Received: 17 February 2024 | Accepted: 13 October 2024
DOI: 10.1111/bjet.13530

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

From recorded to AI-generated instructional videos: A comparison of learning performance and experience

Tao Xu[1] | **Yuan Liu**[1] | **Yaru Jin**[2] | **Yueyao Qu**[2] | **Jie Bai**[2] | **Wenlan Zhang**[2] | **Yun Zhou**[2]

[1]School of Software, Northwestern Polytechnical University, Xi'an, P.R. China

[2]Faculty of Education, Shaanxi Normal University, Xi'an, P.R. China

Correspondence
Yun Zhou, Faculty of Education, Shaanxi Normal University, South Chang'an Road 199, Yanta District, Xi'an, Shaanxi Province 710062, P.R. China.
Email: zhouyun@snnu.edu.cn

Funding information
National Key R&D Program of China, Grant/Award Number: 2022YFC3303600; Research Projects of the National Natural Science Foundation of China, Grant/Award Number: 62077036 and 62377039; CAAI-CANN Open Fund; Higher Education Research Fund of Northwestern Polytechnical University in 2025, Grant/Award Number: No.GJJJM202504

Abstract: Generative AI (GAI) and AI-generated content (AIGC) have been increasingly involved in our work and daily life, providing a new learning experience for students. This study examines whether AI-generated instructional videos (AIIV) can facilitate learning as effectively as traditional recorded videos (RV). We propose an instructional video generation pipeline that includes customized GPT (Generative Pre-trained Transformer), text-to-speech and lip synthesis techniques to generate videos from slides and a clip or a photo of a human instructor. Seventy-six students were randomly assigned to learn English words using either AIIV or RV, with performance assessed by retention, transfer and subjective measures from cognitive, emotional, motivational and social perspectives. The findings indicate that the AIIV group performed as well as the RV group in facilitating learning, with AIIV showing higher retention but no significant differences in transfer. RV was found to offer a stronger sense of social presence. Although other subjective measures were similar between the two groups, AIIV was perceived as slightly less favourable. However, the AIIV was still found to be moderately to highly attractive, addressing concerns related to the uncanny valley effect. This research demonstrates that AIGC can be an effective tool for learning, offering valuable implications for the use of GAI in educational settings.

KEYWORDS
AI-generated content, generative AI, media in education

Practitioner notes

What is already known about this topic

- Instructional videos, especially those featuring a teacher's presence, have been widely used in second language learning to facilitate learning.
- Producing instructional videos is costly and burdensome.
- Generative AI has great potential for generating educational content.

What this paper adds

- An AI-generated instructional video (including generated lecture text, voice and appearance) demonstrated greater improvement in students' retention performance in English word learning than a traditional recorded video.
- Students perceived no significant differences between the AI-generated instructional video and recorded video in satisfaction, motivation, trust, cognitive load, emotions and parasocial interaction dimensions, although the AI-generated instructional video group reported slightly lower values.
- Despite AI-generated instructional video eliciting a significantly lower value of social presence than recorded video, it led to a reduction in cognitive load and better performance.

Implications for practice and/or policy

- We recommend using the AI-generated instructional video in both physical and online classes for its positive effects on both learning achievement and learning experience.
- The findings indicate the equivalence principle in AI-generated content, highlighting that the appearance, voice and lecture text generated by current AI technology have reached a certain level of quality.

INTRODUCTION

Generative AI (GAI) refers to leveraging AI technology to automate the generation of diverse multimodal data, including text, images, videos and audio (Baidoo-Anu & Owusu Ansah, 2023; Ooi et al., 2023). This has triggered increased attention and discussions in the field of education (Jeon et al., 2023; Jeon & Lee, 2023; Kasneci et al., 2023). With advancements in deep learning, computing ability and increased availability of big data, GAI technologies, relying on a foundation model (Bommasani et al., 2021), can achieve impressive content generation results. It is evident that GAI will play an increasingly prominent role in all aspects of students' learning and teachers' instruction. In the future, stakeholders in education may find themselves working in an environment where GAI is omnipresent.

Instructional videos have been widely used in second language learning, especially in English learning (Chen et al., 2020; Hsieh, 2020; Shen et al., 2021), and their outstanding teaching effectiveness has been proven (eg, BavaHarji et al., 2014). This form of instructional medium, by providing visual and auditory stimuli, can convey concepts through vivid images and dynamic demonstrations. However, producing instructional videos with human teachers in the traditional way involves higher financial and time costs, imposing an obvious burden on teachers. The instructional videos contain various types and designs (Crook & Schofield, 2017; Köse et al., 2021). It should be noted that the video design in this study for both 'recorded videos' and 'AI-generated videos' follows the B3 fixed but overlapping design defined by Crook and Schofield (2017). This design features a video narrator at a fixed position overlapping the background sequence, rather than being framed as a picture-in-picture (as illustrated in the bottom right of Figure 1). This type of instructional video includes slides, the teacher's lecture and the teacher's image. The traditional production procedure for this type typically involves the following steps: (1) planning the instructional design, creating slides and preparing scriptwriting; (2) recording, where teachers need repeated practice to ensure fluent language delivery, proficient facial expressions and gestures; and (3) postproduction processes. The first two steps consume a considerable amount of time as teachers iterate through modifications and rehearsals until they are satisfied.

To address this issue, the educational use of AI-generated data and AI-generated content (AIGC) (Bozkurt et al., 2021; Cao et al., 2023) could be a feasible alternative. AIGC can automate the first two steps mentioned above, thus replacing traditional production methods. Large language models enable the generation of lecture scripts (Baidoo-Anu & Owusu Ansah, 2023). Furthermore, through algorithms and models, lecture videos with talking teachers can be rapidly and accurately produced. In these videos, AI-generated characters resemble human teachers delivering lectures (Pataranutaporn et al., 2021). The fast generation of content alleviates the burden on teachers, allowing them to focus more on instructional design and making the production of instructional videos more accessible. Given this development, it may also give rise to certain issues, including how these generated videos impact students' learning and how learners perceive the content generated in this way. However, relatively little research has been conducted on AIGC and its impact on learning, with even fewer studies focusing on AI-generated instructional videos.

The trajectory of ChatGPT in education

Since its launch, ChatGPT has attracted considerable attention, being employed to explore the limitless possibilities and diverse applications of its integration in education (Bozkurt et al., 2021). This exploration has extended to investigating ChatGPT's potential in English language learning, emerging as a focal point in the ongoing discourse. Mohamed (2023) highlighted ChatGPT's potential in improving English language proficiency among EFL (English as a Foreign Language) students, suggesting the necessity for further experimental research to assess its effectiveness. This positive outlook on ChatGPT's potential is also echoed in related studies (Guo & Wang, 2023; Young & Shishido, 2023). Furthermore, the ongoing iteration and improvement of GPT-like technologies have significantly impacted content generation and natural human–computer interaction in the field of education (Lo, 2023; Rahman & Watanobe, 2023; Singhal et al., 2023). These contents include but are not limited to the production and recommendation of materials, customization of materials and provision of cultural knowledge (Jeon & Lee, 2023). Despite these advancements, exploration into AIGC for educational purposes is in its early stages, lacking empirical research validation.

When using ChatGPT to assist in teaching, the design and implementation of prompt engineering are crucial. Research has indicated that well-designed prompts can ensure that AIGC meets the requirements of learning objectives (Lee et al., 2023). Another approach involves the custom versions of ChatGPT. In November 2023, OpenAI introduced GPTs, customized versions of ChatGPT designed for specific purposes, allowing users to tailor ChatGPT for tasks such as learning board game rules, teaching math to kids and sharing

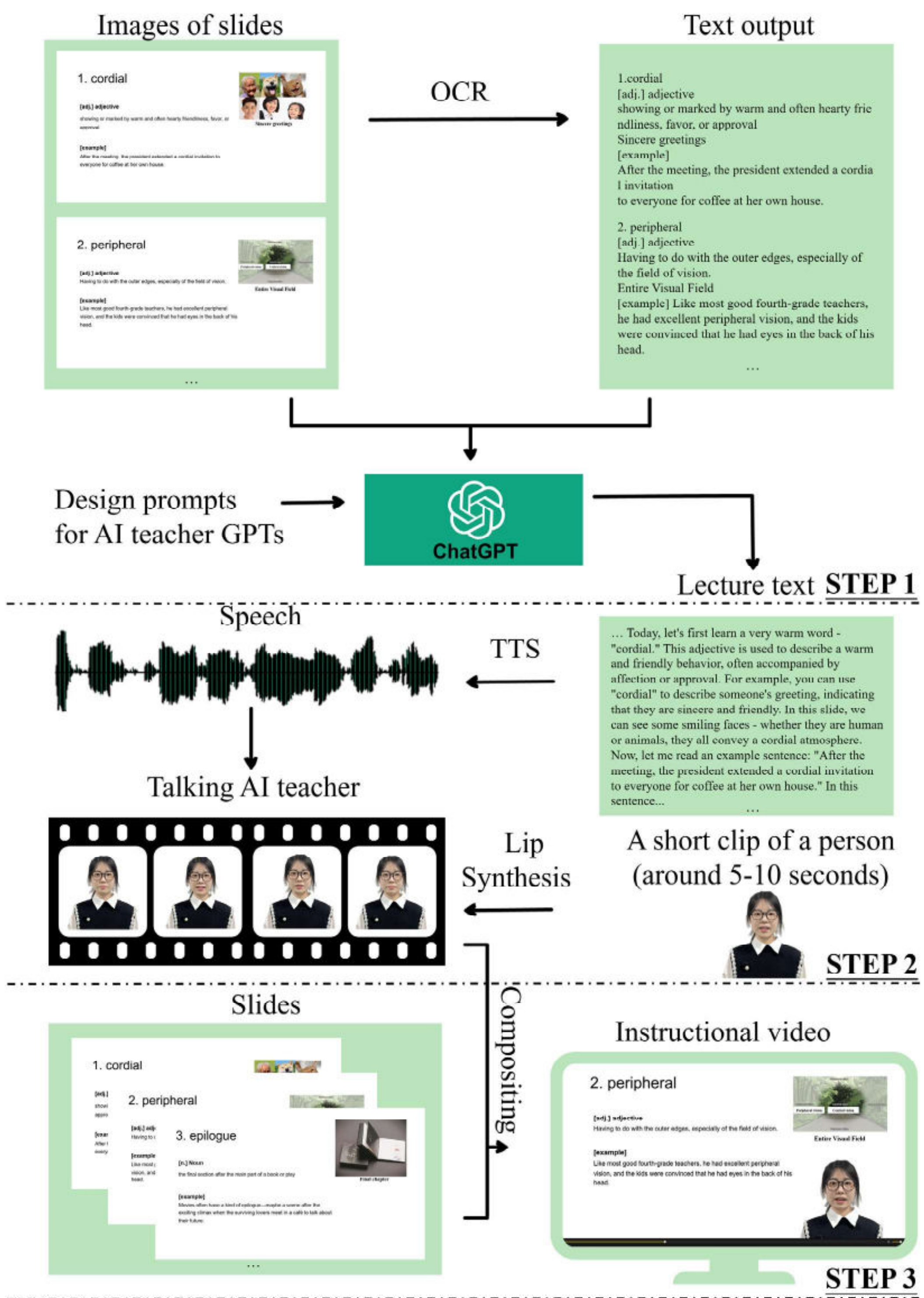

FIGURE 1 The instructional video generation pipeline (with translated examples inside).

these tailored versions with others. Through customized models, generated content can be more professional and accurate (Eloundou et al., 2023). However, even with prompting techniques and customized approaches guiding AIGC, there is still a need to evaluate whether the generated content, such as lecture text, is reliable and indeed promotes learning.

The effects of machine voices on learning

The voice in instructional videos also impacts learners' learning (Lawson & Mayer, 2022). Earlier perspectives suggested that human voices were preferred by learners and more effectively facilitated learning than synthetic voices (Mayer, 2014). This is attributed to human voices having higher social presence and emotional appeal. However, the low effectiveness of old machine voices in learning outcomes may have been due to technological limitations. With advancements in AI text-to-speech technology, machine voices have become increasingly human like, with better naturalness and fidelity. In this context, researchers have reexamined the impact of machine voices on learning. Craig and Schroeder (2017, 2019) found that machine voices could be as effective as human voices, and the modern text-to-speech engine outperformed the older one in facilitating learning (Craig & Schroeder, 2017). Davis et al. (2019) found that in terms of information retention, machine voices were as effective as human voices and the factors influencing voice effectiveness are more complex than a simple human or machine categorization. Chiou et al. (2020) explored how the quality of a virtual human's voice influences learning, perceptions and trust, revealing that voice quality did not significantly impact performance but did impact trust and learners' perceptions of the virtual human. Furthermore, in English vocabulary learning, combining synthesized voices with likeable images can produce motivational effects the same as those with real humans, but the motivational effect is lower when combined with ordinary images (Pi et al., 2022). For the K-12 group, synthesized voices can even enhance retention and transfer scores more than human voices (Deng et al., 2022), which may be due to the novelty effect. Compared to human voices, synthesis technology provides clearer pronunciation and a standard-accented voice. Therefore, modern machine voices have the potential to outperform human voices, as teachers, despite their training, may still have varying degrees of accent. In this study, we need to reexamine this voice effect. Moreover, despite researchers studying the impact of machine voices on credibility, social aspects, learning motivation and cognitive load (Craig & Schroeder, 2019; Edwards et al., 2018; Kim et al., 2022; Liew et al., 2020; Schroeder et al., 2021), such studies are still in their early stages. The effects of machine voices on emotions, the uncanny valley and parasocial interaction are still unknown, and the present study assesses these dimensions of experience extensively.

The effects of the teacher's presence

The positive effects of featuring pedagogical agents or virtual teachers in multimedia learning (Castro-Alonso et al., 2021), or the use of on-screen teachers in instructional videos (Beege et al., 2023), have been observed. That is, the presence of real or virtual teachers in learning resources enhances learning outcomes and experiences (eg, Colliot & Jamet, 2018; Wang & Antonenko, 2017). The instructor's image or appearance is a crucial factor in video learning (Kizilcec et al., 2015; Wilson et al., 2018). A key underlying theory is social presence theory (Martha & Santoso, 2019). Social presence refers to the degree to which individuals feel their existence, attention and recognition by others in a social context (Borup et al., 2012; Lyons et al., 2012; Thomas et al., 2017). The social presence evoked by teachers in instructional videos involves learners' perception of the teacher's presence, interaction and influence in the video, considered a critical factor in promoting learners' active engagement and involvement. When pedagogical agents or virtual teachers have a human-like appearance and body, displaying human-like movements and expressions, participants report higher levels of trust and social presence (eg, Kim et al., 2018; Veletsianos et al., 2009). With the rapid evolution of lip synthesis technology, virtual characters can now be synthesized with more vivid facial expressions, dynamic eye movements and expressive gestures. This increased realism in generation could enhance learning, but it may also trigger the uncanny valley effect (Mori et al., 2012), leading to decreased trust, lower learning experiences and negative emotions. Previous studies on virtual teachers synthesized based on images, such as Pi et al. (2022) and Xu et al. (2021). These generated teachers only had lip synchronization for the lower part of the face, lacking micro-expressions above the nose, like blinking. Therefore, we reexamine the impact of generated appearance on learning, exploring whether improved synthesis triggers the uncanny valley effect, perceived parasocial interaction and other learning experiences.

The present study and research questions

Advancements in AI technology are enabling the generation of more processes in instructional video production, gradually replacing traditional production methods. Given these developments, it is crucial to explore the effects of such AI-generated instructional videos. However, research on the impact of generated content on learning is currently quite limited. Existing generated videos primarily focus on the synthesis of virtual teacher characters and voices (eg, Dao et al., 2021; Leiker et al., 2023; Pi et al., 2022). In these existing systems, the lecture text is written by human teachers. Yet, they have not incorporated the generation of communication, a crucial aspect of instructional content. The present study addresses this gap by evaluating generated videos, including exploring whether the synthesis of virtual teacher characters, voices and communications can effectively enhance learning outcomes, and how learners perceive instructional content generated by AI. First, we proposed an AI-generated lecture video pipeline and implemented the application based on this pipeline. Leveraging text-to-speech, lip synthesis and GPTs, we synthesized virtual teacher characters, speech and lecture text. This allows the generation of lecture videos with slides and a talking teacher, given a driven video and lecture slides.

Building on this, we investigated whether generated instructional videos could facilitate learning as recorded videos, evaluating differences in learners' performance and subjective opinions on generated and recorded videos. Plass et al.'s work (Plass et al., 2020) inspired us to collect subjective opinions from cognitive, emotional, motivational and social perspectives. Given the use of new technology and the nonverbal interactions between students and instructors, it was crucial to include an overall satisfaction dimension. Therefore, we gathered subjective opinions from a broad range of perspectives, including cognitive, emotional, motivational, social and overall satisfaction. The learning performance was evaluated through retention and transfer. The cognitive perspective involves cognitive load dimension. The emotional perspective was analysed in terms of valance, arousal and control. The motivational perspective refers to the dimension of motivation. The social perspective included social presence, trust, uncanny valley and parasocial interaction. And the overall perspective was assessed in the dimension of satisfaction.

We utilized the equivalence principle (Horovitz & Mayer, 2021), the social presence theory (Borup et al., 2012; Lyons et al., 2012; Thomas et al., 2017), emotional aspects (Park et al., 2014; St J. Neill, 1989) and previous research (Deng et al., 2022; Liew et al., 2020; Pi et al., 2022) to make predictions about the effectiveness of AI-generated instructional videos. The equivalence principle (Horovitz & Mayer, 2021) suggests that virtual instructors can serve a similar role to that of human instructors in video-based learning, indicating that people respond to the emotional tone of computerized onscreen agents just as they do to human instructors. Based on this principle, we hypothesize that AI-generated instructional videos will yield learning outcomes comparable to those of traditional recorded videos. Since previous work found that a calm voice prompted a higher germane load than an enthusiastic voice (Liew et al., 2020), we hypothesize that AI instructors will raise a higher cognitive load than human teachers in instructional videos. St J. Neill (1989) found that children's perception of teachers' emotions is significantly influenced by facial expressions, whereas posture and gestures have a relatively minor impact. AI can now generate characters more realistically than before. However, human teachers still display micro-expressions more authentically than AI teachers. Therefore, we expect that AI instructors will evoke fewer emotional effects than human instructors in instructional videos. Pi et al. (2022) observed that synthesized voices could produce motivational effects comparable to or lower than those of real human voices. Thus, we expect that AI-generated instructors will produce fewer motivational effects than human instructors in instructional videos. The uncanny valley effect occurs when robots are highly realistic but lack detailed refinement (Mori et al., 2012). Current technology can better capture and reproduce subtle facial expressions. We believe that our AI teacher is highly realistic and well detailed, and therefore, we hypothesize that AI-generated teachers will not trigger uncanny valley effect. Previous literature suggests that human-like virtual teachers can evoke higher levels of trust and social presence (eg, Kim et al., 2018), which in turn leads to better learning performance (Mayer, 2014). Although AI-generated teachers closely resemble human teachers, they still miss some subtle details that convey liveliness. Therefore, we hypothesize that AI-generated teachers will trigger lower social presence, trust and parasocial interaction than human teachers. A recent study reported that the AI-generated instructor character achieved high interest and satisfaction scores among K-12 students (Deng et al., 2022). Based on its findings, it is reasonable to assume that AI-generated videos will achieve the same level of satisfaction as recorded videos.

Taken together, our general prediction is that AI-generated videos will facilitate learning the same as recorded videos. We proposed the following research questions and hypotheses:

RQ1: How does AI-generated instructional video impact learning outcomes (retention and transfer)?

H1. AI-generated instructional videos will yield learning outcomes comparable to those of traditional recorded videos.

RQ2: What is the impact of AI-generated instructional video on the cognitive perspective?

H2. AI instructors will raise a higher cognitive load than human teachers in instructional videos.

RQ3: How does AI-generated instructional video impact the emotional perspective?

H3. AI instructors will evoke fewer emotional effects than human instructors in instructional videos.

RQ4: What influence does AI-generated instructional video have on the motivational perspective?

H4. AI-generated instructors will produce fewer motivational effects than human instructors in instructional videos.

RQ5: How does AI-generated instructional video affect the social perspective?

H5. AI-generated teachers will trigger lower social presence, trust and parasocial interaction than human teachers.

H6. AI-generated teachers will not trigger uncanny valley effect.

RQ6: What is the impact of AI-generated instructional videos on satisfaction?

H7. AI-generated videos will achieve the same level of satisfaction as recorded videos.

METHODS

Participants and research design

A total of 76 undergraduate and graduate students (32 males and 44 females) were randomly recruited from four universities in western China, aged between 17 and 27 ($M=22.11$, $SD=2.47$). The students were majoring in educational technology, software engineering, electronic information, statistics, religion, biology, geography, chemistry, linguistics or vocational and technical education. The study was carried out in one university located in western China. A power analysis was conducted using G*Power 3.1 (Faul et al., 2007) based on the effect size ($f=0.436$, converted to $d>0.8$) found in Pi et al. (2022). The conversion was calculated using the formula from Cohen (2016). With an effect size of $d=0.8$, power of 0.80 and alpha level of 0.05, a total of 42 participants (ie, 21 for each group) at least were required for two groups to detect an effect. Therefore, the number of participants ($n=76$) in this study met the requirements. All participants had normal or corrected-to-normal visions. This study was approved by the ethics board of our institution. All participants signed an informed consent agreement and were compensated with ¥20 for their time.

The participants were randomly assigned to either the experimental group (AI-generated instructional video condition, abbreviated as AIIV) or control group (recorded video condition, abbreviated as RV), with a balance in gender and English skills. The demographic information of participants is presented in Table 1. The AIIV group ($n=38$) included 13 undergraduate students and 25 graduate students, while the RV group ($n=38$) included 11 undergraduate students and 27 graduate students. All participants were from mainland China. The evaluation of English skills was based on participants' scores in the College English Test Band 4 (CET 4) and College English Test Band 6 (CET 6) exams. CET4 and CET6 are standardized English proficiency exams in China. CET4 assesses basic English skills, while CET6 measures higher-level proficiency. The Mann–Whitney test revealed

TABLE 1 Demographic information of participants.

	AIIV ($n=38$)				RV ($n=38$)						
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	t/χ^2	df	p
Age	22.105	2.502	17	27	21.974	2.284	17	26	$t=-0.239$	74	0.811
Gender	1.579	0.500	1	2	1.579	0.500	1	2	$\chi^2=0.000$	1	1.000
Educational level	1.658	0.481	1	2	1.711	0.460	1	2	$\chi^2=0.244$	1	0.622

Note: Gender: 1, Male; 2, Female; Educational level: 1, Undergraduate; 2, Graduate.

no significant difference in English skills across the two conditions ($U=645$, $z=-0.835$, $p=0.404$, $r=-0.096$).

The learning materials

The AI-generated instructional video pipeline

Figure 1 presents our proposed pipeline for converting slides into generated instructional videos in under 10 minutes, involving three main steps: generating lecture text, creating a digital character and producing instructional videos.

The first step involves using optical character recognition (OCR) and a specialized GPT called AI Teacher to generate lecture text. This process enhances content understanding by combining slide images and OCR-extracted text as input for AI Teachers. Despite GPT-4's image comprehension, solely relying on it does not ensure high-quality scripts. Thus, we first accurately extract text using Tencent API OCR, a tool adept at recognizing various languages and detecting multiple languages in one image. This extracted text, combined with image input for the Language Model, enriches content understanding, aiding in the production of coherent instructional content. Instead of direct prompt engineering with ChatGPT, we simplify interaction for teachers through AI Teacher GPT, which guides users to input PPT images and OCR text in a specific format. The AI Teacher includes five design elements: role definition, task description, image processing, input format and output format. After input, it outputs the lecture text. This content used in this study has been manually validated by two experts for accuracy before proceeding to the next step.

The second step involves generating digital characters, where we leveraged SADtalker (Zhang et al., 2023) and Microsoft Edge's online text-to-speech service. We developed this generation program using Python, with a user-friendly interface. Input for this stage includes the lecture text from the previous step and a dynamic video or photo containing the instructor's image. The output is a complete lecture video with the teacher's image integrated. For input generation in this study, we opted for dynamic video over a static photo. The reason is that the former captures more natural facial expressions and blinking, including those above the mouth, while the latter lacks such micro-expressions.

The third step manually combines the teacher's video with PPT slides using video-editing software, placing the talking teacher at the bottom right of the slides. This finalizes the instructional video, typically within 5 minutes.

AI-generated and recorded instructional videos in this study

Two video lectures were created to teach English vocabulary, with eight words for each. In each video, an instructor on-screen explained the grammatical role of the words, offering explanations in both English and Chinese and using visuals to illustrate its meaning. Furthermore, the instructor provided a sample sentence to demonstrate how the word is used in context. For the RV condition, the video was produced using conventional recording and compositing methods, with an on-screen recorded human teacher. For the AIIV condition, the video was generated according to the instructional video generation pipeline, with an on-screen virtual teacher. The virtual teacher's image was generated from the image of the same human teacher in RV condition. The duration of AIIV is 7 minutes and 46 seconds and the RV lasts for 7 minutes and 58 seconds.

To screen English words that could be used in the formal experiment, a pilot test was conducted. We first selected 60 words from the GRE, TOEFL and IELTS to create a word dataset, which can be used to develop learning materials and tests for experiments. Then, we invited 24 students (*M* age = 21.92, SD age = 1.86; 21 females) who would not participate in the formal experiment to rate their familiarity with these words. We asked them to rate all 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 sampled eight words from the pool of 60 words, and these eight words with low familiarity (*M* = 1.73, SD = 0.30).

Dependent measures

Pre- and posttests of learning performance (retention and transfer)

We measured retention and transfer as indicators of learning performance. The retention test aimed to assess how well students remembered information stated explicitly in the instructional video. The transfer test evaluated students' ability to apply acquired knowledge in one context to a different context.

For the pretest, we assessed participants' familiarity with the words. They were required to fill in the Chinese meaning in the blanks for each word. 1 point was given for a correct answer; otherwise, 0 point was given. The total score of the retention test was 8, with a low score indicating a lower level of prior knowledge about the word to be learned. In the retention posttest, participants were asked to recall and write down the meaning of the words, identical to the pretest.

In the transfer test, participants completed three types of tasks: choosing the correct words, synonyms and antonyms. Twenty-four multiple-choice questions were included. The first task involved choosing the correct words, adapted from the fill-in-the-blank quizzes in Merriam-Webster's Vocabulary Builder. For instance, ‘Her visit in the spring was a kind of ___ to our relationship, which had really ended two months earlier’. The eight choices were as follows: A. cordial; B. peripheral; C. epilogue; D. nemesis; E. voluptuous; F. dystrophy; G. elucidate; and H. opaque. The second task required participants to choose synonyms or related words, and the third task asked them to choose antonyms or near antonyms. For transfer tests, 1 point was given for a correct answer, otherwise, 0 points were given. The total score was 24.

The McDonald's ω values (Hayes & Coutts, 2020; Shaw, 2021) were 0.77, 0.77 and 0.84 for the retention, transfer and overall performance tests, respectively, indicating a high level of internal consistency across all performance tests.

Subjective scales

In our study, subjective measures were employed to assess dimensions including cognitive, emotional, motivational, social aspects and satisfaction. These dimensions were impacted by the intervention, and the goal was to capture the immediate reactions in cognitive, emotional, motivational, social and overall factors due to the intervention. Thus, the subjective scales were administered after participants watched the AI-generated or recorded instructional videos and included a total of 47 questions. Table 2 provides an overview of the question items along with their corresponding perspectives, dimensions, types and references.

Cognitive load, motivation, social presence and satisfaction were assessed using a 5-point Likert scale (ranging from 1, indicating strongly disagree, to 5, indicating strongly agree). Trust and parasocial interaction utilized a 7-point Likert scale (ranging from 1, indicating strongly disagree, to 7, indicating strongly agree). Emotions were evaluated through

TABLE 2 Subjective ratings used in this study.

Perspective	Variable	Subdimensions	Numbers of questions	Type	Reliability (McDonald's omega)	Reference
Cognitive	Cognitive load	Mental load	5	5-point Likert	0.93	Hwang et al. (2013)
		Mental effort	3			
Emotional	Emotions	Valance	1	9-point manikins	0.77	Bradley and Lang (1994)
		Arousal	1			
		Control	1			
Motivational	Motivation		4	5-point Likert	0.85	Revised from (Katuk et al., 2013)
Social	Social presence		4	5-point Likert	0.88	Revised from (Makransky et al., 2017)
	Trust		5	7-point Likert	0.89	Revised from (Jian et al., 2000)
	Uncanny valley	Humanness	5	Semantic differential	0.94	Ho and MacDorman (2017)
		Attractiveness	4			
	Parasocial interaction		6	7-point Likert	0.97	Hartmann and Goldhoorn (2011)
Overall	Satisfaction		5	5-point Likert	0.95	Revised from (Hajhashemi et al., 2018)

9-point manikins, while the uncanny valley was rated using the semantic differential method. We maintained the same rating types as the original scales. Higher scores indicated a higher level of agreement, intensity or semantic relevance.

Cognitive load, emotions, the uncanny valley and parasocial interaction were directly derived from the references. Motivation, social presence, trust and satisfaction were revised, with adjustments made to align them with the instructional video contexts.

Procedure

Figure 2 illustrates the experimental procedure. The video lecture was delivered in a quiet and private room. Participants were first instructed about the experiment, signed the consent form and then completed demographic information (eg, age, gender and major) and retention pretests. The video lectures were displayed via a laptop. After watching the lectures, participants were asked to complete the posttests, fill out subjective scales and answer a few questions about their suggestions and opinions in a short interview. On average, participants spent approximately 25 minutes completing the experiment.

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 two conditions. Therefore, the *t*-test and the Mann–Whitney test were used for further analysis based on the normality of the data. We employed Cohen's *d* (Cohen, 2016) to measure the effect sizes in *t*-tests (small effect = 0.20; medium effect = 0.50; and large effect = 0.80). Like with *d*, the value of *r* represents the effect sizes in Mann–Whitney tests (small effect = 0.10; medium effect = 0.30; and large effect = 0.50).

RESULTS

Table 3 shows an overview of the descriptive statistics across two groups.

Learning performance

The pretest results showed that participants had low prior knowledge of the materials to be learned, and there was no significant difference in prior knowledge across the two conditions

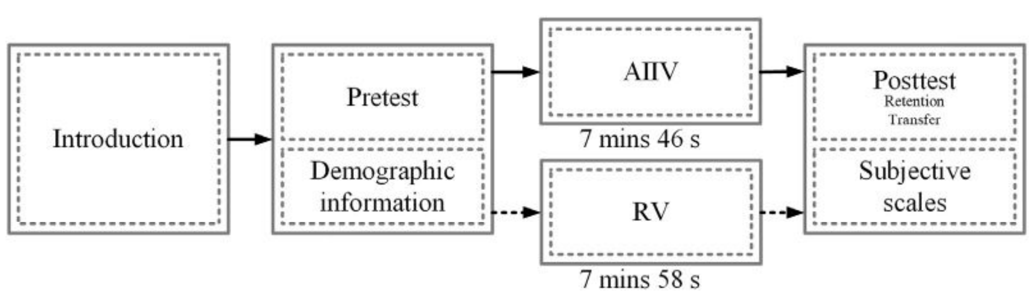

FIGURE 2 The experimental procedure.

TABLE 3 Descriptive statistics, p values and effect sizes across two conditions.

Dependent measures			AIIV ($n=38$; 16 males, 22 females) M (SD)	RV ($n=38$; 16 males, 22 females) M (SD)	p	d	r
Learning performance	Pretest		0.26 (0.95)	0.05 (0.32)	0.295		−0.12
	Posttest	Retention	7.50 (0.86)	6.53 (2.00)	0.038		−0.24
		Transfer	18.11 (3.49)	16.87 (4.36)	0.177	0.31	
	Overall performance		25.6 (4.07)	23.39 (6.08)	0.194		−0.15
Cognitive perspective	Cognitive load	Mental load	2.40 (0.81)	2.78 (0.97)	0.068	−0.42	
		Mental effort	2.19 (0.85)	2.49 (0.93)	0.149	−0.33	
Emotional perspective	Emotions	Valance	5.95 (1.79)	5.92 (1.75)	0.948	0.01	
		Arousal	5.05 (1.94)	5.11 (1.87)	0.905	−0.03	
		Control	5.58 (2.06)	5.58 (2.06)	1.000	0.00	
Motivational perspective	Motivation		3.00 (1.04)	3.34 (0.85)	0.119	−0.36	
Social perspective	Social presence		2.89 (1.07)	3.66 (0.97)	0.002	−0.75	
	Trust		4.16 (1.51)	4.76 (1.22)	0.058	−0.44	
	Uncanny valley	Humanness	2.82 (1.31)				
		Attractiveness	3.24 (1.04)				
	Parasocial interaction		2.55 (1.56)	2.82 (1.28)	0.416	−0.19	
Overall perspective	Satisfaction		3.05 (1.21)	3.51 (1.13)	0.094	−0.39	

($U = 683$, $z = -1.047$, $p = 0.295$, $r = -0.120$). This indicates that participants' familiarity with the knowledge across two conditions was the same.

The 38 participants who received AIIV intervention ($M = 7.50$, SD = 0.86) compared to the 38 participants in the control group ($M = 6.53$, SD = 2.00) demonstrated significantly better retention scores, $U = 543.5$, $z = -2.075$, $p = 0.038$, $r = -0.24$ (see Figure 3).

There was no significant difference in transfer scores between the participants receiving AIIV intervention ($M = 18.11$, SD = 3.49) and the control group ($M = 16.87$, SD = 4.36), $t(74) = 1.364$, $p = 0.177$, Cohen's $d = 0.31$.

There was no significant difference in overall performance between the AIIV condition ($M = 25.60$, SD = 4.07) and the RV condition ($M = 23.39$, SD = 6.08), $U = 597.5$, $z = -1.298$, $p = 0.194$, $r = -0.15$.

Together, these results supported H1.

Subjective scales

Figure 4 shows the differences in the subjective scales excluding the uncanny valley across two conditions.

Cognitive load

Cognitive load included two sub-dimensions: mental load and mental effort. The mental load scores of the control group ($M = 2.78$, SD = 0.97) were higher than those in the group receiving the AIIV intervention ($M = 2.40$, SD = 0.81) but not significant, $t(74) = -1.85$, $p = 0.068$, Cohen's $d = -0.42$ (see Figure 5). Mental effort scores were higher in the control group

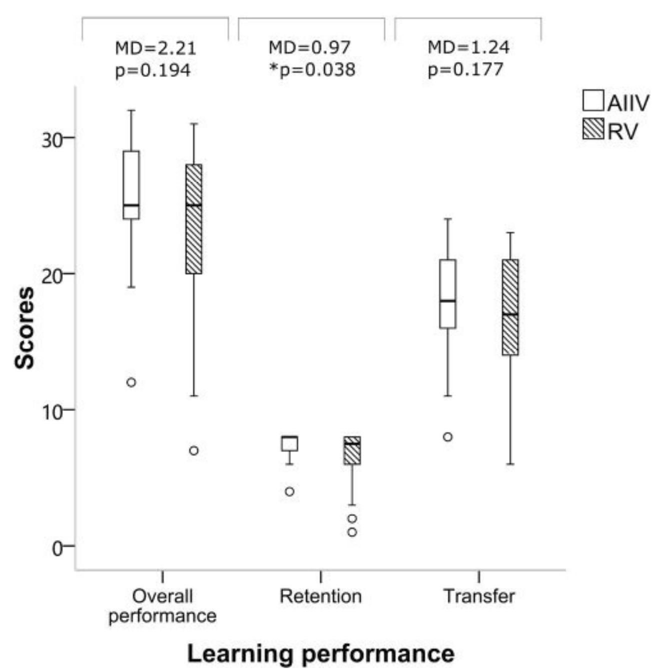

FIGURE 3 Differences in the retention, transfer and overall performance across two conditions.

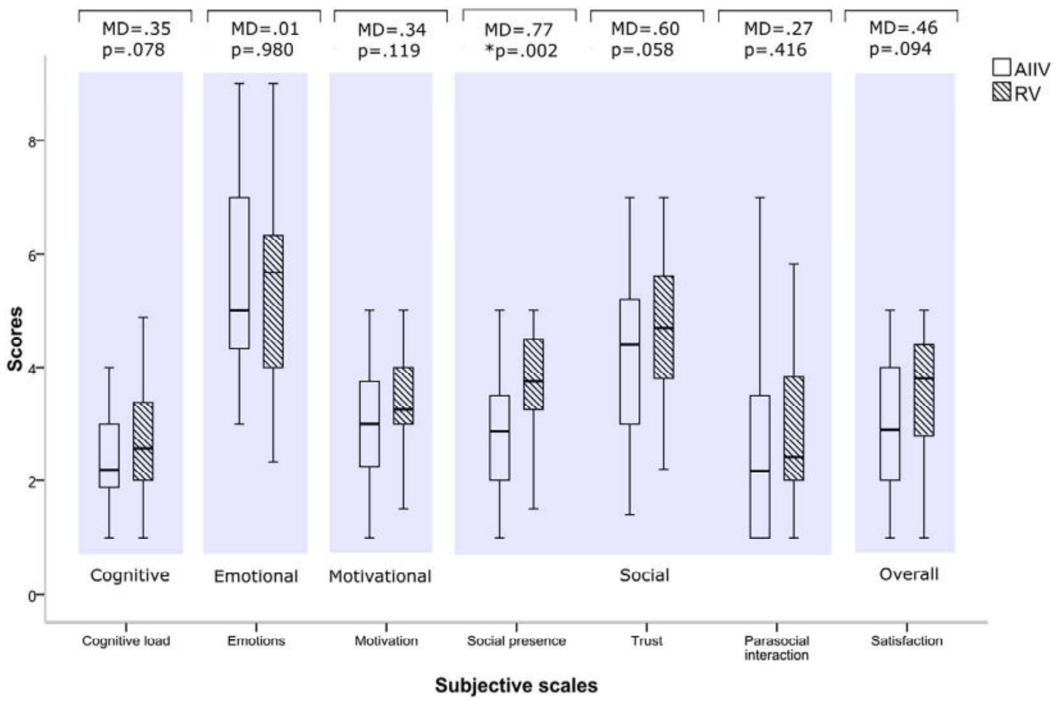

FIGURE 4 Differences in subjective ratings across two conditions.

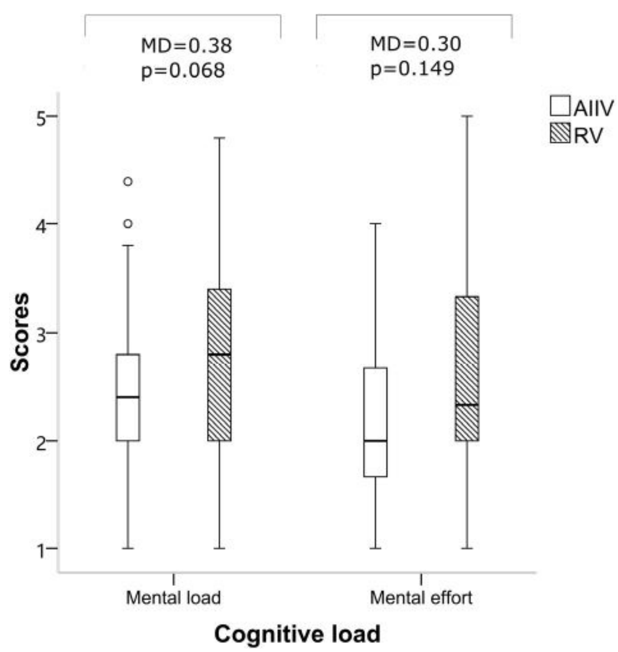

FIGURE 5 Differences in the cognitive load across two conditions.

(M=2.49, SD=0.93) than in the group receiving the AIIV intervention (M=2.19, SD=0.85) but not significant, $t(74)=-1.46$, $p=0.149$, Cohen's $d=-0.33$ (see Figure 5). These results were inconsistent with H2.

Emotions

The emotion scores of the control group ($M = 5.54$, SD = 1.64) were almost the same as the group receiving the AIIV intervention ($M = 5.53$, SD = 1.45), $t(74) = -0.03$, $p = 0.980$, Cohen's $d = -0.0057$. These results were inconsistent with H3.

Motivation

The participants in the RV group ($M = 3.34$, SD = 0.85) had slightly higher motivation scores than those in the AIIV intervention group ($M = 3.00$, SD = 1.04), but without a significant difference, $t(74) = -1.58$, $p = 0.119$, Cohen's $d = -0.36$ (see Figure 4). The result rejected H4.

Social presence, trust, uncanny valley and parasocial interaction

The social presence scores were significantly higher in the control group ($M = 3.66$, SD = 0.97) than in the group receiving the AIIV intervention ($M = 2.89$, SD = 1.07), $t(74) = -3.28$, $p = 0.002$, Cohen's $d = -0.75$ (see Figure 4).

The average humanness score of the AI teacher was 2.82 (SD = 1.31), and the attractiveness score was 3.24 (SD = 1.04). This indicated that the AI teacher was perceived as moderately to highly attractive, and the uncanny valley effect did not occur. The parasocial interaction scores in the control group ($M = 2.82$, SD = 1.28) were slightly higher than those in the AIIV intervention group ($M = 2.55$, SD = 1.56), but the difference was not statistically significant, $t(74) = -0.82$, $p = 0.416$, Cohen's $d = -0.19$.

These results partially supported H5 and fully supported H6.

Satisfaction

The participants in the control group ($M = 3.51$, SD = 1.13) had slightly higher satisfaction scores than those in the AIIV intervention group ($M = 3.05$, SD = 1.21), but without a significant difference, $t(74) = -1.70$, $p = 0.094$, Cohen's $d = -0.39$ (see Figure 4). The result supported H7.

DISCUSSION

Empirical contributions

This study explored the effectiveness of AI-generated instructional videos, encompassing three generation aspects: image, voice and lecture text, in improving learning outcomes, as well as how learners perceive these contents. The results indicated significantly higher retention in the AIIV condition compared to the RV condition. However, no significant differences were observed in transfer scores between the two groups of students, suggesting that while AI-generated videos can boost retention, their impact on the ability to apply learned knowledge may be the same as that of recorded videos. Taken together, AI-generated instructional videos can facilitate learning as traditional recorded videos. This could be explained by the standardized nature of generated resources, where both speech and lecture scripts adhere to high standards. The study focused on the learning of English words, involving pronunciation. Compared to human counterparts, modern machine voices are becoming increasingly realistic (Kaur & Singh, 2023), offering superior clarity in articulation and providing a consistent standard accent. This consistency is crucial in language learning, where the precision of pronunciation can significantly influence comprehension. While human teachers undergo voice training, variations in accents can still occur. Additionally, generated scripts are highly standardized and patterned, featuring smoother transitions and a better logical flow between sentences (Chaka, 2023; Elkhatat et al., 2023). In contrast, human languages may exhibit variations in style and inconsistencies in word usage, which might confuse learners or decrease their learning experience. Therefore, in the context of second language learning, AI-generated videos can facilitate learning the same as human-recorded videos, potentially triggering superior learning outcomes due to increased standardization. However, the generalization of these results to other disciplines such as science and engineering remains uncertain. The standardization effect caused by AI for language learning may not directly transfer to disciplines, such as science and engineering, which do not rely heavily on pronunciation and accent. Additionally, English vocabulary instructional videos do not involve much logical reasoning. Generating lecture scripts for them does not require extensive logical generation. In contrast, science and engineering disciplines require scripts that explain logical reasoning and problem-solving processes. Therefore, the effects may vary across different disciplines. While AI-generated videos show promise in language learning, further research is needed to explore their effectiveness in other disciplines.

Through collecting participants' subjective ratings, we found that scores on dimensions such as satisfaction, motivation, trust, cognitive load, emotions and parasocial interaction for generated videos showed no significant differences compared to recorded videos, but were consistently lower across all these dimensions. This result aligns with some findings from Chiou et al. (2020), although we obtained results in more dimensions. Our findings suggest that despite advancements in AI technology allowing generated videos to enhance learning in terms of voice, communication and appearance, scores on these dimensions still fall below those produced videos with real instructors talking. This study further confirms previous research indicating that humanoid entities are perceived as distinct categories from real human entities (Gong & Nass, 2007).

Individuals in the AIIV group reported lower social presence when interacting with virtual resources, but simultaneously experienced lower cognitive load. The heightened social presence perceived by participants in the RV group was compensated by an increased cognitive load. This finding indicates that real human videos are not the absolute ideal teaching resource. Conversely, generated instructional videos enable learners to focus more on the learning process and still create an environment of companionship by interacting with humanoid characters. The humanness (low) and attractiveness (high) scores in the uncanny valley dimension further emphasized that the presence of virtual beings not only alleviated potential social pressure from real instructors but also provided learners with a sense of genuine companionship. Taken together, this study addresses a gap in existing research, providing a clearer and extensive understanding of learners' subjective opinions on generated instructional video composing generated character, voice and lecture text.

Theoretical contributions

This study has achieved a significant theoretical breakthrough by further extending the applicability of the equivalence principle (Horovitz & Mayer, 2021) into AIGC. Previous research has continuously validated this principle in the application of virtual instructors (eg, Deng et al., 2022; Pi et al., 2022). To go a step further, our study has made advancements, demonstrating that instructional videos generated by AI also adhere to the equivalence principle. Under certain boundary conditions, such as different learning styles or knowledge types, generated videos may outperform recorded videos. This finding not only enriches the scope of the equivalence principle but also provides theoretical support for the practical application of AIGC in the field of education.

Another significant theoretical implication of this study is the evolution of the teacher's role. The findings further confirm that with the advancement of AI technology, the stages involved in producing instructional videos are gradually becoming automatable, allowing more and more parts of instructional videos to be generated. Previously, teachers needed to actively participate in the production of resources, but now they can allocate more time to instructional design. This shift directly results in a transformation of the teacher's role. The idea that the development of AI technology is reshaping the role of teachers and placing additional demands on their capabilities is consistent with existing literature (Chiu et al., 2023).

Practical implications

First, this study serves as an illustrative example of using AIGC for educational purposes. AIGC enables the scalable replication of human intelligence in machines, allowing for the widespread replacement of basic professional tasks. Previously, instructional video platforms depended on manually recorded content, a process both time-consuming and labour intensive. AIGC technology has transformed this process by enabling the rapid production of customized short learning videos, with human efforts mainly focused on expert review. This innovation facilitates the swift conversion of multimedia resources into video content, aligning with the modern learner's preference for microlearning. Therefore, this study opens up new perspectives for the creation of educational AIGC online platforms, potentially representing an innovation in future learning models. Furthermore, AI-generated videos have potential beyond EFL, with applications in science and engineering for complex topics like chemical reactions or mathematical reasoning. However, their effectiveness may vary across disciplines, requiring further research to understand their impact in different contexts, as previously noted.

Second, our research findings encourage instructors to consider using AIGC in both physical and online classes. Video lectures have gained popularity as a learning medium (Van der Meij, 2017). Yet, producing recorded videos is costly and burdensome for instructors. Our study indicates that generated video resources can be as effective as traditional recorded videos in enhancing learning. Given the simpler production process of generated resources, we recommend that educators consider incorporating them to facilitate their teaching. This approach can not only alleviate the burden and costs for teachers but also be applied in intelligent tutoring or adaptive systems, minimizing the effort and resources required for creating educational content. However, it should be noted that using generated resources is meant to complement, not completely replace, traditional methods (Pataranutaporn et al., 2021). To effectively integrate AI-generated instructional videos into their teaching practices, we recommend that educators begin by using AI-generated videos as supplementary materials to illustrate key concepts taught in traditional classrooms. Additionally, it is crucial for educators to undergo training to become proficient with AI tools, enabling them to tailor content to their specific needs. Teachers can also use the full pipeline to create videos with a talking teacher, as required. For more practical use, they can generate lecture audios (using text-to-speech) and combine these with slides to produce videos that do not feature a talking teacher on screen. Furthermore, future research should investigate a broader range of AI-generated video types. The design featuring a video narrator on screen accompanied by slides is not the sole type of instructional videos. Traditional instructional videos encompass various types (Crook & Schofield, 2017; Köse et al., 2021). In this study, we explored only one type of

AI-generated video. Besides this B3 type, our pipeline can partially or fully generate other types of videos, such as A1 Voice over slides, A2 Voice over screencast, A3 Writing over slides, D2 Presence in lecture and D3 Presence in full screen (Crook & Schofield, 2017). A1, A2 and A3 involve a hidden voice narrating content, with A1 and A2 presenting slides and screen recordings, respectively, and A3 adding the narrator's writing or annotations over the slides. D2 and D3 both involve a narrator, with D2 showing the narrator in a traditional lecture setting and D3 presenting a close-up of the narrator. However, certain traditional recorded videos, like A4 Kahn whiteboard videos, remain challenging to produce with current generative technologies. The objective of generated videos is not to replicate traditional recorded videos entirely but to surpass traditional recorded videos in many aspects. Therefore, while the theoretical framework and empirical results may be shared with traditional recorded videos in certain aspects, their impact and boundary conditions should also include unique considerations.

Third, attention should be given to the ethical and responsible use of these technologies. The process of generating instructional videos often involves the use of driven videos that contain personal facial and voice data. Therefore, it is important to ensure the privacy and security of voice and image data used in video generation. This includes implementing robust data protection measures such as access controls and compliance with data protection regulations. Educators and developers should also consider obtaining explicit consent from individuals whose data are used in generating these videos. Additionally, efforts should be made to prevent hallucination produced by AI (Weidinger et al., 2022; Zuccon et al., 2023). Hallucination refers to the risk of AIGC being inaccurate, confusing or misleading (Ji et al., 2023), which raises concerns about inaccurate information, including factual errors, biased viewpoints, logical inconsistencies, erroneous reasoning and potentially manipulative content. These inaccuracies can not only mislead learners but also negatively impact their knowledge base and value systems. A possible solution is to involve human experts in content review to ensure the quality of videos.

Finally, we believe that generated video resources can help reduce the digital divide caused by the application of AI, making education more equitable. The application of GAI requires certain infrastructure, devices and Internet, which may result in unequal access and a lack of resources for education in some regions and communities (Capraro et al., 2023). While GAI may be challenging to distribute widely, generated resources can be rapidly produced in large quantities and distributed to different regions, ensuring that teachers in remote areas can also benefit.

CONCLUSION, LIMITATIONS AND FUTURE DIRECTIONS

Rapid advancements in AI technology are enabling the generation of instructional content. This development offers educational stakeholders the opportunity to move away from the intensive labour of video production and focus more on instructional design. This study is among the first to examine how students use generated instructional videos and thoroughly explore their reactions to such content. We found that the generated instructional videos could facilitate learning as effectively as recorded videos, and in the context of English language learning, they even triggered higher retention scores. Furthermore, AIIV was found to be moderately to highly attractive, addressing concerns related to the uncanny valley effect. These findings extend our understanding of incorporating generated videos into education. To the best of our knowledge, this is the first work on the design, implementation and evaluation of generated instructional videos, offering theoretical insights and practical implications for the application of AIGC in education.

However, it is important to note that the present study has a few limitations that should be addressed in future investigations. First, we did not examine the effects of AI-generated lecture text, voice and teacher's image on learning separately. Future research will consider the influence of each factor and their interaction. Second, this study is a short-term experiment and the results may be impacted by the novelty effect. Future work will explore long-term implications. Third, the learning resources utilized in the current research were centred around second language vocabulary acquisition. However, complex conceptual learning and skill acquisition have not yet been explored. Future research should therefore shift from second language acquisition to science and engineering, placing greater attention on complex learning and exploring variations in the effects. Fourth, most data in the present study were self-reported, which did not reveal the dynamic engagement level over time and the attention participants paid to the teacher and learning material areas. Therefore, in future research, we intend to gather data via electroencephalography (EEG) and eye-tracking technology to better understand these internal cognitive processes. Furthermore, given that learners are central to the learning process, their individual differences would significantly impact the effectiveness of generated videos. Therefore, it is crucial to thoroughly investigate how learners' varying levels of AI literacy, as well as their acceptance and trust towards AI technology impact learning. Also, the design of learning strategies, such as verbal and nonverbal emotional cues, should be studied to establish the principles for AI-generated video design. These principles could then guide the use of generated videos in adaptive systems that cater to individual needs and preferences. Lastly, from the perspective of AI technology, the current pipeline is not fully automated. In the future, we aim to achieve complete automation in the pipeline, replacing more components of instructional video production with AI. Teachers would only need to provide the knowledge to be learned and learning objectives, and AI could generate the instructional video resources automatically. We will further evaluate these fully autogenerated instructional resources extensively.

FUNDING INFORMATION

This work was supported by the National Key R&D Program of China under Grant [2022YFC3303600] and the Research Projects of the National Natural Science Foundation of China under Grants [62077036] and [62377039]; and sponsored by CAAI-CANN Open Fund, developed on OpenI Community.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data are available from the corresponding author upon request.

ETHICS STATEMENT

This research was approved by the Ethics Committee of the Faculty of Education at Shaanxi Normal University (No. 2023-001). The study followed the ethical guidelines of the institutional and national research committee, as well as WMA Declaration of Helsinki. All participants were informed of the voluntary nature of their participation and asked to provide written informed consent. They were assured of anonymity and informed about the purpose of the research, how their data would be used and that no risks were associated with their involvement.

CONSENT TO PARTICIPATE

All participants signed informed consent agreement.

ORCID

Tao Xu https://orcid.org/0000-0002-1721-561X
Yun Zhou https://orcid.org/0000-0002-2306-8986

REFERENCES

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4337484

BavaHarji, M., Kamal Alavi, Z., & Letchumanan, K. (2014). Captioned instructional video: Effects on content comprehension, vocabulary acquisition and language proficiency. *English Language Teaching*, *7*(5), p1. https://doi.org/10.5539/elt.v7n5p1

Beege, M., Schroeder, N. L., Heidig, S., Rey, G. D., & Schneider, S. (2023). The instructor presence effect and its moderators in instructional video: A series of meta-analyses. *Educational Research Review*, *41*, 100564. https://doi.org/10.1016/j.edurev.2023.100564

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A. S., Creel, K. A., Davis, J., Demszky, D., … Liang, P. (2021). On the opportunities and risks of foundation models. *ArXiv*. https://www.semanticscholar.org/paper/On-the-Opportunities-and-Risks-of-Foundation-Models-Bommasani-Hudson/4f68e07c6c3173480053fd52391851d6f80d651b

Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. *The Internet and Higher Education*, *15*(3), 195–203. https://doi.org/10.1016/j.iheduc.2011.11.001

Bozkurt, A., Karadeniz, A., & Baneres, D. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability*, *13*(2), 1–6. https://doi.org/10.3390/su13020800

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023, March 7). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv.Org*. https://arxiv.org/abs/2303.04226v1

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Branas-Garza, P., Butera, L., Douglas, K., Everett, J., Gigerenzer, G., Greenhow, C., Hashimoto, D., Holt-Lunstad, J., Jetten, J., Johnson, S., Longoni, C., Lunn, P., & Viale, R. (2023). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus*, *3*(6), 1–18. https://doi.org/10.31234/osf.io/6fd2y

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*, *27*, 989–1015. https://doi.org/10.1007/s10648-020-09587-1

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching*, *6*(2), Article 2. https://doi.org/10.37074/jalt.2023.6.2.12

Chen, C.-M., Li, M.-C., & Lin, M.-F. (2020). The effects of video-annotated learning and reviewing system with vocabulary learning mechanism on English listening comprehension and technology acceptance. *Computer Assisted Language Learning*, *35*, 1–37. https://doi.org/10.1080/09588221.2020.1825093

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

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

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

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research and Development*, *66*(6), 1415–1433. https://doi.org/10.1007/s11423-018-9594-x

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.

Dao, X.-Q., Le, N.-B., & Nguyen, T.-M.-T. (2021). AI-powered MOOCs: Video lecture generation. In *3rd International Conference on Image, Video and Signal Processing*. Singapore Singapore: (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*, 103605. https://doi.org/10.1016/j.compedu.2019.103605

Deng, L., Zhou, Y., Cheng, T., Liu, X., Xu, T., & Wang, X. (2022). In P. Zaphiris & A. Ioannou (Eds.), *My English teachers are not human but I like them: Research on virtual teacher self-study learning system in K12* (Vol. *13329*, pp. 176–187). Springer International Publishing. https://doi.org/10.1007/978-3-031-05675-8_14

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2018). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior*, *90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*(1), 17. https://doi.org/10.1007/s40979-023-00140-5

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *ArXiv*. https://www.semanticscholar.org/paper/GPTs-are-GPTs%3A-An-Early-Look-at-the-Labor-Market-of-Eloundou-Manning/5501d00310b06e00351295529498cc684187148d

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

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research*, *33*(2), 163–193. https://doi.org/10.1111/j.1468-2958.2007.00295.x

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*, 8435–8463.

Hajhashemi, K., Caltabiano, N., Anderson, N., & Tabibzadeh, S. A. (2018). Multiple intelligences, motivations and learning experience regarding video-assisted subjects in a rural university. *International Journal of Instruction*, *11*, 167–182.

Hartmann, T., & Goldhoorn, C. (2011). Horton and Wohl revisited: Exploring viewers' experience of parasocial interaction. *Journal of Communication*, *61*, 1104–1121. https://doi.org/10.1111/j.1460-2466.2011.01595.x

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, *9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9

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

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Hwang, G.-J., Yang, L.-H., & Wang, S.-Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, *69*, 121–130. https://doi.org/10.1016/j.compedu.2013.07.008

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Jeon, J., Lee, S., & Choi, S. (2023). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*, *32*, 4613–4631. https://doi.org/10.1080/10494820.2023.2204343

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

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

Kim, J., Merrill, K., Jr., 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

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., & Welch, G. F. (2018). Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Munich Germany: (pp. 105–114). https://doi.org/10.1109/ISMAR.2018.00039

Kizilcec, R., Bailenson, J., & Gomez, C. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*, 724–739. https://doi.org/10.1037/edu0000013

Köse, E., Taşlıbeyaz, E., & Karaman, S. (2021). Classification of instructional videos. *Technology, Knowledge and Learning*, *26*(4), 1079–1109.

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

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, *29*, 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

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. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky*. Tokyo Japan: (pp. 523–529). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_81

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 Science*, *121*(3/4), 117–135.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*, 410. https://doi.org/10.3390/educsci13040410

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior*, *28*(1), 181–186. https://doi.org/10.1016/j.chb.2011.08.025

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Martha, D., & Santoso, H. (2019). The design and impact of the pedagogical agent: A systematic literature review. *The Journal of Educators Online*, *16*, n1. https://doi.org/10.9743/jeo.2019.16.1.8

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. *Information Design Journal*, *16*(1), 81–83. https://doi.org/10.1017/cbo9781139547369.017

Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, *29*, 3195–3217.

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

Ooi, K.-B., Tan, G., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y., Huang, T.-L., Kar, A., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N., Sarker, P., Sharma, A., & Fosso Wamba, S. (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

Park, B., Plass, J. L., & Brünken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction*, *29*, 125–127.

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

Plass, J. L., Mayer, R. E., & Homer, B. D. (Eds.). (2020). *Handbook of game-based learning* (pp. 12–13). MIT Press.
Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783
Schroeder, N. L., Chiou, E. K., & Craig, S. D. (2021). Trust influences perceptions of virtual humans, but not necessarily learning. *Computers & Education*, *160*, 104039. https://doi.org/10.1016/j.compedu.2020.104039
Shaw, B. P. (2021). Meeting assumptions in the estimation of reliability. *The Stata Journal: Promoting Communications on Statistics and Stata*, *21*(4), 1021–1027. https://doi.org/10.1177/1536867X211063407
Shen, W.-W., Lin, J.-M., Cheng, W. K., & Hong, Z.-W. (2021). Developing and evaluating an online video-assisted collocation learning system for EFL students. *Interactive Learning Environments*, *31*, 1–15. https://doi.org/10.1080/10494820.2021.2003408
Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., … Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2
St J. Neill, S. R. (1989). The effects of facial expression and posture on children's reported responses to teacher nonverbal communication. *British Educational Research Journal*, *15*(2), 195–204.
Thomas, R. A., West, R. E., & Borup, J. (2017). An analysis of instructor social presence in online text and asynchronous video feedback comments. *The Internet and Higher Education*, *33*, 61–73. https://doi.org/10.1016/j.iheduc.2017.01.003
Van Der Meij, H. (2017). Reviews in instructional video. *Computers & Education*, *114*, 164–174. https://doi.org/10.1016/j.compedu.2017.07.002
Veletsianos, G., Miller, C., & Doering, A. (2009). EnALI: A research and design framework for virtual characters and pedagogical agents. *Journal of Educational Computing Research*, *41*(2), 171–194. https://doi.org/10.2190/EC.41.2.c
Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049
Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022, June). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Seoul Republic of Korea: (pp. 214–229).
Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education*, *122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011
Xu, T., Wang, X., Wang, J., & Zhou, Y. (2021). From textbook to teacher: An adaptive intelligent tutoring system based on BCI. In *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. virtually: (pp. 7621–7624). https://doi.org/10.1109/EMBC46164.2021.9629483
Young, J. C., & Shishido, M. (2023). Investigating OpenAI's ChatGPT potentials in generating Chatbot's dialogue for English as a foreign language learning. *International Journal of Advanced Computer Science and Applications*, *14*(6), 65–72.
Zhang, W., Cun, X., Wang, X., Zhang, Y., Shen, X., Guo, Y., Shan, Y., & Wang, F. (2023, Sadtalker: Learning realistic 3d motion coefficients for stylized audio-driven single image talking face animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, BC, Canada: (pp. 8652–8661).
Zuccon, G., Koopman, B., & Shaik, R. (2023, November). Chatgpt hallucinates when attributing answers. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*. Beijing China: (pp. 46–51).

How to cite this article: Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2025). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*, *56*, 1463–1487. https://doi.org/10.1111/bjet.13530