could be as effective as human voices, with modern text-to-speech engines outperforming older ones in facilitating learning. Davis et al. (2019) demonstrated that machine voices were as effective as human voices for information retention. Taken together, these studies suggest that AI-generated content has reached a quality level where it can be accepted by learners as a viable educational resource and can facilitate learning as effectively, if not even better, than traditionally produced instructional videos.

However, integrating this new technology into education presents several unresolved challenges, one of which concerns the boundaries of instructional design. For example, Pi et al. (2022) investigated the influence of a virtual instructor's voice and appearance on learning in video lectures. They found that the AI likable instructor condition had a significant positive effect on students' learning performance and motivation, better than human teacher and virtual teacher with a normal appearance. Davis et al. (2019) noted that the factors impacting voice effectiveness are more complex than a simple human or machine categorization, indicating the need to consider additional factors. An increasing number of studies are focusing on AI-generated instructors and the design factors involved in AI-generated instructional videos, such as voice and appearance. However, research on other critical factors, such as body expressions related to social cues, remains limited.

1.2. Body expressions and social agency theory

Beat gestures, serve to emphasize verbal information and reveal the speaker's construction of narrative discourse (Cassell & McNeill, 1991). McNeill (1992) established one of the most prominent taxonomies of gestures, identifying beats as hand movements synchronized with the rhythm of speech, such as raising and lowering hands rhythmically. As common gestures used spontaneously by instructors, they not only direct attention to the instructor (Beege et al., 2020) but also highlight the key points emphasized by the instructor (So et al., 2012). In instructional videos, teachers' beat gestures serve as social cues.

Social cues, which include, but are not limited to, facial expressions, body language, and gestures, are verbal or non-verbal signals that facilitate social interactions and promote social presence (Mayer, 2014; Mayer et al., 2003). These cues play a critical role in social agency theory, which posits that individuals act as agents in their social environments, interpreting and responding to social signals (Atkinson et al., 2005). Social agency theory provides a framework for understanding how the social presence of virtual or human instructors in multimedia learning, instructional videos, or virtual reality environments impact learning (Moon & Ryu, 2021). In the context of instructional videos, this

theory highlights the importance of integrating teacher presence to simulate human-to-human interactions. Specially, the inclusion of social cues, such as a teacher's gestures, serves to stimulate learners' social responses, thereby fostering cognitive engagement and enhancing learning (Mayer, 2014). Building on this foundation, Mayer's embodiment principle offers practical guidelines for leveraging social cues to optimize learning. According to Mayer, the natural movements, gestures, and facial expressions of virtual characters or instructors can promote a greater sense of interaction and effectively guide learners' attention to the content in instructional videos (Mayer, 2014; Mayer, 2021).

Beat gestures in instructional videos, as a form of body expression, have demonstrated mixed effects on learning, including neutral (e.g., Beege et al., 2020), negative (e.g., Moon & Ryu, 2021) and positive results (e.g., Gluhareva & Prieto, 2017). A systematic review by Oh et al. (2018) highlighted the importance of behavioral realism in pedagogical agents and identified it as a powerful predictor of social presence. The majority of empirical evidence supporting these findings have been derived from studies involving human instructors or animated pedagogical agents in traditional recorded instructional videos (e.g., Pi, Zhu, Zhang, & Yang, 2021). Recently, synthesized virtual characters have gained popularity on platforms like YouTube and TikTok (Byun et al., 2023; Netland et al., 2025). However, there is a lack of empirical evidence to validate or challenge the applicability of social agency theory in explaining the effects of incorporating social cues, such as gestures, into AI-generated humanoid teachers in instructional videos. It is not yet well understood whether students ignore these social cues of AI teachers, respond negatively to them, or engage with them in a positive and interactive manner. Therefore, research is needed to investigate how the behavioral fidelity of AI-generated instructors influences learners' perceptions of social presence and their learning performance.

1.3. The effects of visual cues on learning

Cueing, also referred to as signaling, serves as a method to direct attention towards or emphasize key points in a textual, visual or mixed form (Van Gog, 2014), employing techniques like spotlighting, coloring, flashing, arrows, referencing, labeling, and textual prompts (Schneider et al., 2018). Through these techniques, cues establish connections across multiple channels of information processing, aiming to reduce cognitive load for students and enhance their integration and organization of information (Mayer, 2005). The cueing principle proposes that incorporating cues like color or arrows into complex graphics, animations, instructional videos, and virtual reality enhances learning performance experience (e.g., Albus et al., 2021; De Koning et al., 2009; Wang et al., 2020). Also, the findings in meta-analysis (Richter et al., 2016; Schneider et al., 2018) have demonstrated the positive effects of cues on learning in multimedia. However, cues intended to scaffold learning are not always effective, as evidenced by studies reporting neutral effects on performance and comprehension (Crooks et al., 2012; Johnson et al., 2015; Yang, 2016). Additionally, the findings of the impact of different types of cues on learning outcomes and experience were divergent (Wang et al., 2020; Zheng et al., 2023). For example, Wang et al. (2020) observed that visual cues and the combination of visual and textual cues have a more significant positive impact on learning compared to textual cues alone. However, Pi et al. (2023) found that textual cues are more effective than visual cues in improving English as a Foreign Language vocabulary learning.

The aforementioned studies primarily focus on various multimedia learning environments, with instructional videos receiving comparatively limited attention (e.g., Johnson et al., 2015; Pi et al., 2023; Wang et al., 2020). The applicability of principles and the effectiveness of these effects are influenced by boundary conditions such as the nature of the learning materials (Mayer, 2010). Little is known about how visual cues influence vocabulary learning in instructional videos. Thus, there remains a lack of clarity regarding the influence of human teacher

behavior and visual cues, as well as their interaction effects on learning outcomes in instructional videos. Furthermore, there is a significant gap in the literature concerning the impact of the behavioral fidelity of AI-generated instructors and visual cues.

1.4. The present study and research questions

Advancements in AIGC technology are enabling the generation of virtual instructors for instructional videos, which could potentially emerge as crucial educational resources in the future. Given this transformation, it is important to explore the effects of such AI-generated instructors and instructional videos. However, research on the impact of generated content on learning is currently quite limited and in its early stages (e.g., Dao et al., 2021; Leiker et al., 2023; Pi et al., 2022). The efficacy of learning facilitation through the mimicry of human teachers' behaviors by generated human-like instructors remains uncertain. The present study addresses this gap by examining the impact of behavioral fidelity and visual cues provided by these generated instructors on learning in instructional videos. The experiment utilized a 2 × 2 within-subjects design, manipulating behavioral fidelity (high/low) and visual cues (with/without) with 36 college student participants. A pre-post design was leveraged to analyze retention and transfer across four groups. Behavioral fidelity, aimed at providing social cues, required the measurement of social presence. Given the learning context involving new AI technology, we included a satisfaction dimension. Furthermore, this study measured cognitive load and motivation related to learning. Eye-tracking technologies were employed to investigate attention and information organization and integration, corresponding to cognitive processing.

Based on social agency theory (Atkinson et al., 2005) and the embodiment principle (Mayer, 2014; Mayer, 2021), we hypothesize that AI-generated instructional videos embedded with virtual instructors with high behavioral fidelity will lead to higher social presence, thereby improving student performance. Drawing on previous research on visual cues in instructional videos (Wang et al., 2020) and the cueing principle (Mayer, 2021; Van Gog, 2014), we predict that visual cues will enhance learning and support AI-generated instructors in facilitating the learning process. Thus, we propose the following research questions (RQ) and hypotheses (H):

RQ1. How do the behavioral fidelity of AI-generated instructors and the presence of visual cues affect learning performance?

H1. High behavioral fidelity is expected to significantly enhance retention (H1a) or transfer (H1b). The presence of visual cues is expected to improve retention (H1c) or transfer (H1d).

RQ2. How do the behavioral fidelity of AI-generated instructors and the presence of visual cues affect satisfaction, social presence, cognitive load, motivation, and humanness?

H2. High behavioral fidelity is predicted to increase student satisfaction (H2a), social presence (H2b), motivation (H2c), and humanness (H2d), potentially leading to an increased cognitive load (H2e). The absence of visual cues is expected to diminish satisfaction (H2f), while leading to an increased cognitive load (H2g).

RQ3. How do the behavioral fidelity of AI-generated instructors and the presence of visual cues affect attention?

H3. High behavioral fidelity is expected to draw increased attention towards instructors (H3a). The presence of visual cues is hypothesized to enhance the attention direction (H3b).

RQ4. How do the behavioral fidelity of AI-generated instructors and the presence of visual cues affect organizing and integrating information?

H4. The presence of visual cues is anticipated to facilitate the organization and integration of information.

2. Methods

2.1. Participants

A total of 36 undergraduate and graduate students (21 males and 15 females) were randomly recruited, aged between 19 and 29 ($M = 23.22$, $SD = 1.69$). A power analysis was conducted using G*Power 3.1 (Faul et al., 2007) based on the effect size ($\eta^2 = 0.16$, converted to $f = 0.4364358$) found in (Pi et al., 2022). The effect size can be determined through a priori power analysis (Faul et al., 2007), using effect sizes from prior studies, pilot studies, or empirical data (Kang, 2021). Since this study shares a within-subjects design and a similar theme on AI-generated instructors in instructional videos with Pi et al.'s study (Pi et al., 2022), we have estimated the effect size based on their findings. The conversion was calculated using the formula from (Cohen, 2016). With an effect size of $f = 0.4364358$, power of 0.80, and alpha level of 0.05, a total sample size of 9 participants at least were required to detect an effect. Therefore, the number of participants ($n = 36$) in this study met the requirements.

This study was approved by the ethics board of our institution. All participants signed an informed consent agreement and were compensated for their time. They were native Chinese speaker with normal or corrected-to-normal vision. Their average CET4 and CET6 scores were 488.46 ($SD = 60.46$) and 462.75 ($SD = 46.78$). CET4 and CET6 are standardized English proficiency exams in China. CET4 assesses basic English skills, while CET6 measures higher-level proficiency. Participants' scores of CET4 and CET6 showed that they could read, listen to, and write in English.

2.2. Experimental design

The present study used a 2×2 within-subjects design to control for individual differences in memory, learning ability, and foundational knowledge of English. As shown in Table 1, we manipulated the independent variables of behavioral fidelity (high and low) and visual cues (with and without). Concerning high behavioral fidelity, the virtual instructor exhibited beat gestures, body movement, facial expressions, and eye-contact, as shown in Fig. 1 on the bottom. In contrast, low behavioral fidelity featured a static virtual instructor with only synchronized lip movements, without any gestures or body movements (see Fig. 1, top). The dynamic visual cues presented were aligned with the

Table 1
Experimental and control conditions.

Conditions		Descriptions
Experimental conditions	High behavioral fidelity with visual cues (HBC)	<ul style="list-style-type: none">The talking virtual teacher has human-like appearance, beat gestures, body movement, facial expression, and eye-contact.Depending on the video content, highlighted square boxes, a different color for the font, and a directed icon will be displayed dynamically.
	High behavioral fidelity without visual cues (HB)	<ul style="list-style-type: none">The talking virtual teacher has human-like appearance, beat gestures, body movement, facial expression, and eye-contact.
	Low behavior fidelity with visual cues (LBC)	<ul style="list-style-type: none">The talking virtual teacher is almost static only with synchronized lip movement.Depending on the video content, highlighted square boxes, a different color for the font, and a directed icon will be displayed dynamically.
The control condition	Low behavior fidelity without visual cues (LB)	<ul style="list-style-type: none">The talking virtual teacher is almost static only with synchronized lip movement.

learning content. When the instructor pronounced a word, it was highlighted (see Fig. 1 left). Subsequently, when the instructor explaining the word's part of speech and definition, the relevant section of the slides was marked in red. An illustration was highlighted with a pointing hand icon when the instructor describing the context-appropriate usage of the word. Finally, the word was emphasized within an example sentence. The three experimental conditions were HBC (high behavioral fidelity with visual cues, see Fig. 1C), HB (high behavioral fidelity without visual cues, see Fig. 1D), and LBC (low behavior fidelity with visual cues, see Fig. 1A), with LB (low behavior fidelity without visual cues, see Fig. 1B) serving as the control condition.

Each participant was required be exposed to all four conditions, watching four short instructional videos to learn 24 English words (six words for each video). The order of four conditions was counterbalanced.

To screen English words that could be used in the formal experiment, a pilot test was conducted. Previous research indicates that word familiarity serves as an indicator of word difficulty, with strong correlations observed between the two (Crossley et al., 2014; Leroy & Kauchak, 2014). Accordingly, we utilized word familiarity to predict word difficulty. We first selected 60 words from the GRE, TOEFL, and IELTS to create a word pool. This pool can be used to develop learning materials and tests for experiments. Then we invited a different 24 students (M age = 21.92, SD age = 1.86; 21 females) to rate their familiarity in these words. We asked them to rate all total 60 words from 1 ("extremely unfamiliar") to 7 ("extremely familiar"). A lower score indicates a greater likelihood that the word was not learned by students recruited in the formal experiment. We ranked the words according to their scores, from low to high, and then stratified them into four groups containing six words each, with balanced number of the part of speech, length, and score. Stratification is the process of dividing members of the population into homogeneous subgroups before sampling.

The descriptive results showed that students were not familiar with these 24 target words used in instructional videos ($M = 1.76$, $SD = 0.32$). To evaluate these words, we considered two dimensions: familiarity and the number of students unfamiliar with the word. The analysis revealed no significant differences in word familiarity across the four groups, $F(3,15) = 0.028$, $p = 0.993$, $\eta^2 = 0.006$. Additionally, there were no significant differences in the number of students unfamiliar with the word across the groups, $F(3,15) = 0.281$, $p = 0.838$, $\eta^2 = 0.053$. These findings indicate that the words used in all four instructional videos were comparable.

2.3. AI-generated instructors and instructional videos

We utilized the Wav2Lip model (Prajwal et al., 2020) and text-to-speech to generate lip-syncing videos. The generation application, implemented in Python, includes a user-friendly interface for inputting images, videos and text, and outputs a generated video with talking virtual instructor. In this study, we leveraged a video of a human teacher instead of a static image as the input to guide the AI generation process. This input video is used to align the AI-generated face with the facial features and body gestures of the human instructor. By incorporating both facial expressions and body movements, this approach creates more natural and dynamic generated content. All input videos across the experimental conditions were sourced from the same human instructor. The design of video in this study follows the B3 Fixed but Overlapping design (Crook and Schofield, 2017), featuring a video narrator at a fixed position overlapping the background (as illustrated in Fig. 1).

In this study, we used the 24 English words mentioned above to create the learning materials, with each video containing six words. Each video presented the instructor in high/low behavioral fidelity with/without visual directed cues: (1) HBC, (2) HB, (3) LBC, and (4) LB. The duration of the video lectures were about 8 min, that is, 460 s (HBC), 476 s (HB), 470 s (LBC) and 495 s (LB). For each word and each video, an instructor lectured in the following four organized steps. First, the

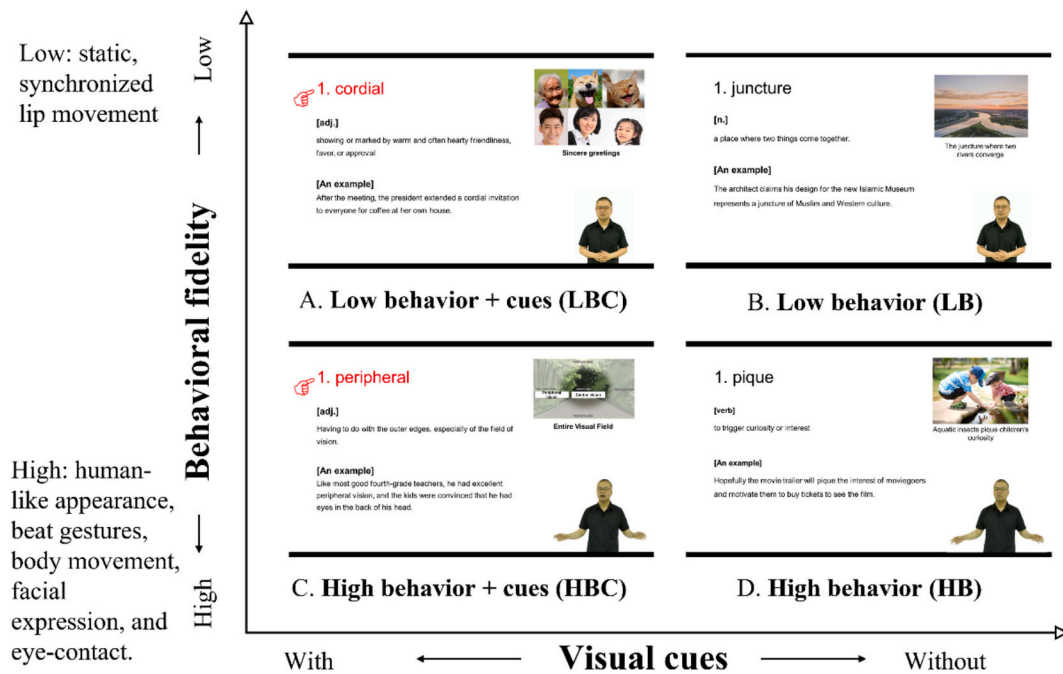


Fig. 1. The experimental design of the study and illustrative examples (translated). High behavioral fidelity featured human-like appearance, beat gestures, body movement, facial expression, and eye-contact (bottom), whereas low behavioral fidelity featured only synchronized lip movement (top). The talking virtual teacher lectures with (left) and without (right) visual cues to direct students' attention.

instructor read the word three times to let students be familiar with pronunciation. Second, the instructor presented the part in speech of the word and illustrated the definition in English and then Chinese. Third, the instructor explained the usage of the word in the context, and finally gave an example sentence as well as the Chinese translation.

2.4. Dependent measures

We collected data on learning performance (retention and transfer), social presence, satisfaction, cognitive load, humanness, and motivation. Additionally, we tracked eye movements and conducted brief interviews.

2.4.1. Pre- and post-tests of learning performance

In the pretest, we assessed participants' familiarity with the words by requiring them to fill in the Chinese meanings for each word in the provided blanks. A correct response received one point, while an incorrect response received zero points. The pretest for each group had a maximum possible score of six, with lower scores indicating a lower level of prior knowledge regarding the words to be learned.

In the performance posttest, participants were required to complete four types of tasks: (1) recalling and writing down the meanings of words, (2) selecting the correct words, (3) identifying synonyms, and (4) choosing antonyms. Each group was presented with twenty-four multiple-choice questions. The first task, which involved recalling and writing down the meanings of the words, was identical to the pretest and aimed to assess retention. This task measured the retention. The second task, adapted from fill-in-the-blank quizzes in Merriam-Webster's Vocabulary Builder, required participants to select the correct word to complete a sentence. For example, one question asked: "1. ___ vision is part of what most eye doctors test in their patients. (peripheral, dyslexia, panoply, nescience, omnivore, dissuade)." The third task involved selecting synonyms or related words, and the fourth task required the selection of antonyms or near antonyms. Research indicates that synonyms and antonyms have both positive and negative transfer effects on language learning (Kanungo, 1967). Therefore, we classify Tasks 3 and 4 as involving transfer. Taken together, the second, third, and fourth tasks

were designed to measure the application of vocabulary knowledge, i.e., transfer. Each correct answer in the performance posttest was awarded 1 point, while incorrect answers received 0 points. The total possible score for each group was twenty-four points.

2.4.2. Subjective measures

Table 2 presents an overview of the subjective measures, detailing the number of questions, types, and references. The scales used for subjective measures were originally written in English and then translated into Chinese for use in this study. To ensure the accuracy of the translation, we used forward and backward translation and involved two bilingual experts for review. To further assess the psychometric properties of the scales, we conducted an exploratory factor analysis. The results indicated strong construct validity, with a KMO value of 0.910 and a significant p -value (< 0.001).

Satisfaction, social presence, cognitive load, motivation, and humanness were evaluated using a five-point Likert scale, with responses ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores represented greater levels of agreement. The subjective measures demonstrated strong internal consistency, with Cronbach's α values of 0.91.

2.4.3. Measures of eye movement data and eye-tracking device

The eye-mind link hypothesis suggests that eye movement patterns

Table 2
Subjective ratings used in the present study.

Variable	Numbers of questions	Reliability	Type	Reference
Satisfaction	5	0.90	5-point Likert	(Hajhashemi et al., 2018)
Social presence	2	0.84		(Katuk et al., 2013)
Motivation	2	0.69		
Humanness	1	N/A		
Cognitive load	2	0.61		(Paas et al., 1994)

are closely tied to cognitive processes and can reflect an individual's mental activity to some extent (Just & Carpenter, 1980). Eye-tracking data offers a real-time view of learners' visual attention, revealing how they allocate cognitive resources during information processing (Holmqvist et al., 2011; Lai et al., 2013). By tracking eye movements and fixations, researchers gain insights into attention distribution and the cognitive processes of different activities (Pi et al., 2022; Wang et al., 2020). To compare attention allocation, we defined two Areas of Interest (AOIs): the instructor AOIs and the learning materials AOIs, as shown in Fig. 2A. To compare the use of cues, we defined four AOIs. The visual cues will appear sequentially in these four areas, as shown in Fig. 2B, including gesture indicator icons, highlighted text, borders, and the learning content. We exported the following eye-tracking metrics to measure attention allocation, information organization and integration, and attention on the visual cues.

2.4.3.1. Percentage fixation duration on an AOI. The ratio of the total fixation duration within an AOI to the total duration. This metric reflects the extent of attention allocated by participants to particular regions on the screen. A higher value indicates a greater degree of attention directed towards the AOI (Lai et al., 2013). This study measures the percentage of fixation duration on the AI instructor, learning materials, and visual cues.

2.4.3.2. Percentage Saccade on an AOI. The ratio of saccades that either begin or terminate within an AOI to the total number of saccades. Saccades are defined as the total number of times that a fixation transits from one AOI to another. These transitions are indicative of visual scanning and efforts to organize and integrate information (Wang et al., 2020). Thus, this metric serves as an indicator of the extent of cognitive attempts at organizing and integrating information on an AOI. A higher value suggests a greater effort.

While each participant watched the instructional video on English words, their eye movements were captured using a Tobii Pro Nano eye tracker with a sampling rate of 60 Hz. The Nano, a screen-based eye tracker, was attached to a 21.5-in. monitor. Participants were allowed to move their heads freely and sat approximately 60 cm from the eye tracker. The Tobii Prolab software was used for participant calibration and data analysis. The screen frame rate in Tobii Pro Lab is set to 25 Hz. The screen frame rate and eye tracker sampling rate are distinct: the former refers to how often the screen updates (25 Hz), while the latter indicates how often eye positions are recorded (60 Hz). A higher sampling rate than the frame rate allows multiple eye-tracking data points to be captured within each screen frame, enabling more precise analysis of eye movements.

Eye tracking data is processed using the Tobii I-VT (Fixation)

algorithm (Olsen, 2012), with the following settings: a minimum fixation duration of 60 ms (fixations shorter than 60 ms are discarded), a maximum event interval of 75 ms between consecutive fixations (intervals shorter than 75 ms are merged), and a maximum angular difference of 0.5° between consecutive fixations (angles smaller than 0.5° are merged). The threshold for eye movement speed is set at 30°/s (movements below this speed are classified as fixations, and those above it as saccades). The time window for calculating eye movement speed is 20 ms, and 3 data samples are processed in each calculation.

2.5. Suggestion and comments collection

The objective of this qualitative data collection and analysis is to assist in finding the underlying reasons for the observed effects and to inform improvements in the design of AI instructors and their cues. After completing the subjective scales, participants were asked to provide their suggestions and comments on the instructional video they had just watched and learned from by answering a single prompt: "Please fill in your suggestions and comments on this instructional video." They could provide their responses either in writing or orally; however, all participants chose to write down their comments. For analysis, we collected the written responses, organized the text data into a spreadsheet, and applied open coding.

To ensure the reliability of the qualitative analysis, two independent coders were involved in the process. Both coders first familiarized themselves with the data by reading through all participant responses. They then independently generated preliminary codes, focusing on recurring themes and key concepts. After completing the initial coding, the two coders compared their results to identify similarities and differences. Discrepancies were resolved through discussion to achieve consensus, and a final set of codes was established. Through this process, four key themes emerged from the data (i.e., they were not pre-determined): "behavioral fidelity design," "visual cues design," "unclassified design," and "outside design context." These codes were then applied to analyze participant responses for each condition.

To measure the reliability of the coding process, Cohen's Kappa was calculated as an indicator of inter-coder agreement (Cohen, 1960). Initially, the two coders achieved a Cohen's Kappa value of 0.88, indicating substantial agreement. Following the discussion and resolution of discrepancies, the final coding achieved full consensus.

2.6. Procedure

Before starting the experiment, participants completed the demographic information (e.g., age, gender, and major) and the retention pretests. Then, they watched four videos, filled out the learning



Fig. 2. The screenshot of Areas of Interest (AOIs). A. Instructor and the learning materials AOIs. B. Dynamic AOIs for dynamic visual cues (The dynamic visual cues appear sequentially at positions ①, ②, ③, and ④ in the slides).

performance posttests, subjective scales in a counterbalanced order while their eye movements were recorded in real-time, as shown in Fig. 3. Then, participants answered one single question in a short interview. On average, participants took about 70 min to complete the experiment.

2.7. Data analysis

Following the results of the Kolmogorov-Smirnov test and visual evaluations of histograms, normal Q-Q plots, and box plots, it was determined that not all the observed values exhibited a normal distribution across the four conditions. The following variables violated the normality assumption: pretest, performance task 1 (retention), social presence, cognitive load, motivation, humanness, percentage fixation duration on the AI instructor, and percentage saccades on learning materials. The variables performance tasks 2–4 (transfer), satisfaction, percentage fixation duration on learning materials, and percentage fixation duration on cues were normally distributed. Therefore, the repeated ANOVA tests and the Friedman ANOVA tests were used for further analysis based on the normality of the data. Post-hoc analysis was performed using the Dunn-Bonferroni test (default in SPSS) for Friedman ANOVA, or the Least Significant Difference (LSD) test for repeated measures ANOVA, depending on the type of analysis conducted. We employed η^2 to measure the effect sizes in repeated ANOVAs, with thresholds for small (0.01), medium (0.06), and large (0.14) effects.

3. Results

Table 3 shows an overview of the descriptive statistics of learning performance, subjective scales, and eye-tracking metrics across the four groups. Tables 4 and 5 present the results of Friedman's ANOVA (used for non-normally distributed data) and repeated measures ANOVA (used for normally distributed data) respectively. Table 6 shows a summary of predictions and tests.

3.1. Learning performance

3.1.1. Pretest

The participants possessed low prior knowledge of the materials to be learned as shown in Table 3. The results of the Friedman analysis showed no significant differences in prior knowledge across the four groups ($\chi^2(3) = 1.286, p = 0.733$). This indicated that participants'

Table 3

Descriptive statistics across five conditions.

Measures	LBC	LB	HBC	HB
	(n = 36)	(n = 36)	(n = 36)	(n = 36)
	M (SD)	M (SD)	M (SD)	M (SD)
Pretest	0.03 (0.17)	0.03 (0.17)	0.03 (0.17)	0.00 (0.00)
Performance task 1 (Retention)	4.22 (1.61)	3.92 (1.63)	4.36 (1.44)	4.22 (1.62)
Performance task 2–4 (Transfer)	11.22 (3.92)	10.11 (3.69)	12.56 (3.71)	11.83 (3.81)
Satisfaction	2.98 (1.12)	2.87 (0.86)	3.03 (1.02)	2.88 (0.98)
Social presence	3.18 (0.97)	3.29 (1.10)	3.53 (0.98)	3.32 (1.01)
Cognitive load	3.44 (0.78)	3.57 (0.82)	3.51 (0.70)	3.51 (0.77)
Motivation	2.82 (1.10)	2.79 (1.02)	2.88 (0.94)	2.74 (0.94)
Humanness	2.92 (1.30)	2.89 (1.21)	2.61 (1.27)	2.64 (1.27)
Percentage fixation duration on learning materials	0.54 (0.08)	0.54 (0.10)	0.52 (0.09)	0.53 (0.10)
Percentage fixation duration on the AI instructor	0.08 (0.07)	0.08 (0.08)	0.10 (0.08)	0.08 (0.07)
Percentage fixation duration on cues	0.60 (0.13)	0.43 (0.12)	0.55 (0.15)	0.39 (0.11)
Percentage saccades on learning materials	0.83 (0.06)	0.84 (0.09)	0.82 (0.05)	0.83 (0.07)

Table 4

The results of Friedman ANOVA tests for processing non-normal distribution data.

Dependent variables	χ^2	p
Pretest	1.286	0.733
Performance task 1 (Retention)	3.433	0.330
Social presence	8.908	0.031*
Cognitive load	1.833	0.608
Motivation	1.858	0.602
Humanness	0.973	0.808
Percentage fixation duration on the AI instructor	14.900	0.002**
Percentage saccades on learning materials	10.233	0.017*

Note: * $p < 0.05$, ** $p < 0.01$.

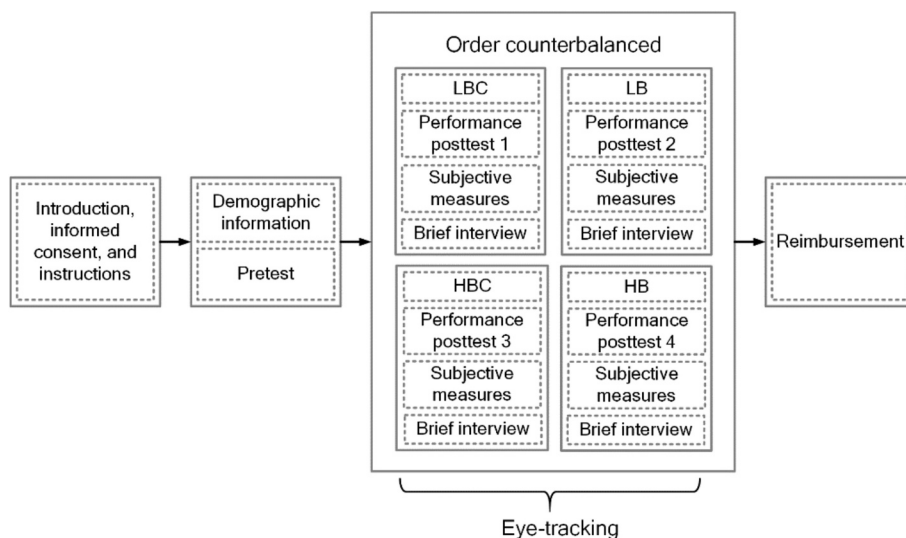


Fig. 3. The experimental procedure.

Table 5

The results of repeated measures ANOVA tests for processing normal distribution data.

Dependent variables	One-way	Two-way		
	Group	Behavior fidelity	Visual cues	Behavior fidelity * visual cues
	$p (\eta_p^2)$	$p (\eta_p^2)$	$p (\eta_p^2)$	$p (\eta_p^2)$
Performance task	<0.001**	0.001**	0.022*	0.658
2–4 (Transfer)	(0.156)	(0.290)	(0.142)	(0.006)
Satisfaction	0.371	0.728	0.067	0.797
	(0.029)	(0.003)	(0.092)	(0.002)
Percentage fixation duration on learning materials	0.291	0.157	0.668	0.337
	(0.035)	(0.056)	(0.005)	(0.026)
Percentage fixation duration on cues	<0.001**	0.001**	<0.001**	0.984
	(0.641)	(0.277)	(0.779)	(<0.001)

Note: * $p < 0.05$, ** $p < 0.01$.

familiarity with the words was consistent across all groups.

3.1.2. Performance posttests

3.1.2.1. Retention. The results of the Friedman analysis revealed no significant differences in retention across the four groups ($\chi^2(3) = 3.433$, $p = 0.3d30$). The highest performance was observed in the HBC group, with a mean score of 4.36.

3.1.2.2. Transfer. The results of the one-way repeated measures ANOVA analysis indicated significant differences in the transfer of learning among the four groups ($F(3,105) = 6.446$, $p < 0.001$, $\eta_p^2 = 0.156$). The HBC group demonstrated the highest performance, with a mean score of 12.56. Post hoc analysis using the LSD test revealed that the HBC group significantly outperformed the LBC group (Mean Difference = 1.333, $p = 0.031$) and the LB group (Mean Difference = 2.444, $p < 0.001$), while the HB group outperformed the LB group (Mean Difference = 1.722, $p = 0.006$), as shown in Fig. 4.

The results of the two-way repeated measures ANOVAs indicated significant main effect of cueing on transfer scores ($F(1,35) = 5.778$, $p = 0.022$, $\eta_p^2 = 0.142$). Moreover, a significant main effect of behavioral fidelity on transfer scores was found ($F(1,35) = 14.313$, $p = 0.001$, $\eta_p^2 = 0.290$), with groups experiencing high behavioral fidelity achieving significantly higher transfer scores compared to those exposed to low behavioral fidelity. The interaction effect between cueing and behavioral fidelity on transfer scores was not significant ($F(1,35) = 0.199$, $p = 0.658$, $\eta_p^2 = 0.006$).

Hypothesis 1 posited that high behavioral fidelity is expected to significantly enhance retention (H1a) or transfer (H1b). The presence of visual cues is expected to improve retention (H1c) or transfer (H1d). We observed significant differences in transfer, showing that the HBC outperformed the LBC and LB groups, and the HB group outperformed the LB group. Behavioral fidelity and cueing had main effects on transfer performance. However, no significant differences were found in retention. Therefore, these results partially supported H1.

3.2. Subjective measures

3.2.1. Satisfaction

The results of the one-way repeated measures ANOVA revealed that there were no statistically significant differences in satisfaction among the four groups ($F(3,105) = 1.056$, $p = 0.371$, $\eta_p^2 = 0.029$). Of all the conditions, the HBC group exhibited the highest mean performance score ($M = 3.03$).

The two-way repeated measures ANOVA results indicated that there was no significant main effect of cueing on satisfaction ($F(1,35) = 3.567$, $p = 0.067$, $\eta_p^2 = 0.092$). There was no significant main effect of

Table 6

Predictions and tests.

Hypothesis	Prediction	Measures	Results	Upheld
H1a	High behavioral fidelity is expected to significantly enhance retention.	Retention	Non-significant	Rejected
H1b	High behavioral fidelity is expected to significantly enhance transfer.	Transfer	Significant	Supported
H1c	The presence of visual cues is expected to improve retention.	Retention	Non-significant	Rejected
H1d	The presence of visual cues is expected to improve transfer.	Transfer	Significant	Supported
H2a	High behavioral fidelity is predicted to increase student satisfaction.	Satisfaction	Non-significant	Rejected
H2b	High behavioral fidelity is predicted to increase student social presence.	Social presence	Significant	Supported
H2c	High behavioral fidelity is predicted to increase student motivation.	Motivation	Non-significant	Rejected
H2d	High behavioral fidelity is predicted to increase student humanness.	Humanness	Non-significant	Rejected
H2e	High behavioral fidelity is predicted to increase student cognitive load.	Cognitive load	Non-significant	Rejected
H2f	The absence of visual cues is expected to diminish satisfaction.	Satisfaction	Non-significant	Rejected
H2g	The absence of visual cues is expected to an increased cognitive load.	Cognitive load	Non-significant	Rejected
H3a	High behavioral fidelity is expected to draw increased attention towards instructors.	Percentage fixation duration on learning materials Percentage fixation duration on the AI instructor	Non-significant Significant	Supported
H3b	The presence of visual cues is hypothesized to enhance the attention direction.	Percentage fixation duration on cues	Significant	Supported
H4	The presence of visual cues is anticipated to facilitate the organization and integration of information.	Percentage saccades on learning materials	Significant	Partially supported

behavioral fidelity on satisfaction ($F(1,35) = 0.123$, $p = 0.728$, $\eta_p^2 = 0.003$). The interaction effect between cueing and behavioral fidelity on satisfaction was also not significant ($F(1,35) = 0.067$, $p = 0.797$, $\eta_p^2 = 0.002$).

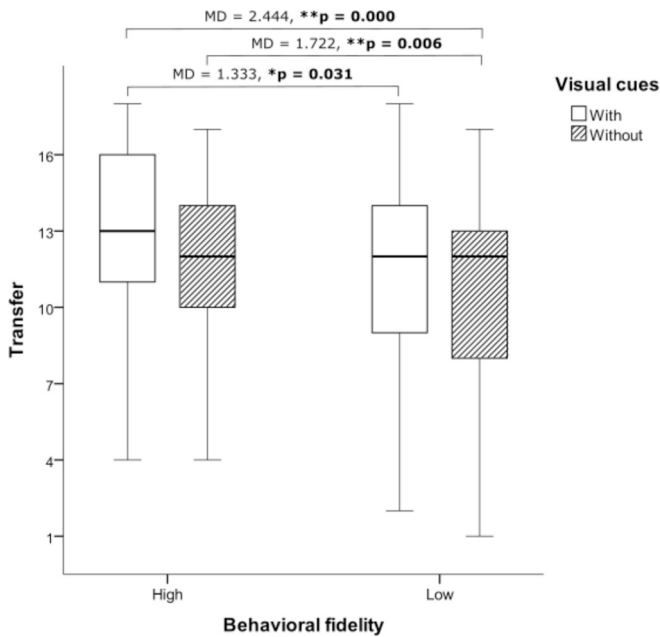


Fig. 4. Differences in the transfer performance across four conditions.

3.2.2. Social presence

The results of the Friedman analysis revealed significant differences in the social presence across the four groups ($\chi^2(3) = 8.908, p = 0.031$). Among the conditions, the HBC group gained the highest performance with a mean score of 3.53.

3.2.3. Cognitive load

The Friedman analysis revealed no significant differences in the cognitive load of students across the four groups ($\chi^2(3) = 1.833, p = 0.608$). Among the four conditions, the LB group demonstrated the highest mean performance ($M = 3.57$).

3.2.4. Motivation

The Friedman analysis revealed no significant differences in student motivation across the four groups ($\chi^2(3) = 1.858, p = 0.602$). Among the conditions, the HBC group exhibited the highest mean motivation score ($M = 2.88$).

3.2.5. Humanness

The Friedman analysis revealed no significant differences in the perceived humanness across the four groups ($\chi^2(3) = 0.973, p = 0.808$). Among the conditions, the LBC group exhibited the highest mean score ($M = 2.92$).

Hypothesis 2 was that high behavioral fidelity is predicted to increase student satisfaction (H2a), social presence (H2b), motivation (H2c), and humanness (H2d), potentially leading to an increased cognitive load (H2e). The absence of visual cues is expected to diminish satisfaction (H2f), while leading to an increased cognitive load (H2g). We observed significant differences in social presence across the four groups, with learners in the high-fidelity instructor groups reporting higher social presence compared to those in the low-fidelity instructor groups. However, no significant differences were found in other subjective measures. Collectively, these findings support H2b and reject the other sub-hypotheses.

3.3. Results based on eye-tracking data

3.3.1. Attention allocation and direction

3.3.1.1. Percentage fixation duration on learning materials.

The one-way repeated measures ANOVA analysis revealed no significant differences in the percentage of fixation duration on learning material among students across the four groups ($F(3,105) = 1.263, p = 0.291, \eta^2 = 0.035$). Among the various conditions, the LBC group exhibited the highest mean percentage of fixation duration ($M = 0.54$).

The results of the two-way repeated measures ANOVA indicated that there was no significant main effect of cueing ($F(1,35) = 0.187, p = 0.668, \eta^2 = 0.005$) and behavioral fidelity ($F(1,35) = 2.087, p = 0.157, \eta^2 = 0.056$) on the percentage of fixation duration on the learning materials. The interaction effect between two independent factors on the percentage of fixation duration on the learning material was not significant ($F(1,35) = 0.949, p = 0.337, \eta^2 = 0.026$).

3.3.1.2. Percentage fixation duration on the AI instructor. The Friedman analysis revealed significant differences in the percentage of fixation duration on the AI teacher across the four groups ($\chi^2(3) = 14.900, p = 0.002$). The HBC group demonstrated a significantly higher percentage of fixation duration on the AI instructor compared to the LBC group (Mean difference = 0.020, $p = 0.008$) and the LB group (Mean difference = 0.021, $p = 0.003$), as shown in Fig. 5.

3.3.1.3. Percentage fixation duration on cues. The one-way repeated measures ANOVA analysis revealed significant differences in the percentage fixation duration on cues across the four groups ($F(3,105) = 62.510, p < 0.001, \eta^2 = 0.641$). Post hoc LSD tests showed significant differences between all pairwise comparisons across the four groups. Among them, the LBC group had a significantly higher fixation duration on cues compared to the LB group (Mean Difference = 0.166, $p < 0.001$). Additionally, the HBC group exhibited significantly higher fixation durations than the HB group (Mean Difference = 0.166, $p < 0.001$), as shown in Fig. 6. These findings indicated that visual cues improved attention direction and effectively guided students to focus more on the relevant areas, including the cues and the learning content.

The two-way repeated measures ANOVA revealed significant main effects of behavioral fidelity ($F(1,35) = 13.383, p = 0.001, \eta^2 = 0.277$) and visual cues ($F(1,35) = 123.671, p < 0.001, \eta^2 = 0.779$) on the percentage of fixation duration on cues, but no significant interaction effect ($F(1,35) < 0.001, p = 0.984, \eta^2 < 0.001$).

Hypothesis 3 was that high behavioral fidelity is expected to draw increased attention towards instructors (H3a) and the presence of visual cues is hypothesized to enhance the attention direction (H3b). The HBC

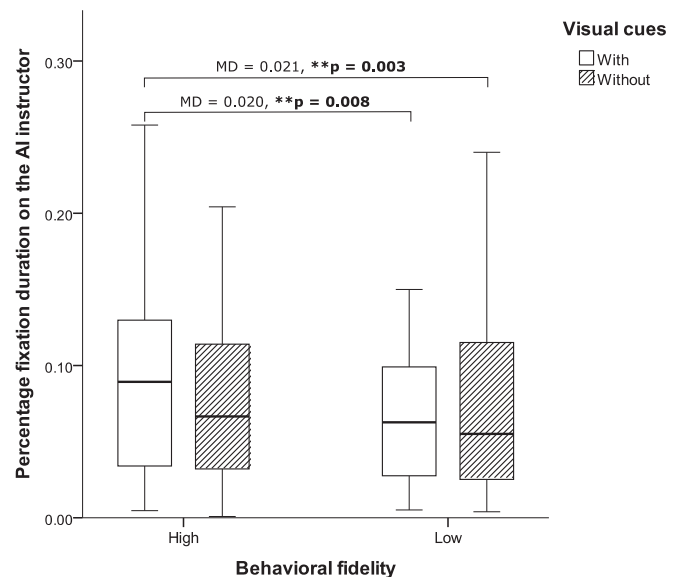


Fig. 5. Differences in the percentage fixation duration on the AI instructor across four conditions.

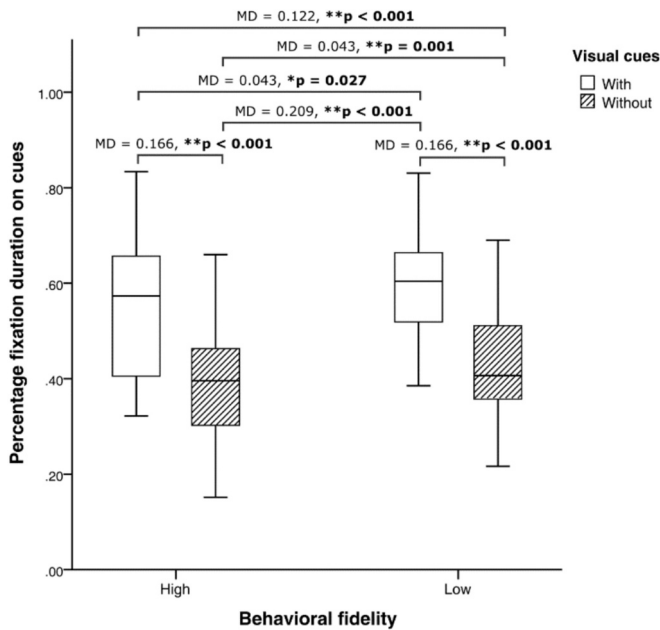


Fig. 6. Differences in the percentage fixation duration on cues across four conditions.

group demonstrated a significantly higher percentage of fixation duration on the AI instructor compared to the LBC group. This suggests that the high behavioral fidelity of the AI teacher captured the participants' attention more, leading them to focus more on and listen to the AI teacher's lectures. The LBC and HBC groups showed significantly higher percentage of fixation durations on cues compared to the LB and HB groups, respectively. This indicates that visual cues effectively directed students' attention to signaling areas, helping students focus on critical learning content. The findings supported H3.

3.3.2. Organizing and integrating information

The Friedman analysis revealed significant differences in the percentage of saccades on learning materials across the four groups ($\chi^2(3) = 10.233, p = 0.017$). The HBC group exhibited a significantly higher percentage of saccades directed towards learning materials compared to

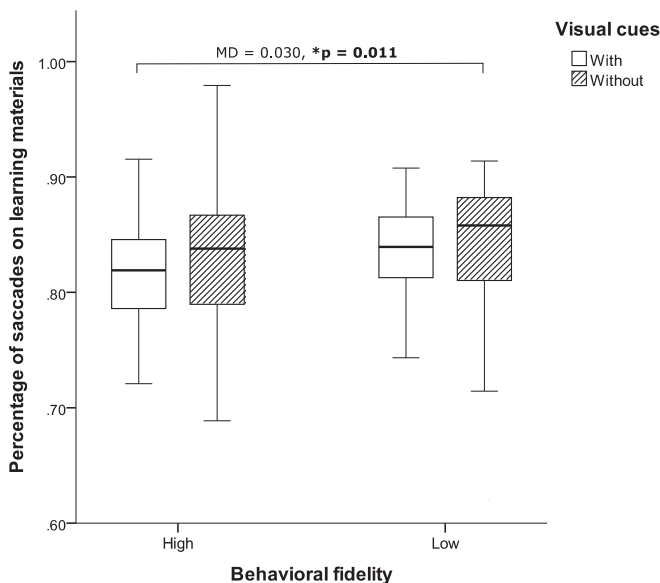


Fig. 7. Differences in the percentage of saccades on learning materials across four conditions.

the LB group (Mean difference = 0.030, $p = 0.011$), as shown in Fig. 7.

Hypothesis 4 predicted that the presence of visual cues is anticipated to facilitate the organization and integration of information. The HBC group demonstrated a significantly lower percentage of saccades compared to the LB group, indicating that the cues when paired with a high-fidelity AI instructor helped participants organize and integrate information. The findings partially supported H4.

3.4. Suggestions and comments

Through coding and analyzing their comments, we found a greater emphasis on behavioral fidelity than visual cues. Besides, participants indicated the need for enhancements in the machine voice, facial expressions, and body movements. Notably, ten participants from the HBC and HB groups highlighted the necessity for improvements in facial expressions and body movements. For example, Participant 7 expressed, "...All attention is on the arms..." Participant 17 said, "...The teacher's movements are very mechanical..." Participant 30 reported, "This video includes the teacher's body movements, but I find them too simplistic." Participant 36 suggested, "...The instructor's gestures could be more varied..." The original participant comment quotes and their translations are presented in Appendix Table A2 for reference."

4. Discussion

4.1. Empirical contributions

4.1.1. High behavioral fidelity increased social presence and more effectively engages students, leading to better transfer performance

The results revealed no significant differences in retention across conditions; however, transfer differed significantly. The HBC group outperformed significantly the LBC group and LB groups, and the HB group outperformed significantly the LB group. Behavioral fidelity demonstrated a significant main effect on transfer. Regarding subjective scales, social presence differed significantly across the four groups, with learners in high-fidelity instructor groups reporting higher social presence compared to those in low-fidelity instructor groups. The eye-tracking data revealed that participants allocated significantly more attention to the high-fidelity instructor than the low-fidelity instructor. The HBC group exhibited a significantly higher percentage of saccades directed towards learning materials compared to the LB group, indicating high-fidelity instructors are beneficial for organizing and integrating information. Our findings regarding the impact of high behavioral fidelity in AI instructors on performance partially align with previous research on human instructors (Gluhareva & Prieto, 2017; Pi, Zhu, Zhang, & Yang, 2021). This can be explained by the social agency theory (Atkinson et al., 2005). Higher behavioral fidelity in virtual instructors strengthened the social interactions perceived by students. These enhanced social cues then increased learners' attentiveness to the instructors, resulting in increased focus on and engagement with the AI teacher's lectures (Mayer, 2014). Despite the greater attention directed towards the high-fidelity instructors, there was no observed decrease in attention to the learning materials. The percentage of fixation duration on learning materials across the four groups showed no significant differences. As a result, learners become more engaged in their learning, leading to better transfer performance.

However, the results from the suggestions and comments of participants indicated that while high behavioral fidelity enhances social presence and improves performance, the mechanical rhythm of beat gestures may disturb students. Therefore, the design of the AI-generated instructor's behavioral fidelity requires further refinement.

4.1.2. Dynamic visual cues were found to facilitate the organization and integration of information

Visual cues demonstrated a significant main effect on transfer. High behavioral fidelity with the presence of visual cues yielded the greatest

enhancements in transfer. Additionally, the results, which indicated no significant differences in retention but significant differences in transfer, suggested that while behavioral fidelity and visual cues did not notably aid in English word recall, they actively supported learners in integrating and organizing knowledge. These findings proved the positive effects of visual cues (Van Gog, 2014) and showed that visual cues can augment AI-generated instructors to provide indicative assistance for students.

The eye-tracking data on fixation durations and saccades provide clear evidence of how cues influence retention and transfer. The LBC and HBC groups exhibited significantly higher fixation durations on cues than the LB and HB groups, respectively. This supports the attention-guiding hypothesis (Betancourt, 2005), showing that cues effectively directed attention to signaling areas, helping students focus on critical learning content. Dynamic visual cues not only highlighted the structure of learning materials but also provided a clear guide for participants to follow. Additionally, participants in the HBC group showed significantly fewer saccades on learning materials than LB group, indicating that the cues facilitated the organization and integration of information. These findings align with previous research, which has demonstrated that visual cues facilitate the integration of image information by providing visual-spatial guidance (Moon & Ryu, 2021). This effect was significant when paired with a high-fidelity AI instructor. The cue in this study, synchronized with oral language, plays an organizational and integrative role (De Koning et al., 2009). It effectively supports the integration of visuals (including text and charts) with oral instruction, thereby enhancing the students' performance to transfer knowledge effectively. This aligns with the underlying mechanism of cues: they influence performance by regulating attentional control and selective visual processing (Zhou & Wu, 2024). Taken together, these findings demonstrated that visual cues could enhance the effectiveness of AI-generated instructors by offering visual scaffolding for students. They fill a critical gap in understanding how dynamic visual cues impact vocabulary learning in instructional videos (Pi et al., 2023; Wang et al., 2020).

4.2. Theoretical contributions

This study makes significant contributions to our understanding of AI-generated instructors' body expressions and visual cues in instructional videos. To our knowledge, this is one of the first studies to explore the boundary design of AI-generated instructors in this context. Existing theoretical proposals and previous studies on instructor gestures have primarily focused either on real teachers in video lectures (Beege et al., 2020; Pi, Zhu, Zhang, & Yang, 2021) or on pedagogical agents that are not as realistic and human-like (e.g., Moon & Ryu, 2021). Research on the impact of generated content and synthesized teachers on learning in instructional videos remains scarce (e.g., Dao et al., 2021; Leiker et al., 2023; Pi et al., 2022). Social cues and visual cues are critical design factors in instructional videos. This study's examination on the effects of behavioral fidelity of realistic and human-like AI-generated instructors and visual cues addresses an important gap in the existing research, paving the way for a deeper understanding of AI-generated instructors' potential in education.

The findings highlight the importance of human-like behavior in AI-generated instructors, and contribute to social agency theory and the embodiment principles by providing empirical evidence that high behavioral fidelity enhances the effectiveness of AI-generated instructors. Realistic beat gestures, body movements, and facial expressions were shown to improve transfer performance. This study expands these the social agency theory (Atkinson et al., 2005) and the embodiment principles (Mayer, 2014) related to animated instructors or human instructors in classical video learning to the context of AI-generated instructors. Furthermore, this research contributes to the development of the cueing principle (Van Gog, 2014) and the attention-guiding hypothesis (Betancourt, 2005) by providing evidence of the positive effects of incorporating visual cues to assist AI-generated instructors in

instructional videos. The inclusion of dynamic visual cues significantly improved learners' focus on critical content areas, and facilitated the organization and integration of information. This study refines the hypothesis by showing that the effectiveness of visual cues is amplified when paired with high-fidelity behavioral features, suggesting that attention guidance is optimized through a combination of dynamic visuals and human-like behavior.

Nonetheless, it should be noted that this study conducted in the context of vocabulary learning, which involved more memorization and knowledge integration. Other subjects like STEM education often emphasize reasoning and problem-solving. Thus, further research is required to determine whether these findings can be applied more broadly across different subjects.

4.3. Practical implications

The findings of the present study provide practical implications for the design of an AI-generated instructors' body expressions and visual cues in instructional videos. First, this study revealed that students' learning performance, satisfaction, social presence, and motivation were most enhanced when the instructor was generated with high behavioral fidelity. Therefore, we recommend using instructors with high behavioral fidelity in instructional videos, featuring realistic beat gestures, body movements, facial expressions, and eye contact. Second, this study observed that the effectiveness of high-behavioral fidelity AI-generated instructors was further improved when visual cues were incorporated. Visual cues helped reduce students' cognitive load. Thus, we recommend augmenting AI-generated instructors with visual cues, especially when indicative gestures cannot be easily generated.

5. Conclusion, limitations and future directions

This study is one of the first to investigate the impact of AI-generated content in education, which explored the effects of behavioral fidelity of AI-generated instructors and the presence of visual cues on learning. It provided empirical evidence on how these factors influence educational outcomes. The findings highlight the transformative potential of AIGC in instructional videos. By generating talking instructors from text and images or videos, AIGC offers a groundbreaking approach to the development of educational content. Our research provides critical insights into the instructional design boundaries of such AI-generated content and demonstrates how different levels of behavioral fidelity and the presence or absence of visual cues impact learning outcomes.

However, this study faces a few limitations that should be addressed in future investigations. First, in our current study, we investigated the impact of two levels of behavioral fidelity on learning outcomes, utilizing driving videos that included only beat gestures. In future research, we aim to refine the dimensions of behavioral fidelity by incorporating a broader range of teacher gestures and facial expressions. We will examine the effects of these additional behaviors on learning and explore the interaction effects between facial expressions and gestures. Additionally, this study did not include a direct comparison between a real human instructor and an AI-generated teacher. Although prior studies (Netland et al., 2025; Pi et al., 2022; Xu et al., 2024) have explored the effectiveness of human and AI instructors in instructional videos, further comprehensive evaluations are required to deepen our understanding of their comparative impacts on learning outcomes, experience and student behaviors. Second, the learning intervention lasted only about an hour. Future research should extend these studies to cover longer periods, such as a semester, to evaluate the long-term effects of AI-generated content. Third, the study focused solely on university students, limiting its applicability to other populations. Future studies should include K-12 students to explore the impact across different age groups. Next, the research was confined to vocabulary building. Further research should examine other subject areas like engineering or STEM. Lastly, the study relied mainly on quantitative

methods, which may overlook the complexity of interactions with these generated contents. Future research should incorporate more qualitative methods to gain deeper insights.

Consent to participate

All participants signed informed consent agreement.

Ethics approval statement

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

Funding statement

This work was supported by the Research Projects of the National Natural Science Foundation of China under Grant [62077036] and [62377039]; the National Key Research and Development Program of China under Grant [2022YFC3303600]; CAAI-CANN Open Fund,

developed on OpenI Community; Higher Education Research Fund of Northwestern Polytechnical University in 2025, No. GJJJM202504.

CRedit authorship contribution statement

Tao Xu: Conceptualization, Methodology, Writing – original draft, Funding acquisition. **Qian Chen:** Formal analysis, Software. **Zhiwei Zhang:** Software. **Bingyu Dong:** Formal analysis, Visualization. **Hui Zhang:** Formal analysis, Visualization. **Jie Bai:** Methodology. **Yun Zhou:** Conceptualization, Supervision, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of competing interest

Authors have no competing interests to declare.

Acknowledgements

We appreciate the participants for their engagement in this research.

Appendix A

Table A1
The 60-word pool and 24 selected words.

Words	Number of students unfamiliar with the word	Familiarity	Selected	Condition
Anachronism	23	1.35	YES	HB
Destitute	19	2.3	YES	HB
Grandiloquent	22	1.57	YES	HB
Pique	21	1.87	YES	HB
Sporadic	21	1.83	YES	HB
Succumb	19	2.17	YES	HB
Dissuade	20	2	YES	HBC
Dyslexia	21	1.48	YES	HBC
Nescience	20	2.04	YES	HBC
Omnivore	21	1.65	YES	HBC
Panoply	22	1.39	YES	HBC
Peripheral	21	1.65	YES	HBC
Amortize	22	1.57	YES	LB
Eulogy	22	1.35	YES	LB
Fiduciary	22	1.39	YES	LB
Juncture	20	1.65	YES	LB
Malevolent	20	2.13	YES	LB
Protrude	18	2.09	YES	LB
Cordial	19	2.35	YES	LBC
Dystrophy	22	1.39	YES	LBC
Elucidate	22	1.65	YES	LBC
Epilogue	22	1.43	YES	LBC
Extrapolate	19	2.09	YES	LBC
Nemesis	20	1.74	YES	LBC
Apparatus	16	2.83	NO	
Appendage	18	2.26	NO	
Baseless	5	4.87	NO	
Cease	11	3.78	NO	
Circumspection	17	2.39	NO	
Comport	8	4.09	NO	
Contaminate	16	3.04	NO	
Contradict	11	3.96	NO	
Corpus	18	2.35	NO	
Counsel	14	3.22	NO	
Embankment	14	2.87	NO	
Emissary	22	1.61	NO	
Eviscerate	19	2.35	NO	
Fabricate	12	3.17	NO	
Fatal	18	2.83	NO	
Fluctuate	12	3.39	NO	
Friction	14	3.13	NO	
Goggle	10	4.09	NO	
Handicap	13	3.04	NO	
Inarticulate	16	2.7	NO	

(continued on next page)

Table A1 (continued)

Words	Number of students unfamiliar with the word	Familiarity	Selected	Condition
Indulge	15	3.04	NO	
Instinctively	10	3.83	NO	
Intimidate	11	3.7	NO	
Opaque	19	1.91	NO	
Plagiarism	21	1.87	NO	
Ponderous	18	2.52	NO	
Portfolio	17	2.43	NO	
Relentless	18	2.61	NO	
Seductive	16	2.61	NO	
Segment	16	2.83	NO	
Siren	20	1.91	NO	
Somnambulism	21	1.43	NO	
Translucent	16	2.57	NO	
Turbulent	15	2.78	NO	
Voluptuous	21	1.74	NO	
Withstand	6	4.7	NO	

Note: A lower score of word familiarity reflects a lower level of familiarity with the word among students. The number of students unfamiliar with the word was determined by counting those who rated the word below 4.

Table A2

The original and the translation of participants' comments in Section 3.4, suggestions and comments.

Participant number	Condition	Original (Chinese)	Translation
P7	HBC	注意力全在胳膊上	...All attention is on the arms...
P17	HBC	老师的动作很机械	...The teacher's movements are very mechanical...
P30	HB	这个视频添加了教师的肢体动作，但是自我感觉肢体动作过于简单。	This video includes the teacher's body movements, but I find them too simplistic.
P36	HBC	授课教师的手势可以多样一点	...The instructor's gestures could be more varied...

Data availability

The analysis code and data are available at the following link: https://osf.io/u6sfb/?view_only=7f5227c50ab0460280ef55b78a57730d; The Chinese version of subjective scales is available from the corresponding author upon reasonable request.

References

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing work productivity through generative artificial intelligence: A comprehensive literature review. *Sustainability*, 16(3), 1166.

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, 166, Article 104154. <https://doi.org/10.1016/j.compedu.2021.104154>

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.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62.

Beege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., ... Rey, G. D. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education*, 156, Article 103955. <https://doi.org/10.1016/j.compedu.2020.103955>

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). Cambridge University Press.

Byun, K. J., Ahn, S. J., & (Grace). (2023). A systematic review of virtual influencers: Similarities and differences between human and virtual influencers in interactive advertising. *Journal of Interactive Advertising*, 23(4), 293–306. <https://doi.org/10.1080/15252019.2023.2236102>

Cassell, J., & McNeill, D. (1991). Gesture and the poetics of prose. *Poetics Today*, 12(3), 375–404.

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

Cohen, J. (2016). *A power primer* (p. 284). American Psychological Association. <https://doi.org/10.1037/14805-018>

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>

Craig, S. D., & Schroeder, N. L. (2019). Text-to-speech software and learning: Investigating the relevancy of the voice effect. *Journal of Educational Computing Research*, 57(6), 1534–1548. <https://doi.org/10.1177/0735633118802877>

Crook, C., & Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, 34, 56–64.

Crooks, S. M., Cheon, J., Inan, F., Ari, F., & Flores, R. (2012). Modality and cueing in multimedia learning: Examining cognitive and perceptual explanations for the modality effect. *Computers in Human Behavior*, 28(3), 1063–1071. <https://doi.org/10.1016/j.chb.2012.01.010>

Crossley, S., Salsbury, T., Titak, A., & McNamara, D. (2014). Frequency effects and second language lexical acquisition: Word types, word tokens, and word production. *International Journal of Corpus Linguistics*, 19(3), 301–332.

Dao, X.-Q., Le, N.-B., & Nguyen, T.-M.-T. (2021). AI-powered MOOCs: Video lecture generation. In *Proceedings of the 2021 3rd international conference on image, video and signal processing* (pp. 95–102). <https://doi.org/10.1145/3459212.3459227>

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>

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, 21(2), 113–140. <https://doi.org/10.1007/s10648-009-9098-7>

Demirel, H. O., Ahmed, S., & Duffy, V. G. (2022). Digital human modeling: A review and reappraisal of origins, present, and expected future methods for representing humans computationally. *International Journal of Human-Computer Interaction*, 38(10), 897–937. <https://doi.org/10.1080/10447318.2021.1976507>

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 P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies. Novel technological environments* (pp. 176–187). Springer International Publishing. https://doi.org/10.1007/978-3-031-05675-8_14

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavior, and biomedical sciences. *Behavior Research Methods, Instruments, & Computers*, 39, 175–191. <https://doi.org/10.3758/BF03193146>

Frenkel, J., Cajar, A., Engbert, R., & Lazarides, R. (2024). Exploring the impact of nonverbal social behavior on learning outcomes in instructional video design. *Scientific Reports*, 14(1), Article 12867.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, 21(5), 609–631. <https://doi.org/10.1177/1362168816651463>

- Guo, D., Chen, H., Wu, R., & Wang, Y. (2023). AIGC challenges and opportunities related to public safety: A case study of ChatGPT. *Journal of Safety Science and Resilience*, 4 (4), 329–339. <https://doi.org/10.1016/j.jssr.2023.08.001>
- Hajhashemi, K., Caltabiano, N., Anderson, N., & Tabibzadeh, S. (2018). Multiple intelligences, motivations and learning experience regarding video-assisted subjects in a rural university. *International Journal of Instruction*, 11, 167–182. <https://doi.org/10.12973/iji.2018.11112a>
- Henderson, M. L., & Schroeder, N. L. (2021). A systematic review of instructor presence in instructional videos: Effects on learning and affect. *Computers and Education Open*, 2, Article 100059. <https://doi.org/10.1016/j.caeo.2021.100059>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. oup Oxford.
- Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, 32, 47–53. <https://doi.org/10.1016/j.chb.2013.11.013>
- 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>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329.
- Kang, H. (2021). Sample size determination and power analysis using the G*power software. *Journal of Educational Evaluation for Health Professions*, 18. <https://doi.org/10.3352/jeehp.2021.18.17>
- Kanungo, R. (1967). Meaning mediation in verbal transfer. *British Journal of Psychology*, 58(3–4), 205–212.
- Katuk, N., Kim, J., & Ryu, H. (2013). Experience beyond knowledge: Pragmatic e-learning systems design with learning experience. *Computers in Human Behavior*, 29 (3), 747–758. <https://doi.org/10.1016/j.chb.2012.12.014>
- 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>
- 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.
- Leiker, D., Gyllen, A. R., Eldesouky, I., & Cukurova, M. (2023). Generative AI for learning: Investigating the potential of learning videos with synthetic virtual instructors. In N. Wang, G. Rebollo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), 1831. *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (pp. 523–529). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_81
- Leroy, G., & Kauchak, D. (2014). The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association*, 21(e1), e169–e172. <https://doi.org/10.1136/amiajnl-2013-002172>
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.004>
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171. <https://doi.org/10.1016/j.learninstruc.2009.02.012>
- Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In 16. *The Cambridge handbook of multimedia learning* (pp. 345–370).
- Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition*, 10(2), 229–240.
- 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>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Moon, J., & Ryu, J. (2021). The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction. *Journal of Computing in Higher Education*, 33(1), 39–63. <https://doi.org/10.1007/s12528-020-09255-x>
- 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, Article 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. <https://doi.org/10.3389/frobt.2018.00114>
- Olsen, A. (2012). The Tobii i-VT fixation filter. *Tobii Technology*, 21(4–19), 5.
- Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., ... Wong, L.-W. (2023). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, 1–32. <https://doi.org/10.1080/08874417.2023.2261010>
- Paas, F. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1), 419–430. <https://doi.org/10.2466/pms.1994.79.1.419>
- 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>
- 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>
- Pi, Z., Liu, C., Wang, L., Yang, J., & Li, X. (2023). Cues facilitate foreign language vocabulary learning from instructional videos: Behavioral and neural evidence. *Language Teaching Research*. <https://doi.org/10.1177/13621688231164724>, 13621688231164724.
- Pi, Z., Zhu, F., Zhang, Y., & Yang, J. (2021). An instructor's beat gestures facilitate second language vocabulary learning from instructional videos: Behavioral and neural evidence. *Language Teaching Research*, 28(5), 1997–2025. <https://doi.org/10.1177/13621688211039023>
- Prajwal, K. R., Mukhopadhyay, R., Namboodiri, V., & Jawahar, C. V. (2020). A lip sync expert is all you need for speech to lip generation in the wild. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 484–492). <https://doi.org/10.1145/3394171.3413532>
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. <https://doi.org/10.1016/j.edurev.2015.12.003>
- Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, 23, 1–24. <https://doi.org/10.1016/j.edurev.2017.11.001>
- So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language & Cognitive Processes*, 27(5), 665–681. <https://doi.org/10.1080/01690965.2011.573220>
- Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.014>
- Wang, X., Lin, L., Han, M., & Spector, J. M. (2020). Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos. *Computers in Human Behavior*, 107, Article 106279. <https://doi.org/10.1016/j.chb.2020.106279>
- 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, H.-Y. (2016). The effects of attention cueing on visualizers' multimedia learning. *Journal of Educational Technology & Society*, 19(1), 249–262.
- Zhang, C., Wang, Z., Fang, Z., & Xiao, X. (2024). Guiding student learning in video lectures: Effects of instructors' emotional expressions and visual cues. *Computers & Education*, 218, Article 105062. <https://doi.org/10.1016/j.compedu.2024.105062>
- Zheng, X., Ma, Y., Yue, T., & Yang, X. (2023). Effects of different types of cues and self-explanation prompts in instructional videos on deep learning: Evidence from multiple data analysis. *Educational Technology Research and Development*, 71(3), 807–831. <https://doi.org/10.1007/s11423-023-10188-2>
- Zhou, W., & Wu, X. (2024). The impact of internal-generated contextual clues on EFL vocabulary learning: Insights from EEG. *Frontiers in Psychology*, 15, Article 1332098.

Tao Xu, Ph.D., is an associate professor from the School of Software at Northwestern Polytechnical University. His research interests include AI in education, AI applications, and brain-computer interfaces.

Qian Chen is a master student from the Faculty of Education at Shaanxi Normal University. Her research interests include AI in education and media in education.

Zhiwei Zhang is a master student from the School of Software at Northwestern Polytechnical University. His research interests include AI applications and AI in education.

Bingyu Dong is a master student from the Faculty of Education at Shaanxi Normal University. Her research interests include AI in education and media in education.

Hui Zhang is a master student from the Faculty of Education at Shaanxi Normal University. Her research interests include AI in education and media in education.

Jie Bai is a Ph.D. student from the Faculty of Education at Shaanxi Normal University. Her research interests include AI in education and media in education.

Yun Zhou, Ph.D., is a full professor from the Faculty of Education at Shaanxi Normal University. Her research interests include AI in education, media in education, and educational brain-computer interfaces.