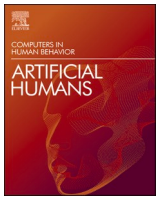
Computers in Human Behavior: Artificial Humans 3 (2025) 100117



Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans





Learning through AI-clones: Enhancing self-perception and presentation performance

Qingxiao Zheng [a,*], Zhuoer Chen [b], Yun Huang [c]

[a] Institute for Artificial Intelligence and Data Science (IAD), Department of Computer Science and Engineering, University at Buffalo, SUNY, NY, USA
[b] HeyGen, 12130 Millennium Dr. Ste 300, Los Angeles, CA, 90094, USA
[c] School of Information Sciences, University of Illinois at Urbana Champaign, 614 E Daniel St, Champaign, IL, 61820, USA

ARTICLE INFO

Keywords:
Social comparison theory
Regulatory-focus
AI avatar
Deepfake
Voice clone
Experiment
Self-kindness

ABSTRACT

This study examines the impact of AI-generated digital clones with self-images (AI-clones) on enhancing perceptions and skills in online presentations. A mixed-design experiment with 44 international students compared self-recording videos (self-recording group) to AI-clone videos (AI-clone group) for online English presentation practice. AI-clone videos were generated using voice cloning, face swapping, lip-syncing, and body-language simulation, refining the repetition, filler words, and pronunciation of participants' original presentations. The results, viewed through the lens of social comparison theory, showed that AI clones functioned as positive "role models" for encouraging positive social comparisons. Regarding self-perceptions, speech qualities, and self-kindness, the self-recording group showed an increase in pronunciation satisfaction. However, the AI-clone group exhibited greater self-kindness, a wider scope of self-observation, and a meaningful transition from a corrective to an enhancive approach in self-critique. Moreover, machine-rated scores revealed immediate performance gains only within the AI-clone group. Considering individual differences, aligning interventions with participants' regulatory focus significantly enhanced their learning experience. These findings highlight the theoretical, practical, and ethical implications of AI clones in supporting emotional and cognitive skill development.

1. Introduction

A large body of work in the social sciences has investigated the influence of celebrity role models, and in the context of education, several disciplines have a rich research history in this area, e.g. medical education (Jo et al., 2018). However, in the context of learning, research on role models has largely remained on the periphery (Muir et al., 2021). Foreign language speakers, especially international students, frequently struggle with unique challenges in online presentation delivery, facing language barriers and difficulties in conveying nuanced messages to global audiences online (Alene, 2013; Du, 2015; Reinhardt, 2019; Szyszka, 2017). Existing technologies, such as in situ and delayed feedback systems, although beneficial in providing feedback, can sometimes overwhelm or inadequately serve foreign language speakers to address their language and cultural challenges (Alene, 2013; Wang et al., 2020; Zhou et al., 2020). Moreover, these solutions may reinforce "negative performance," posing risks for individuals dealing with public speaking anxiety or self-esteem issues (Karl & Kopf, 1994; Zhou et al., 2020). This situation signifies an urgent need for better feedback systems specifically designed to meet the challenges of international speakers. In this study, through the lens of *social comparison theory*, we examine how AI can support foreign language speakers in practicing online presentations.

Over 70 years ago, Festinger (Leon, 1954) introduced the concept of social comparison theory, positing that individuals have an innate drive to evaluate their opinions and abilities for accurate self-assessment. Festinger noted that while people ideally use objective standards for self-evaluation, they often compare themselves to others in the absence of such information. This theory has been extensively applied in education and learning, with numerous studies investigating its effects on learning outcomes (Dijkstra, Kuyper, Van der Werf, Buunk, & van der Zee, 2008; Hanus & Fox, 2015; Leon, 1954; Micari & Drane, 2011). Among the facets explored is the *"near-peer role model"* (Kang & Liu, 2019; Michinov & Michinov, 2001; Murphey, 1998), which suggests that comparisons are influenced by similarities in gender, race, experience, etc (Julius et al., 1990; Luong et al., 2020; Wolff et al., 2010),

* Corresponding author.
E-mail addresses: qingxiao@buffalo.edu (Q. Zheng), joy@heygen.com (Z. Chen), yunhuang@illinois.edu (Y. Huang).

https://doi.org/10.1016/j.chbah.2025.100117
Received 6 March 2024; Received in revised form 3 January 2025; Accepted 6 January 2025
Available online 3 February 2025
2949-8821/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

indicating a tendency for individuals to compare themselves with those who share similar characteristics. This body of work highlights the complexity of social comparisons and their nuanced impact on personal development and learning.

Many learning tools have been designed and created to support speech and language practice in online environments, offering mechanisms for learners to compare their performance with their peers (Guerra et al., 2016; Tabuenca et al., 2015), or with students from previous sessions (Davis et al., 2016; Pérez-et al., 2017). However, this method has led to a scarcity of accessible role models, particularly for international or non-native English speakers struggling to find relatable role models because of the lack of representation or connections to others with similar backgrounds (Latu et al., 2013; Lockwood, 2006). Recent advancements in AI technology have made AI-clones increasingly capable, with capabilities ranging from text-based chatbots to immersive virtual reality experiences (Freeman & Maloney, 2021; Lee et al., 2023; Neelkumar et al., 2019). While these clones present a host of ethical and psychological considerations (Hogan, 2010; Mori et al., 2012; Stutzman & Hartzog, 2012; Zhao et al., 2013), their potential benefits, particularly in educational contexts, remain underexplored. One such potential could be the creation of role models *with self-images*, or highly personalized AI-clones, which may support skill development, especially in addressing the distinct challenges faced by international speakers.

In this paper, we propose a novel approach leveraging AI-generated videos to create role models for international speakers practicing online presentations. The proposed intervention allows users to observe their "improved" (potentially ideal) self in AI-clones, thereby providing an aspirational target and promoting self-observation from an audience's perspective. The study journey is structured in several key phases: initial self-concept mapping, a first presentation followed by self-assessment, video-stimulated recall for performance review, a second presentation attempt, and a final exit evaluation. Employing a randomized design, we separated participants into a self-recording (control) group or an AI-clone (experiment) group. We ask, **RQ1**: *How do learning outcomes differ between individuals watching self-recorded* versus *AI self-clone videos?* The AI-clone group was further divided into two types of regulatory focus: promotion-focused and prevention-focused. The promotion focus centers on the pursuit of goals and achievements, while the prevention focus emphasizes the avoidance of setbacks and errors (Tory Higgins et al., 2001). We ask, **RQ2**: *How does regulatory focus influence learning outcomes when using AI self-clone videos?*

To evaluate, we adopt a multi-dimensional approach to examine how self-cloned video feedback can improve presentation efficacy, within the broader context of skill improvements. It involves three stages–*Think*, *Feel*, and *Act*– to investigate the influence of the AI self-clones on speakers' online presentation practice practice. This approach is grounded in the concept of self-regulation, which describes the interplay between the environment's impact on an individual and their consequent behavior, mediated through processes of *self-observation*, *self-judgment (or evaluation)*, and *self-reaction* (Albert Bandura, 1986; Lin et al., 2022).

This study contributes to the existing body of research on AI-supported online presentation delivery in three ways. First, we demonstrate that AI-generated self-clone videos can function as "role models," in line with the social comparison and regulatory focus theory, without the need to explicitly prime a role model as seen in previous studies (Latu et al., 2013; Lockwood, 2006; Lockwood et al., 2002; Zhang, 2016). Second, our research delves into the effects of generative AI self-clones on learners' self-observation, perceptions of their speech and learning experiences, and their performance. Participants using the AI demonstrated enhancements particularly in the emotional aspects (e.g., showing greater kindness to self) and self-perception (e.g., perceived higher improvements in smiles and communication) within a presentation practice learning context (Orii et al., 2022; Pataranutaporn et al., 2022). Our findings indicate that AI's influence varies on learners' regulatory focus. Promotion-focused individuals experienced increased pronunciation confidence, reduced speech anxiety, and found AI more enjoyable and useful, whereas prevention-focused individuals did not experience the same benefits. While both groups improved in reducing filler words, only the prevention-focused group also showed enhancements in managing repetition and weak words. These insights open new avenues for applying AI into online speech and language learning.

2. Related work

2.1. Urgent needs and approaches for practicing online self-presentation

The digital age, marked by the ubiquity of social media platforms like Facebook, TikTok, RedNote, Instagram, YouTube, and LinkedIn, has underscored the importance of effective online self-presentation, or in the perpetual act of curating their online persona (Ann, 2020; Jacobson, 2020). This has become particularly crucial for enhancing social and professional opportunities (Li et al., 2021; Ngai, 2019). The widespread adoption of remote work practices and platforms such as Zoom, Teams, and Webex has made online self-presentation a necessary skill for social interaction and work proficiency (Gunawan et al., 2021). However, international students face unique challenges in online self-presentation due to non-native speech (Alene, 2013), including but not limited to, language barriers such as accents and pronunciation (Szpyra-Kozłowska & Szpyra, 2014; Szyszka, 2017), vocabulary limitations (Hazenberg & Hulstun, 1996; Roberts et al., 2005), grammar difficulties(Du, 2015), and the capability to convey nuanced messages that resonate with a global audiences (Du, 2015; Reinhardt, 2019).

2.1.1. Limitations of current feedback systems

Existing feedback systems significantly advanced the domain of public speaking and online presentation training, and they mainly fall into three categories: in-situ, delayed, and self-corrective feedback systems.

In-situ feedback systems provide real-time cues that improve speakers' awareness of nonverbal communication and pacing. For instance, these systems can deliver live feedback on body openness, energy, and speech rate to help users improve their physical expressiveness (Damian et al., 2015). They can also assist speakers by providing immediate prompts, such as showing text cues like *"louder"* or *"faster"* in AR environments, enabling on-the-fly adjustments to volume and speaking rate (Barmaki & Hughes, 2018; Iftekhar et al., 2015). Some systems also display script coverage to aid in oral presentations (Asadi et al., 2017). However, real-time feedback can overwhelm international speakers who may already face challenges related to language and cultural nuances. Moreover, these systems often lack tools to facilitate structured deliberate practice or enable performance comparisons (Wang et al., 2020).

Delayed feedback systems usually offer quantitative metrics to guide future improvements (Wang et al., 2020). These include assessment, analytical, or tagged "videotape playback" tools such as dashboards that visualize performance metrics or physiological responses like electrodermal activity (e.g., heart rate) linked to stress and anxiety (Croft et al., 2004; Giraud et al., 2013). Automatic performance scoring systems assess speech quality using vocal characteristics like pitch, speech rate, and loudness (Wang et al., 2020). While such systems support novice users in reflecting on their performance (Zhou et al., 2020), they often rely on predefined, context-insensitive metrics. This approach can be particularly limiting for international speakers who may be unfamiliar with the rhetorical or cultural norms of the presentation language (Wang et al., 2020; Zhou et al., 2020).

Self-corrective feedback systems, such as auditory feedback mechanisms, provide immediate remedies for presentation issues, and also widely incorporate playback as a mechanism. For example, computer-aided personalized pronunciation training with exaggerated corrective feedback improves speakers' awareness (Bu et al., 2021). Altered

auditory feedback systems resynthesize voice inputs with modifications to create more confident-sounding voices, helping speakers release tension and manage emotions (Naruse et al., 2019; Orii et al., 2022). Despite their effectiveness for short-term error correction, intermittent feedback has been shown to have limited impact on long-term retention (Schmidt et al., 1989; Wang et al., 2020).

A major drawback of the above feedback systems is their tendency to highlight performance deficiencies, also referred to as "negative performance" (Karl & Kopf, 1994). For example, in-situ video feedback systems have been reported to distract speakers by drawing attention on their imperfections (Dermody and Sutherland, 2019). Similarly, video or audio playback in delayed feedback systems and self-corrective feedback systems frequently exacerbates public speaking anxiety (Myles et al., 1970). Individuals with public speaking anxiety, who might benefit the most from playback feedback, often avoid it due to self-esteem and self-efficacy concerns (Karl & Kopf, 1994; Zhou et al., 2020). These issues can be particularly severe for international speakers, who may interpret negative feedback as criticism of their linguistic or cultural proficiency. To overcome this limitation, it is crucial to develop feedback mechanisms that prioritize positive reinforcement for international language learners.

2.1.2. Challenges of learning from role models

Role modeling is widely recognized as an effective mechanism for improving self-presentation skills, allowing learners to observe and internalize strategies for success (Orii et al., 2022). Observing peers, lecturers, or exemplary speech performances has been shown to enhance self-presentation efficacy and encourage informative communication (Dale, 1987; Sonnenschein & Whitehurst, 1980). To leverage the role modeling mechanism, prior work has primarily relied on "*priming*" to set role models encouraging learners to picture a role model who is distinct from oneself, yet shares similarities, to achieve their impact in many of the learning contexts (Lockwood, 2006; Lockwood et al., 2002; Zhang, 2016). However, there is a lack of access to "real human role models" who can be available on deman.

Research studies also indicate that role models who share attributes with learners are particularly effective in enhancing learning outcomes (Latu et al., 2013; Lockwood, 2006). These near-peer role models are defined as "individuals who are "near" to oneself in several ways, such as age, ethnicity, sex, interests, past or present experiences (Murphey, 1998, pp. 201–205), a concept often referenced in second language learning context. For example, prior research highlights the impact of near-peer role models on learners' perceptions of speech performance (Latu et al., 2013; Lockwood, 2006). Similarly, gender plays a significant role in the effectiveness of role models, with a study reporting that women achieved better self-evaluations and presentation outcomes when observing female role models with similar attributes(Latu et al., 2013).

For international speakers or non-native English speakers, however, finding role models who share similar attributes can be challenging, particularly in foreign environments where the majority of their connections and interactions are with native speakers. This lack of relatable role models can hinder their ability to fully engage with or imitate the role models, thereby reducing the effectiveness of the priming approach (Orii et al., 2022). Despite its promise, this approach has yet to be widely integrated into the feedback systems design.

The limitations of existing feedback systems, particularly their tendency to reinforce negative performance, together with the lack of resources to identify relatable role models, motivate us to think: C*an the role-model approach be adapted to feedback systems to provide presentation feedback that avoids negative reinforcement while enhancing reliability?*

2.2. The potential of using AI-clones as role models

An AI-clone refers to the digital replication of an individual's appearance and behavior, made increasingly feasible through recent advancements in deep-learning technologies such as DeepFake, voice conversion, and virtual avatars (Lee et al., 2023; Mika, 2019). These AI-clones exist in various formats, ranging from straightforward text-based chatbots (Neelkumar et al., 2019) to immersive virtual reality experiences that include multi-modal feedback mechanisms (Freeman & Maloney, 2021). AI-clones are created by training machine learning algorithms on data specific to a person, including their vocal characteristics and visual appearances, to emulate their speech and behavior (Maras & Alexandrou, 2019). As a result, AI-clones can convincingly replicate or even refine the unique traits of real people (Fletcher, 2018). This makes it feasible to create a role model using an AI-clone that mirrors speaker-related characteristics while refining specific aspects of their performance, such as pronunciation or voice.

Existing studies on AI-clones have primarily focused on the negative implications of deepfake technologies. However, recent research has begun exploring their potential applications in learning contexts, particularly in simulating peers or instructors. For instance, studies on AI-generated virtual instructors modeled after admired individuals have demonstrated their ability to boost student motivation and foster positive emotions, despite having no significant impact on test scores (Pataranutaporn et al., 2022). These findings suggest that AI-clones, particularly those tailored to individual needs, could provide a unique and *personalized* educational experience. Additionally, AI co-presenters in virtual presentations have been found to reduce anxiety and improve presentation quality, especially among participants who perceive the agents favorably (Kimani et al., 2019, 2021). These findings highlight the potential for AI-clones to serve as role models or co-presenters, offering support to learners facing public speaking anxiety or challenges like linguistic and cultural barriers.

Despite these advances, most studies have focused on using AI-clones to simulate others rather than employing them as self-representations for skill development. Given the growing accessibility and popularity of this technology, there is an urgent need to understand how AI-generated personalized self-images can benefit learning. In addition, while AI-clones offer promising opportunities, they also present challenges, such as concerns about impression management (Hogan, 2010; Stutzman & Hartzog, 2012; Zhao et al., 2013), identity and authenticity (DeVito et al., 2017; Holly, 2015; Michikyan et al., 2015), and the uncanny valley effect (Mori et al., 2012), which can provoke unease in users encountering digital replicas of themselves (Mori et al., 2012; Weisman & Peña, 2021). These complexities raise important questions about the ethical implications of AI-clones and their potential drawbacks in learning contexts.

By addressing these challenges and leveraging the distinct capabilities of AI-clones, this research aims to explore the potential and limitations of using AI-clones as a personalized feedback system that supports learning while fostering positive reinforcement.

2.3. Theoretical lens: social comparison theory and individual regulatory focus

Role models serve as key reference points for social comparison, a psychological process described by *Social Comparison Theory*, which posits that people have an innate drive to evaluate themselves by comparing their performance to others (Leon, 1954). In learning contexts, role models significantly influence self-perceptions, recognition, and productivity, shaping the efficiency of learning experiences and outcomes (Gary et al., 2010; Matthias et al., 2010; Sheldon et al., 2008). Social comparison can take two forms: upward comparison, where individuals compare themselves to others who perform better, and downward comparison, where they compare themselves to those performing worse (Dijkstra, Kuyper, Van der Werf, Buunk, & van der Zee, 2008; Fleur et al., 2023; Hanus & Fox, 2015; Micari & Drane, 2011).

Within this framework, the concept of *Regulatory Focus* provides a deeper understanding of how individuals respond to social comparison information. Regulatory Focus identifies two motivational orientations:

"promotion focus", which emphasizes the pursuit of aspirations and accomplishments, and *"prevention focus"*, which centers on avoiding failures and mistakes (Tory Higgins et al., 2001). These orientations determine whether upward or downward social comparisons are more inspiring and motivating to an individual (Emily et al., 2016; Watling et al., 2012; Zhang, 2016).

Studies have shown that promotion-focused individuals are motivated by *"upward comparisons"*, drawing inspiration from role models who exemplify success and highlight pathways to achievement (Lockwood, 2006; Lockwood et al., 2002; Pasko Arigo et al., 2021). Conversely, prevention-focused individuals are more inspired by *"downward comparisons"*, as they emphasize avoiding mistakes and minimizing risks (Lockwood & Pinkus, 2008; Tory Higgins et al., 2001). This alignment between an individual's regulatory focus and the type of social comparison or role model is referred to as congruence, while a mismatch is termed incongruence (Tory Higgins et al., 2001). For example, Fleur et al. demonstrated that a learning analytics tool displaying students' performance alongside peers with similar goal grades fostered extrinsic motivation and improved academic achievement, aligning comparisons with students' regulatory focus (Fleur et al., 2023). In the context of learning, promotion-focused individuals are most motivated by positive role models who showcase strategies for success. Meanwhile, prevention-focused individuals benefit from negative role models who emphasize avoiding mistakes.

To evaluate the effectiveness of AI-clones as role models for international speakers in online presentation practice, we adopt the concept of *"upward comparison"* by presenting AI-clones as positive role models. Upward comparison is relevant here as the task focuses on inspiring learners to improve, aligning with goals of skill development and aspirational growth, unlike downward comparison, which emphasizes avoiding negative outcomes (e.g., staying above a minimum threshold or avoiding below-average scores). Moreover, this focus addresses the "negative reinforcement" prevalent in current feedback systems (Karl & Kopf, 1994; Zhou et al., 2020), ensuring the emphasis remains on learners' potential rather than their shortcomings. Thus, **H1** examines the comparative effect of AI-clone and the traditional playback approach used in mainstream feedback systems. Recognizing that the impact of social comparison varies based on an individual's regulatory focus, we also investigate how promotion-focused and prevention-focused learners may respond differently to AI-clone. This leads us to **H2**, which explores individual differences in using AI-clone for presentation practice.

2.3.1. H1 (self-recording group v.s. AI-clone group)

To examine the comparative effect of AI-clones and the traditional playback approach used in mainstream feedback systems, we developed several hypotheses based on previous studies on online speech and language training.

2.3.1.1. Self-perception. Supporting speakers in improving their self-perception of speech is crucial, especially since enhanced self-perceptions have a positive effect on social well-being (Cecilia, 2018). Specifically, recent research in speech and language training has investigated how listening to recordings of role-model speakers affects self-perception of speech. Researchers found that exposure to model speakers significantly improved confidence in tone (Orii et al., 2022), when measuring four aspects of self-perceptions on *confidence, satisfaction, communication, and self-expression*. Building on these findings, we aim to compare the effectiveness of using different role models in enhancing speech perception metrics. We hypothesize, **H1 (a)**: *Using AI-clone videos as a role model for speech practice will be more effective than using self-recording videos in improving self-perception of speech performance.*

2.3.1.2. Speech quality. Moreover, while the effects of role models in fostering self-perceptual changes have been studied (Nash et al., 2016; Orii et al., 2022; Scherer et al., 1973), the direct impact of role models on immediate behavioral changes in fluency, pronunciation, or delivery was also explored (Orii et al., 2022; Pataranutaporn et al., 2022). Thus, we hypothesize, **H1 (b)**: *Using AI-clone videos as a role model for speech practice will be more effective than using self-recording videos in improving speech quality.*

2.3.1.3. Self-kindness. Additionally, addressing the issue of "negative reinforcement" in current feedback systems is crucial. Previous studies evaluated the self-kindness (Shepherd & Cardon, 2009) which examined how individuals respond to observed mistakes or imperfections. We aim to understand whether the AI-clone can alleviate the "negative reinforcement" issue previously discussed (Karl & Kopf, 1994; Zhou et al., 2020). By presenting an aspirational version of the self, AI-clones may foster self-kindness, encouraging learners to reflect more positively on mistakes and imperfections, Thus, we hypothesize, **H1 (c)**: *Using AI-clone videos as a role model for speech practice will be more effective than using self-recording videos in improving self-kindness.*

2.3.2. H2 (AI-clone group: promotion focus v.s. prevention focus)

Second, to examine how individual differences influence the effectiveness of AI-clones in speech practice, we employed constructs related to regulatory focus within educational contexts.

2.3.2.1. Learning experience. Previous research has demonstrated that aligning an individual's regulatory focus with the traits of their role models positively impacts their perceptions of the *learning experience* (Järvelä et al., 2023; Reinhold et al., 2021; Zhao et al., 2024). Studies have thoroughly investigated how the alignment between an individual's regulatory focus and the traits of their role models can positively influence their assessment of the learning experience (Järvelä et al., 2023; Reinhold et al., 2021; Zhao et al., 2024). It has been found that learners' experience is considered often boosted when their role models have similar goals and reduced when the goals differ (Jiang & Papi, 2022; Lockwood et al., 2002; Zhang, 2016). A prior study found that when a role model is aligned with learners' regulatory focus, it increases learners' *motivation* to learn and their evaluation of the learning tool as helpful, but they typically do not yield positive evaluations regarding the enjoyment (Zhang, 2016). Another study also reported that adopting a promotion-focused approach has been found to reduce *anxiety* in foreign language practice contexts (Jiang & Papi, 2022), an aspect which is critical in presentation (Nash et al., 2016). Building on this foundation, we hypothesize, **H2 (a)**: *Using AI-cloned videos as a role model for speech practice, individuals with a promotion focus will demonstrate higher evaluation of the learning experience.*

2.3.2.2. Self-perception. Expanding on the rationale in H1, where role models are expected to enhance self-perception, we further examine how regulatory focus moderates these effects, we further hypothesize, **H2 (b)**: *Using AI-cloned videos as a role model for speech practice, individuals with a promotion focus will demonstrate higher improvements in self-perception of speech performance.*

2.3.2.3. Speech quality. Similarly, building upon H1, where role models are expected to speech performance, we extend our investigation into how regulatory focus moderating factor, we hypothesize, **H2 (c)**: *Using AI-cloned videos as a role model for speech practice, individuals with a promotion focus will demonstrate higher improvements in speech quality, compared to those with a prevention focus.*

2.3.3. Measurements

We present details on the measurement process, including the pretest of regulatory focus, in Table 1. We calculated the mean scores for the multi-dimensional variables and reported their reliability.

Table 1
Measurement areas, variables, sample items, and scales used. Note: Yoodli is a third-party, time-stamped, automated delayed speech analysis feedback generating tool on word selection and delivery nuances for presentation assessment used in prior research studies (Joel Whalen & Drehmer, 2023; Panke, 2023).

Areas	Variables	Sample Items	Notes
Self-perception (Orii et al., 2022)	Confidence	I am satisfied and fulfilled with my voice and the way I speak. Confidence is the ability to effectively perform a task.	7-point Likert scale: 1 = strongly disagree to 7 = strongly agree; same statement of each item used for rating for all 9 aspects of speech qualities; collected at 2 time-points
	Satisfaction	I am satisfied with my voice and the way I speak. Satisfaction refers to the fulfillment of happiness, i.e., "Are you happy with what you are evaluating?"	
	Communication	I can communicate what I want to say or convey to the listener, through my voice and the way I speak.	
	Self-Expression	I can express myself through my voice and the way I speak.	
Speech qualities (Scherer et al., 1973)	Word Choice	Repetition, filler words, and weak words, along with their respective percentages.	Presented as statistics; collected at 2 time-points; analyzed by Yoodli
	Delivery	Eye contact percentage, pacing in words per minute, pacing variation, minimum/ maximum pause duration, mean pause duration.	
Self-kindness (Shepherd & Cardon, 2009)		Was your response to yourself regarding the mistakes or imperfections observed in your video positive, negative, or neutral?	Single-choice
Learning experience	Motivation (Lockwood et al., 2002)	I want to integrate AI videos' knowledge with my own knowledge for online speech practice.	7-point Likert scale: 1 = strongly disagree to 7 = strongly agree; 4-item for each variable; Cronbach's alpha ranges from 0.84 to 0.94.
	Helpfulness (Zhang, 2016)	I think learning from AI videos to improve my online speech was helpful.	
	Enjoyment (Zhang, 2016)	I think learning from my AI video was a happy experience.	
	Anxiety (Bartholomay & Houlihan, 2016)	AI videos helped me manage concerns like, "My speech won't impress the audience."	
Regulatory Focus Questionnaire (RFQ) (Tory Higgins et al., 2001)		I feel like I have made progress toward being successful in my life. (Promotion) How often did you obey rules and regulations that were established by your parents? (Prevention)	11 items using 5-point scales with scoring keys
Manipulation Check		How would you rate the performance of the AI-clone for your speech practice, compared to your performance?	7-point scale ranging from 1 (AI's performance is much worse) to 7 (AI's performance is much better)

3. Method

3.1. Multi-modal AI-clone manipulation

To create an AI-clone, or digital replication that is programmed to exhibit the behaviors, language skills, and presentation styles that the learner aims to achieve, providing a positive stimulus, we gathered a 2-minute audio sample and multiple snapshots from each participant as the input data for model training. The goal of this training was to accurately replicate both the vocal and facial characteristics of the participants in the generated videos, as illustrated in Fig. 1, a comparison of *self-recording* and *AI-clones* videos.

The AI-clone is generated through a three-step process. First, we leverage the text-to-speech ElevenLabs toolkit[1] to capture the unique vocal traits of participants and create AI-generated audios (ElevenLabs, 2023). Second, we refine the transcripts from participants' initial presentations by removing repetitive and filler words. These optimized transcripts are then used to produce synthetic audio that maintains a Standard English Accent while preserving participants' original vocal characteristics. Lastly, we use Heygen,[2] a commercial platform that generates customizable AI avatars to replicate participants' facial talking and gestures (HeyGen, 2023), which are captured through snapshots. This synthesized audio is then paired with a participant-selected preset avatar template, and face-swapping techniques are applied to generate AI-clones of the participants.

The manipulation check results (M = 6.69, SD = 0.79) indicate the AI-clones generally function as a positive role model, see measurement in Table 1.

3.2. Study procedure

The experiment lasted approximately 100 minutes and followed the study procedure outlined in Fig. 2. Before the formal experiment, participants read a sample script and recorded a 2-minute video for AI-clone generation. Then, all participants engaged in the following sequential activities.

Self-Concept: To warm up, participants listed 5–10 attributes that represented their *"Actual Self"* (e.g., "I tend to repeat myself in online presentations"). They also described their envisioned self – an *"Ideal Self"* for promotion-focused individuals (e.g., "I want to be persuasive") or an *"Ought-to-Self"* for prevention-focused individuals (e.g., "I must maintain consistent eye contact"). This information was documented in a shared online file. Then, participants selected an outfit from a template

[1] https://elevenlabs.io.
[2] https://www.heygen.com.

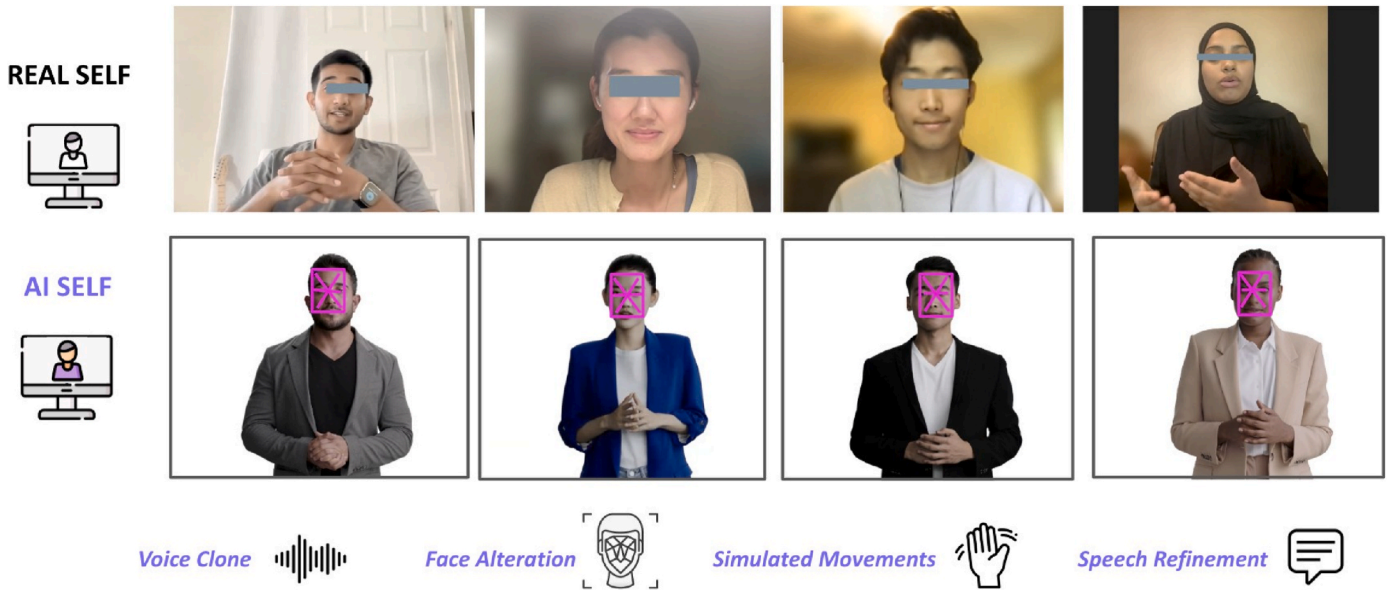


Fig. 1. AI-clones with participants' chosen Avatar template (Images are shared with participants' permission).

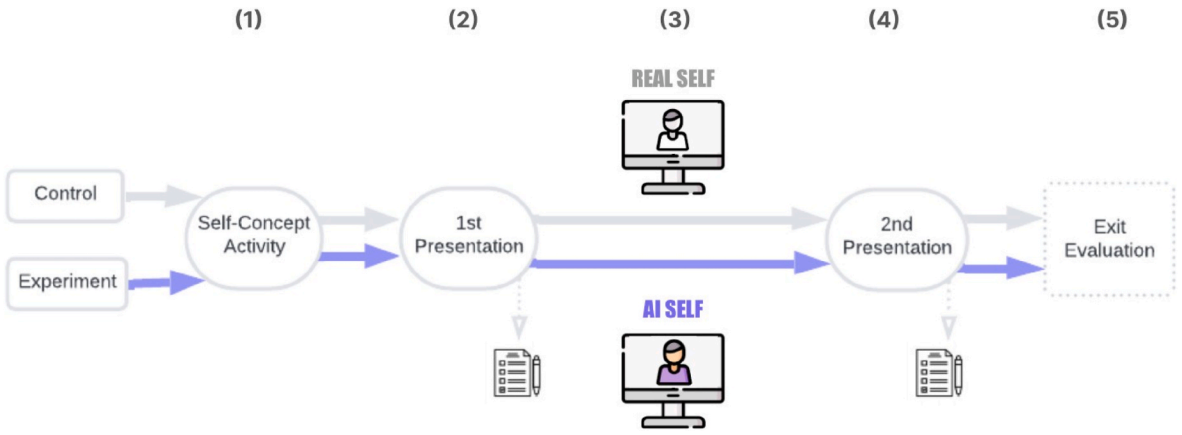


Fig. 2. Study procedure: all participants engaged in the five steps (detailed in Section 3.2).

gallery that aligned with their online speech. ***The* 1st *Presentation:*** Participants had 10 minutes to prepare for a 1-minute short speech in English, to concisely summarize an idea, project, product, or oneself for online business, networking, interviews, or quick essence conveyance – without relying on a script. The speech was recorded for later analysis of presentation quality. Afterward, they finished a questionnaire assessing their self-perception of speech performance. ***Video Replay:*** A stimulated recall (Lyle, 2003) allowed participants to think-aloud while reviewing a video interpretation. Participants first watched a sample video showing the process and were then asked to share their screen as they played back their video, with the option to replay as needed. The AI-clone group viewed AI-clone video based on their initial speeches, while the self-recording group watched replays of their presentations. After this, participants assessed whether the video is a positive role model as part of the manipulation check. ***The 2nd Presentation:*** Reflecting on their self-concept, participants *wrote their goals* for their second presentation and were allotted 10 minute for preparation. Upon completion of their second 1-minute speech, they finished a questionnaire asking about self-perception of speech performance, and evaluating their learning experience. ***Exit Evaluation***: For the exit evaluation, participants discussed their experience watching their video playbacks (either self-recording or AI-clone). Then, they watched a counterpart video—either AI-clone or self-recording, depending on their group—and repeated the step (3). Finally, they were interviewed about using AI-generated videos for learning, discussing its benefits and drawbacks.

3.3. Data analysis

The socio-cognitive view of human behavior describes self-practice as the reciprocal determinism of the environment on the person mediated through behavior (self-observation, self-evaluation, and self-reaction) (Albert Bandura, 1986; Lin et al., 2022). Based on this view, we adopt a holistic evaluation involving Think, Feel, and Act to explore how AI-clones influence speakers' presentation practice experience.

3.3.1. Quantitative data

We employed statistical approaches to evaluate the result for between-group analyses. To ensure a robust evaluation of training outcomes, we selected statistical methods based on the timing and nature of the collected data.

3.3.1.1. Two time-point measures (pre-post). When data were collected both before and after the practice, we used paired sample t-tests to examine within-group changes across various dependent variables. This method allows us to analyze how each group's performance changed due to the practice. Additionally, to control for initial performance levels

and isolate the impact of starting points, we employed the Analysis of Covariance (ANCOVA). This "post-adjust" approach is particularly effective for making comparative assessments between the two groups in randomized pre-post designs where outcomes are measured both at baseline and post-intervention (Wan, 2021). For ANCOVA, we conducted post-diagnostic plots to confirm the linearity and homogeneity of regression slopes, and residual analysis indicated normal distribution and equal variances in these variables, thereby meeting the data assumptions.

3.3.1.2. Single time-point measures. When data were collected or coded only at a single-time point, independent t-tests were applied to assess differences in continuous outcomes between the groups. This "between" approach allows us to compare group means for each dependent variable at that specific time point. If the outcomes are categorical with small sample sizes, Fisher's exact tests were utilized to evaluate differences between groups (Aylmer Fisher, 1970; Teachey et al., 2022). This method provides a more precise assessment of group differences for non-continuous variables.

3.3.2. Qualitative data

Transcripts provided insights into participants' cognitive processes as they interacted with the AI system. Through iterative coding and constant comparison, we captured the nuances of participants' opinions, challenges, and suggestions.

3.3.2.1. Think-Aloud Transcripts. Transcripts have been gathered for the think-aloud (stimulated recall) sessions of each participant when replaying videos. All transcriptions were then carefully reviewed. An experienced researcher, along with two undergraduate research assistants, independently applied and tallied codes to the segmented transcripts based on the predefined coding scheme, which covered various dimensions participants remarked on, such as eye contact, pace, and pronunciation. Subsequent analysis focused on the frequency of these specific codes, referred to as "number of reflections," and the diversity of dimensions discussed, termed "dimension counts." This process enabled the identification of patterns and trends among the groups.

3.3.2.2. Written Goals. Participants wrote goals for the second presentation. First, by reading through their goals, we identified that all participants exhibited some level of attention to the following topics: *technical* proficiency (e.g., pause, word choice, pace, grammar, time, content, pronunciation), and *emotional* resonance in delivery (e.g., eye contact, tone, gesture, emotion, facial expression), echoed with the previous work on two dimensions of evaluating presentation (De et al., 2012). We also found participants, used different verb choices: the *correction words* that show errors, e.g., *"avoid," "don't," "fix," "improve"*, and the *enhancement words* that show efforts, e.g., *"add," "more," "should," "try"*. Then, we quantified the frequency of participants' disclosures tied to these four categories. We compared the differences between technical proficiency and emotional resonance, and correction approach and enhancive approach, and used a binary coding (Y/N) to determine if participants demonstrated a stronger focus for either aspect. These results were then organized into a 2x2 contingency table and proportions were analyzed.

3.3.2.3. Semi-structured interviews. In the semi-structured interviews, participants were guided to express their experience, presentation preparation challenges, and insights on integrating AI videos for presentation skill improvements, They identified AI benefits, proposing integration strategies, and addressing concerns. Interviews were audio-recorded and transcribed through Zoom.

3.4. Participants

3.4.1. Recruitment

We recruited participants through university-hosted events organized by the International Student and Scholar Services (ISSS) and connections established through various social media platforms. Participants were randomly selected based on their responses to a screening survey. Booking invitations were extended to individuals who met the following criteria: 1) Non-native English speakers, 2) Individuals hoping to improve online presentations in English, 3) Age exceeding 18 years. Participants received a copy of consent before the study and our research was approved by the IRB. Participants received $15 of compensation upon the completion of the study.

3.4.2. Randomization

At the start of the study, participants were asked to complete an initial registration survey that collected information on their age range and native language. In addition, they were required to fill out the *"Regulatory Focus Questionnaire"* (Tory Higgins et al., 2001) to assess their regulatory focuses. Using this information, the study implemented a randomized design for dividing participants into two primary groups, each initially targeting 15 participants. However, due to the absence of one participant from the self-recording group, the allocation was slightly adjusted, resulting in 14 participants in the self-recording group (control) and 15 in the AI-clone group (experiment), bringing the total count in these initial groups to 29 participants.

To explore how individual regulatory focus might influence the impact of AI-clone videos, the study further divided the AI-clone group based on participants' regulatory focuses. This division aimed to form two additional groups with a balanced representation of the regulatory focuses to ensure statistical integrity. Stratified random sampling was utilized for this purpose, drawing from the remaining pool of registrants. This approach facilitated the creation of two groups: the promotion-focused group ($n = 15$) and the prevention-focused group ($n = 15$). The final participant count in the two new groups was 30 participants. Consequently, this segmentation increased the total participant count for this study to 44 participants.

3.4.3. Demographics

Of our participants, 45.24% are female and 54.76% are male. Participants' age ranges from 18 to 24 (44.2%) to 25–34 (55.8%). For their native languages, the majority are Mandarin, making up approximately 52.38% of the total. Indians are the second most common, making up approximately 9.52%, followed by Koreans at 11.36%. All other native languages make up 29.55% of the sample, including Bangladeshi, Cantonese, Ecuadorian, French, Hispanic/Latino, Indonesian, Japanese, Jordanian, Kyrgyz, Nepali, Nigerian, Peruvian, and Vietnamese.

4. Findings

4.1. Differences between self-recording and AI-clone groups (RQ1)

4.1.1. H1 (a), rejected

The self-recording group demonstrated significantly higher self-perception of satisfaction in pronunciation than the AI-clone group.

By comparing self-perceptual measures (i.e., confidence, satisfaction, expressiveness, and communication) at two-time points, the ANCOVA showed that when accounting for the baseline, significant *between-group* differences were observed in satisfaction with pronunciation: $F(1, 26) = 10.50$, $p = 0.003$, effect size: $\eta^2 = 0.257$, large effect. The self-recording group showed a higher adjusted mean (*adjusted mean* = 5.70) than the AI-clone group (*adjusted mean* = 4.53) in satisfaction with pronunciation, thus H1a was rejected. See Fig. 3, and the complete result in Appendix Table 5.

A follow-up analysis of *within-group* changes was also conducted, see details of the paired-sample t-tests in Appendix Table. 3. Results showed

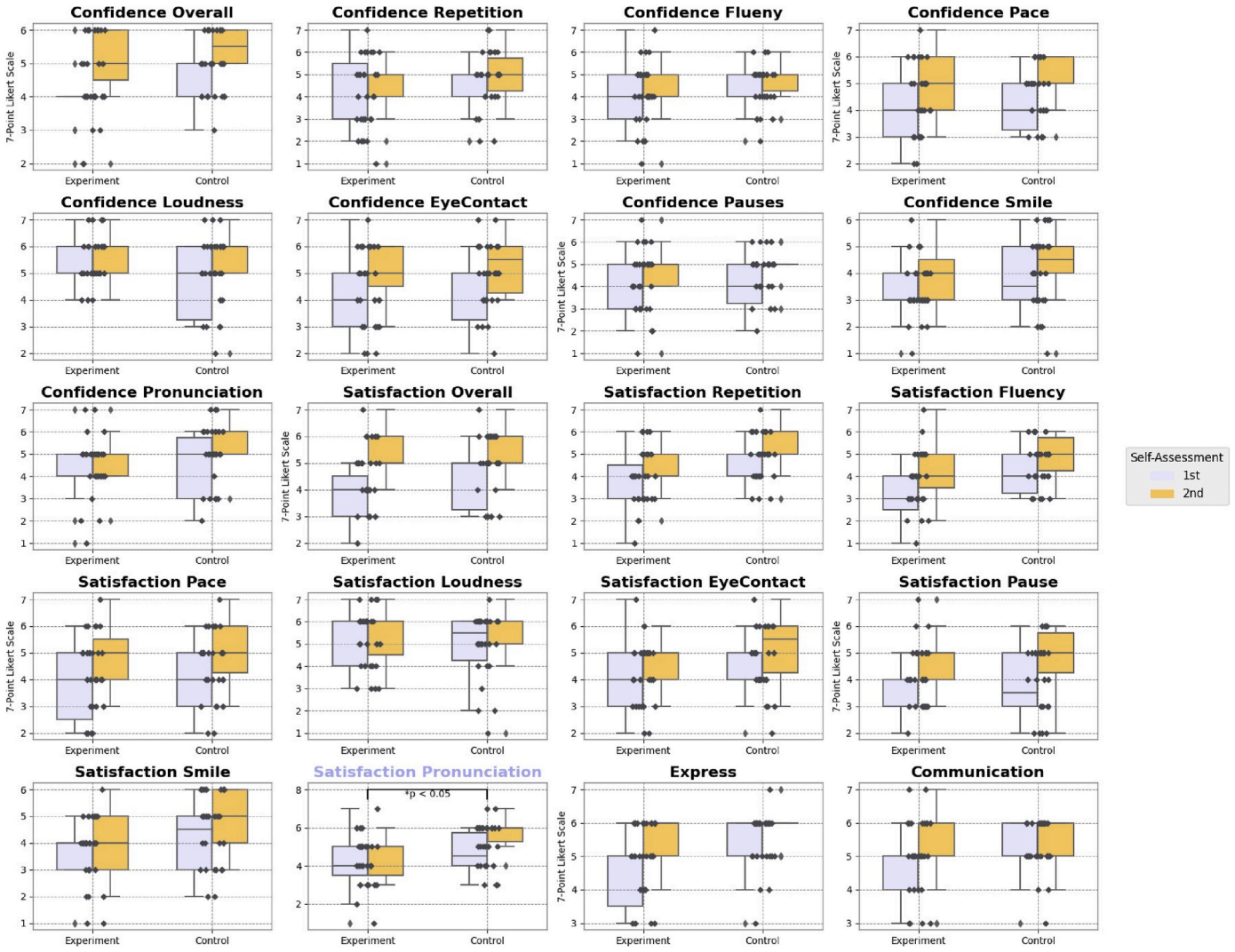


Fig. 3. Pre-test and Post-test *self-perceptions* towards the rated 9 items of speech performance. While many improvements were found within both groups, only *Satisfaction in Pronunciation* yielded a significant difference between groups.

that: 1) both groups exhibited statistically significant improvements in overall confidence, confidence in pace and eye contact, overall satisfaction, satisfaction in fluency, eye contact, pace, and pauses, and expressiveness; 2) The self-recording group showed unique improvements in both confidence and satisfaction of pronunciation ($t = -4.32$), and confidence of Loudness ($t = -2.88$); and 3) the AI-clone group showed unique improvements in communication ($t = -2.48$), satisfaction of repetition ($t = -2.75$), confidence in managing pauses ($t = -2.96$), and confidence of smiles ($t = -2.44$).

4.1.2. H1 (b), partially supported

The AI-clone group demonstrated a significant immediate improvement in speech quality of pace control than the self-recording group.

ANCOVA analysis showed that significant differences between the self-recording and the AI-clone group in controlling pace: $F(1, 26) = 4.50$, $p = 0.044$, effect size: $\eta^2 = 0.04$, small to medium effect. The covariate-adjusted means revealed that the AI-clone group (*adjusted mean* = 114.93 w/m) reported a lower speech speed compared to the self-recording group (*adjusted mean* = 122.71 w/m) during their second performance after adjustments to their baseline performance, thus H1 (b) was partially supported. Post-diagnostic plots and residual analysis confirmed that the data met the assumptions. See details in Fig. 4 and Appendix Table 5.

A follow-up analysis of *within-group* changes was also conducted, see details of the paired-sample t-tests in Appendix Table 3. Results showed that: Only immediate speech performance improvements were found in the AI-clone group, whereas the self-recording group found no actual immediate speech improvement. Specifically, the AI-clone group exhibited significant reductions in filler words ($t = 3.10$), weak words ($t = 3.07$), and their respective percentages ($t = 3.03$, $t = 2.78$), all with p-values below 0.05. In contrast, within the self-recording group, results showed no significant differences in these performance metrics.

4.1.3. H1 (c), supported

The AI-clone group showed higher self-kindness than the self-recording group. Fisher's exact analysis revealed a statistically significant difference in the likelihood of negative self-ratings ($p < 0.05$). Individuals in the self-recordings group were significantly more likely to rate themselves negatively (Odds Ratio: 26.0, 95% CI: 2.61 to 259.29) compared to those in the AI-clone group. This statistical significance (p-value: 0.0017) suggested a strong association between self-recording feedback and an increased likelihood of negative self-assessment, compared to AI-generated feedback. Thus, H1 (c) was supported.

Summary: No significant differences were observed between groups in self-perception measures, except that the self-recording group exhibited significantly greater improvements in pronunciation satisfaction compared to the AI-clone group. Follow-up within-group analysis reinforced this finding, as participants in the self-recording group

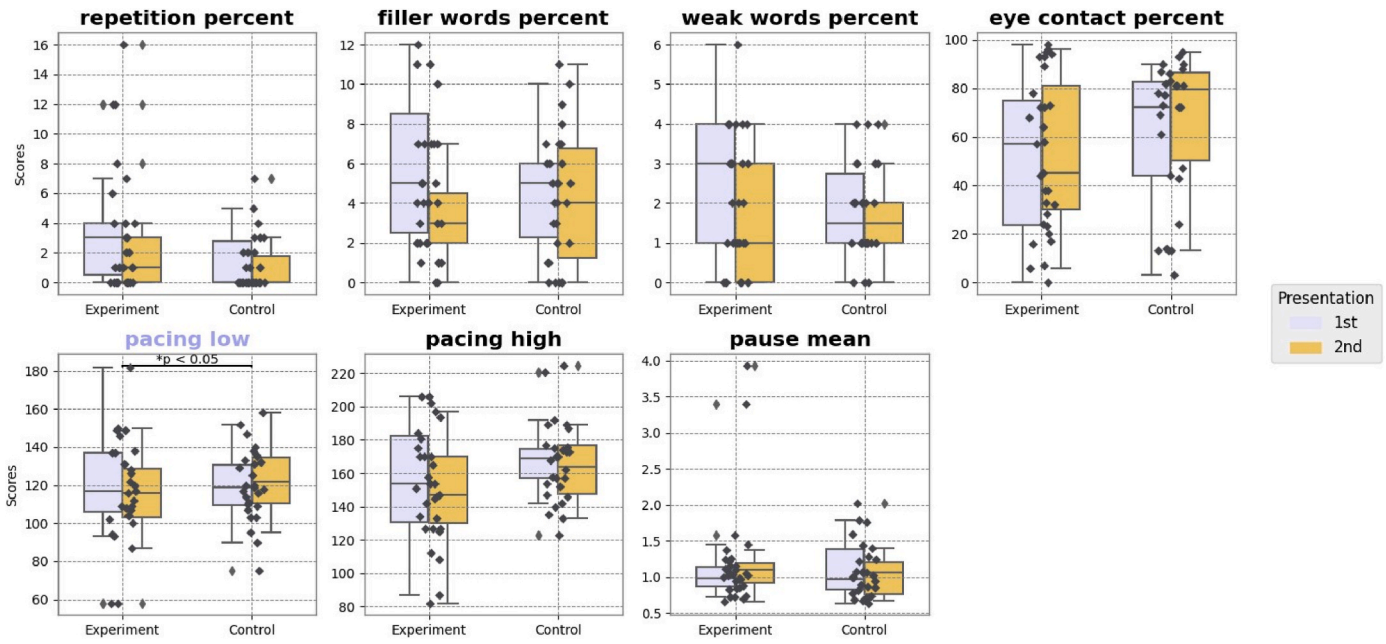


Fig. 4. Pre-test and Post-test Scores of *Speech Quality* for the Self-Recording and AI-Clone Group: Only *pacing control* yielded a significant difference between groups. 1st is original, 2nd is after trained.

showed unique enhancements in both confidence and satisfaction with pronunciation, a trend not seen in the AI-clone group.

Meanwhile, the AI-clone group uniquely improved in self-perception in communication, repetition, pauses, and smiling, which were not observed in the self-recording group. Also, the AI-clone group was significantly more inclined to extend self-kindness. Moreover, participants' actual speech quality only revealed enhancements in the AI-clone group, especially in word choice and pacing control.

4.2. Differences within the AI-clone group (RQ2)

We further explored the influence of AI-clone videos to assess potential individual differences based on regulatory focus, involving participants from both promotion-focused and prevention-focused groups.

4.2.1. H2 (a), partially supported

Promotion-focused individuals reported higher enjoyment, helpfulness, and anxiety reduction than prevention-focused individuals. However, no significant differences were found for motivation, thus, H2 (a) was partially supported. These differences are detailed in Table 2, with statistical evidence provided by the independent *t*-test results with medium to large effects.

4.2.2. H2 (b), partially supported

Promotion-focused individuals exhibited significantly higher confidence in pronunciation than prevention-focused individuals.

ANCOVA analysis indicated significant differences *between-group* in perceived confidence of pronunciation ($F(1, 27) = 5.434$, $p = 0.027$), with a large effect size ($\eta^2 = 0.168$). Promotion-focused participants reported higher confidence levels (*adjustedmean* = 5.40) compared to their prevention-focused counterparts (*adjustedmean* = 4.53) in their second speech, highlighting a notable difference between the groups. Thus, H2 (b) is partially supported. See Appendix Table 6 for details.

Table 2
Independent T-test evaluation of learning experience.

	Prevention (n = 15)	Promotion (n = 15)	t	p
Enjoyment	5.43 ± 1.23	6.33 ± 0.67	−2.486	0.019*
Helpfulness	5.48 ± 1.19	6.30 ± 0.52	−2.431	0.022*
Less Anxiety	4.73 ± 0.96	5.73 ± 0.89	−2.966	0.006**
Motivation	5.68 ± 1.36	6.20 ± 0.54	−1.368	0.182

A follow-up analysis of *within-group* changes was also conducted, see details of the paired-sample t-tests in Appendix Table 4. We found: 1) both groups showed improvements in overall satisfaction, satisfaction in repetition, satisfaction in pace control, satisfaction in eye contact, satisfaction in pauses, and expressiveness; 2) The promotion-focused participants showed unique improvements in 7 aspects: overall confidence, confidence in repetition, confidence in eye contact, confidence in pauses, confidence in pronunciation, satisfaction in fluency, and satisfaction in pronunciation; and 3) the prevention-focused participants demonstrated unique improvements in three aspects: communication, confidence in fluency, confidence in pace.

4.2.3. H2 (c), rejected

Prevention-focused individuals had more immediate improvement dimensions than promotion-focused individuals.

ANCOVA found no significant *between-group* differences in immediate speech performance, thus, H2 (c) was not supported. See details in Appendix Table 6.

A follow-up analysis of *within-group* changes was also conducted, see details of the paired-sample t-tests in Appendix Table 4. We found: 1) both groups had a decrease in filler words and filler words percentage; and 2) unique to the prevention group, there was a significant decrease in repetition and weak words, both in terms of count and percentage.

Summary: Promotion-focused individuals exhibited significantly more reductions in speech anxiety, along with increased enjoyment and perceived usefulness when using AI-clone video for presentation practice, compared to prevention-focused individuals. Also, they reported significantly higher perceived confidence in pronunciation. However, no significant difference in immediate speech performance was found. Both promotion-focused and prevention-focused individuals improved in reducing filler words, but only prevention-focused individuals improved in controlling repetition and weak words.

4.3. More insights from qualitative analysis

Using the data collected from participants' *Think-Aloud Transcripts*, and *Written Goals*, we found that participants in the self-recording and

AI-clone groups diverged in their observations, priorities, and approaches.

4.3.1. AI-clone videos encouraged more coverage of self-observations than self-recordings

Using the transcripts from the video think-aloud sessions, we coded the dimensions participants mentioned. The independent t-tests were conducted to compare the differences between groups watching self-recording and the AI-clone videos. We found that the dimensions observed by AI-clone group ($M = 1.33$ $SD = 1.53$) were statistically higher than the self-recording group ($M = -0.32$, $SD = 2.55$), $t = -2.133$, $p < 0.05$, Cohen's $d = 0.78$, medium effect. The results suggested the AI-clone group was more adept at identifying and reflecting upon various nuanced dimensions of their speech performances (e.g., repetition, filler words, and eye contact) than their counterparts in the self-recording group.

Participants also expressed their desire to emulate specific styles or qualities that they admire or find effective in the AI-clones. We found explicit references to AI, e.g., *"I want to* pace *myself like the AI (excluding the first introduction). I would like to do gestures similar to the AI and also pronounce some words like it."* (P16). The statement not only recognizes a certain style as effective but actively seeks to incorporate elements of that style into the presenter's style. Speakers also extracted valuable lessons from the AI-clone. For example, while some participants recognized the AI's limitation in terms of energy, they were inspired to combine the AI's other strengths with their own desired improvements. For example, *"I think I will try to use a more energetic voice than AI and make my speech more engaging and fun"* (P32). This reflects a nuanced approach to self-improvement.

4.3.2. Goal setting indicates different priorities for emotional resonance

For the written goals for the second delivery, both the self-recording and the AI-clone groups emphasized the importance of technical proficiency and emotional resonance in presentation skills. Some participants primarily targeted the "nuts and bolts", or foundational elements of a presentation—focusing on fluency (P21, P23, P25, P26, P28, P29, P30, P42), pacing (P24, P37, P42, P43), vocabulary (P25, P28, P37, P42), grammar (P30), less repetition (P21), pronunciation (P24), time control (P24), content (P22, P26, P27, P44) and minimizing filler words (p26). Their goals were inherently technical, striving for a well-tuned presentation that was polished in its clarity, form, and structure. For example, one participant said, *"Control my speaking* pace. *Don't speak too fast and give an unclear speech"* (P24), while another aimed to *"Speak fluently, use fewer filler words, think and speak simultaneously, and diversify my vocabulary"* (P25).

Some participants aimed for a well-rounded presentation experience that extended beyond technical prowess to include emotional resonance and audience engagement. They set goals for technical aspects such as, fluency (P14, P16, P18, P32, P34, P36), word choice (P34), pacing (P16, P31, P38), content (P14, P33, P35), pronunciation (P16), and minimizing filler words (P16, P35, P36, P38). However, some of them also mentioned utilizing logical pauses for emphasis (P16), and more also prioritized body language (P14, P15, P16, P31, P33), facial expression (P18, P20, P31, P39, P41), eye contact (P17, P18, P32, P34, P35, P39, P41), tone (P32), and emotional expressiveness (P15, P32, P41) to not just to present information but to make it resonate on a more emotional to forge a deeper connection with the audience. For example, participants prioritize emotive expression to connect with the audience, such as *"I want to appear to be more emotional by adding some gestures during speaking"* (P15); *"be more confident, making eye contact with audiences better, smiling, energetic, and emotional"* (P41).

Further fisher's exact test comparing the goal-setting tendency within each of the self-recording and AI-clone indicated a significant statistical difference in delivery-related tendencies (emotional resonance v.s. technical proficiency). The odds of setting a emotional goal were significantly lower in the self-recording group compared to the AI-clone group (odds ratio = 0.08, 95% CI: 0.008 to 0.85, $p < 0.05$). Notably, even though the AI-clone group observed more observations during their think-aloud sessions than the self-recording group—as seen in the think-aloud observation findings—they still laid significantly more emphasis on emotional resonance aspects for their presentations goal. The findings suggest that participants watching AI-clone recording are significantly more likely to aim goals in emotional aspects compared to those in the self-recording group. This could be attributed to the perceived feedback provided by AI, which may provide them an alternative perspective. However, given the small sample, future research should explore this association in larger, more diverse populations to evaluate the strength of this association.

4.3.3. Linguistic traits indicating different approaches to self-critique

We also coded the two types of linguistic traits (corrective and enhancive) participants used. Some participants took a corrective stance, aiming to rectify perceived weaknesses or limitations in their existing skill set. The aim is to bring the presentation up to a certain standard by eliminating errors or gaps in proficiency and adhering to an outline, among other things. For example, participants were concerned with minimizing mistakes to improve the overall quality of the presentation: *"I should stick to my outline of what to say in paper so I don't take too many pauses and don't look like I am unprepared or don't know what to say"* (P29), *"avoid the repeated words and pause if possible"* (P21), and *"Use some complex vocabs if possible"* (P28). The underlying assumption here is that repeating words, taking unnecessary pauses, or using not complex vocabulary is flaws that need to be corrected for a more polished presentation to avoid weaknesses. Other participants, on the other hand, demonstrated a seemingly more enhancive approach, where the focus was not merely on fixing shortcomings but on elevating already-existing competencies – focusing on amplifying and expanding existing skills. These goals of adding emotional layers, utilizing gestures, and regulating the pace more effectively amplify the strengths of a presentation by adding new layers of complexity and engagement. For example, *"Use one more example or illustrations, and maybe add a bit of hand movement when appropriate"* (P14), and *"Deliver more useful information"* (P20). Rather than correcting flaws, these goals aimed at enriching the content to provide added value to the audience. The implicit understanding is that the basic presentation is already solid; what's needed is the "extra mile" to make it more engaging or more informative.

Fisher's exact test was conducted between the self-recording and the AI-clone groups on the average number of the two approaches (enhancive v.s. corrective) each group took. The result indicated a significant statistical difference in goal-setting approaches between the self-recording and the AI-clone (odds ratio = 0.11, 95% CI: 0.017 to 0.72, $p < 0.05$), indicating the self-recording group is more oriented toward corrective measures, aiming to fix existing weaknesses, while the AI-clone group leans more toward enhancive strategies, seeking to build on existing strengths, although both groups engage in a mix of goal-setting approaches. However, given the small sample, future research should explore this association in larger, more diverse populations to evaluate the strength of this association.

Summary: The qualitative data reveal variations between the groups that observed self-recordings and those that viewed AI-clone videos. These differences are evident in the observation coverage (what participants noticed in their recording), goal-setting priorities (corrective v. s. enhancive strategies), and self-critique approaches. The findings illuminate how using AI-clones as underlying mechanisms (e.g., observation attention, linguistic traits) helps identify and discuss these "unconscious" nuances and concerns associated with their use.

5. Discussion

Our study offers a multi-dimensional exploration of the impact of using AI-clone for online presentation practice and reveals several important theoretical, practical, and ethical implications of leveraging

AI clones in self-presentation contexts.

5.1. Advancing social comparison theory: new insights on AI clones as role models

Our findings contribute to the empirical understanding of social comparison theory in the context of using AI clones of oneself as social comparison entities, a field which has traditionally focused mainly on human role models or using others as role models (Dijkstra, Kuyper, Van der Werf, Buunk, & van der Zee, 2008; Hanus & Fox, 2015; Leon, 1954; Micari & Drane, 2011). In today's digital landscape, where online presentations are prevalent across various platforms—from workplace virtual conferences to self-promotion on social media (Pasquini & Eaton, 2021), for foreign language speakers engaged in self-directed online presentation practice, the quest for relatable role models poses a significant challenge. We found that using AI-clone could potentially serve as an accessible positive role model when practicing online self-presentation.

On one hand, in our investigation, AI-clones were utilized as a means of upward comparison in presentation practice learning. To the best of our knowledge, this research is the first to explore applying *social comparison theory* (Leon, 1954) to the use of one's self-image as a primed role model, marking a departure from previous approaches that used "others" as role models (Latu et al., 2013; Lockwood, 2006; Lockwood et al., 2002; Zhang, 2016). This method of self-comparison does not rely on external cues or suggestions to identify comparison targets, which could otherwise introduce bias or affect the authenticity of the comparison (Ratcliff and McKoon, 1995). Our manipulation check confirmed that the AI clone was indeed perceived as a positive role model, and this is not a self-evident result because individuals tend to perceive positive role models as similar to themselves, rather than superior (Schokker et al., 2010; Suls & Wheeler, 2013). This was supported by manipulation check ratings deviating from the scale's neutral midpoint, indicating that participants viewed their AI-clones' performance as surpassing their original performance. These findings extend existing literature that emphasizes the effectiveness of relatable role models in learning (Latu et al., 2013; Lockwood, 2006).

On the other hand, we investigated the hypothesis that an individual's regulatory focus moderates the influence of role models in contexts featuring positive role models. Our findings validate the theory that alignment between an individual's regulatory focus and their goals leads to increased learning engagement in the new AI-clone context, evidenced by significant increases in *enjoyment* and perceived *helpfulness* of the learning process. This observation is consistent with previous research examining regulatory focus theory within online learning environments (Zhang, 2016). Additionally, we noted a significant reduction in *anxiety*, a vital factor for effective presentation practice, echoing findings from studies that employed different approaches to role modeling in learning and training contexts (Jiang & Papi, 2022; Leitzelar et al., 2020). These insights advance our understanding of how individual differences mediate the effectiveness of AI-based feedback.

5.2. Identifying the practical potentials of AI clones for self-presentation training

We discussed the practical implications of using AI clones for self-presentation training as follows.

5.2.1. Four benefits of adopting AI clones identified

5.2.1.1. Immediate improvements in speech quality (H1-b). H1-b was partially supported, as the AI-clone group demonstrated measurable enhancements in speech pace and word choice, particularly through the reduced use of filler and weak words. These results highlight the potential of AI clones to deliver immediate, actionable feedback that refines technical aspects of speech. This finding complements prior research focused on perceptual changes, offering a more granular understanding of AI's impact on immediate speech performance improvements (Orii et al., 2022; Pataranutaporn et al., 2022).

5.2.1.2. Promoting self-kindness and counteracting negative reinforcement (H1-c). Significant findings showed that the AI-clone group exhibited higher levels of *self-kindness* compared to the self-recording group. This suggests that AI-clone videos may mitigate the "negative reinforcement" effect often associated with traditional playback systems (Dermody & Sutherland, 2019; Karl & Kopf, 1994; Zhou et al., 2020). Previous studies have reported that self-compassion techniques, including self-kindness, help athletes view mistakes as learning opportunities, reducing the negative impact of self-criticism and enhancing well-being (Omori-Miyake, 2021). This transformative approach positions AI clones as valuable tools for enhancing both performance and emotional well-being in learning contexts. Future studies could explore the mechanisms underlying the observed increase in self-kindness, such as the role of reliability, or deeper knowledge about how it may foster acknowledging errors without self-condemnation, allowing for constructive reflection and growth as discussed by previous work (Neff, 2022; Neff & Dahm, 2014, pp. 121–137).

5.2.1.3. Fostering engagement skills (H1-a, follow-up analysis). The AI-clone group demonstrated within-group improvements in engagement skills, such as communication, pauses, and smiling, areas that did not improve in the self-recording group. These results suggest that structured feedback from AI clones provides specific insights and cues that are difficult to identify through self-recording. This aligns with AI's role in enhancing self-reflection and engagement, helping learners refine their naturalness and expressiveness during presentations.

5.2.1.4. Enhancing self-reflection and emotional resonance (qualitative insights). Qualitative analysis revealed that the AI-clone group engaged in deeper *self-reflection*, emphasizing not only technical skills but also *emotional resonance* and *enhancive strategies*. AI clones enabled learners to explore subtle or "unconscious" elements of their performance that traditional playback methods often overlook. By providing objective feedback, AI clones helped learners recognize both their unique *strengths* and areas for improvement, potentially fostering autonomy and competence. These findings align with Self-Determination Theory (SDT), which highlights autonomy, competence, and relatedness as key drivers of motivated behavior (Edward et al., 2012). Additionally, comparing oneself to an AI-enhanced "ideal self" supports autonomy by empowering learners to control their learning trajectory and connect their current abilities to future potential.

On the other hand, results also showed advantages of self-recording on pronunciation (H1-a) and discrepancies between perception and performance (H1-b):

H1-a was rejected, as the self-recording group demonstrated significantly higher satisfaction with pronunciation compared to the AI-clone group. This advantage likely stems from the direct feedback mechanism inherent to self-recording, where participants closely listen to and critique their own pronunciation and volume. This reflective process facilitates awareness of these two aspects of speech, highlighting the unique benefits of self-recording for enhancing awareness of these aspects. While previous studies focused on perceived confidence in tone using recordings (Orii et al., 2022), our findings reveal that self-recording offers new advantages for pronunciation satisfaction and technical skill refinement. Interestingly, a notable discrepancy emerged between self-perception and objective performance metrics: While self-recording participants reported higher satisfaction with pronunciation, pace, and pauses, their performance metrics revealed faster speech and shorter pauses, potentially signaling anxiety rather than genuine improvement. This underscores the importance of integrating

objective feedback mechanisms, such as those offered by AI assessment tools, to complement self-assessment and provide a more balanced understanding of performance.

5.2.2. Impacts of Individual regulatory focus when using AI clones for training

5.2.2.1. Regulatory fit benefits more in learning experience (H2-a). Promotion-focused individuals reported a significantly more positive learning experience when using AI-cloned videos, demonstrating higher levels of enjoyment and perceived helpfulness of the AI tool, along with a notable reduction in anxiety related to speech performance. These findings align with prior research on regulatory fit (Tory Higgins et al., 2001), which suggests that a promotion-focused orientation—centered on seeking gains—when paired with a positive role model enhances the subjective value and emotional benefits of learning systems. This regulatory alignment makes the experience more engaging and emotionally supportive for promotion-focused learners.

5.2.2.2. Regulatory fit demonstrated higher self-perceptions (H2-b). Promotion-focused individuals also exhibited significantly greater perceived confidence in their pronunciation compared to their prevention-focused counterparts. This aligns with the tendency of promotion-focused individuals to respond positively to aspirational and success-oriented feedback, which is a core feature of AI clones.

5.2.2.3. Enhanced speech quality for prevention-focused individuals (H2-c). The results suggested that while the promotion- and prevention-focused individuals showed no differences in performance improvements when comparing their first and second presentations, the prevention-focused participants exhibited immediate improvements in some speech metrics that the promotion-focused group did not show. Specifically, prevention-focused participants showed immediate improvements in specific speech quality metrics, such as reducing filler words, controlling repetition, and minimizing weak words. These improvements were not observed in the promotion-focused group, reflecting the prevention focus's orientation toward avoiding losses and errors (Tory Higgins et al., 2001; Zhang, 2016). This immediate enhancement highlights how AI-based learning systems can cater effectively to prevention-focused individuals by emphasizing precision and error reduction. Given that H2 (c) was not supported, this finding, unreported in prior work, prompted us to hypothesize that AI's most powerful impact might be in "shaping perception" rather than delivering "immediate performance improvements". Nevertheless, perceptual gains should not be undervalued, as they may be the precursors to long-term performance improvements (Schouten et al., 2007; Williams & Ericsson, 2005).

5.2.3. Bringing inspirations for future research on self-representation

Previous studies have demonstrated the effectiveness of role models in various training contexts, including but not limited to social skills (Ladd & Mize, 1983), career development (Lamb et al., 2022), therapy (Rochlin, 1982), leadership, and medical training (Spaans et al., 2023). For example, role models can be integrated to support individuals in addressing challenges such as social anxiety issues with role-playing exercises, allowing them to practice conversations or presentations with different versions of their self-images, without the pressure of real-world interactions. This aligns with findings that highlight the effectiveness of virtual reality and simulated agents in reducing anxiety and improving therapeutic outcomes (Horigome et al., 2020; Wong et al., 2023). By providing contextually relevant feedback and simulating realistic reactions based on a self-presentation, AI clones have the potential to support these applications while fostering self-reflection.

Building on our findings of four key benefits of AI clones—enhancing speech quality, promoting self-kindness in the face of errors, improving engagement skills in interpersonal communication, and fostering emotional resonance—we envision future research further exploring their role in developing personal self-representation skills. Researchers could also examine additional variables relevant to specific contexts or explore potential negative effects on self-image, such as those discussed by Farrar et al. (Farrar et al., 2015). These effects might include an overemphasis on transforming the self to align with the perceived ideal characteristics of the AI clone, potentially leading to unrealistic self-perceptions or reliance on external validation (Kang et al., 2009). Exploring these areas is essential for future studies to deepen our understanding of the broader implications and applications of AI clones.

5.3. Ethical considerations

We reported various concerns based on interview studies, addressing issues related to personal identity and selfhood, concerns about excessive complexity, expectations for alignment in human-AI interactions, and anxieties about inadvertently presenting inconsistent self-images. For detailed qualitative insights, see Appendix B.

Besides the widely discussed ethical considerations surrounding privacy, data security, informed consent, and the potential misuse of AI-generated clones, such as the abusive potential of AI, the displacement of self-identity (Oliver et al., 2015), and negative emotions (Lee et al., 2023), we posit that any AI technology should respect individual differences: Ignoring individual differences in AI design can inadvertently violate ethical obligations. A single approach may not address the unique learning needs, cultural sensitivities, and emotional responses of different users (e.g., individual regulatory focus). This is especially true for international speakers who already face linguistic and cultural barriers (e.g., preferences for maintaining unique accents even though they are not *"standard"*). Customizing AI-clone to individualized needs is not just a matter of educational efficacy but also an ethical obligation to ensure equitable access to educational opportunities.

Our findings also point out that people have varying emotional and cognitive responses to AI technologies. Designers must take these psychological factors into account to avoid unintended negative consequences. For instance, while AI-clone could alleviate public speaking anxiety for some, they might exacerbate feelings of inadequacy or self-consciousness for others, e.g., *Feels too good to be true"* (P41). The potential for negative psychological impact becomes even more acute in educational settings (e.g., while most participants react positively to AI, we noted that two participants mentioned being *"Uncomfortable to face the AI because it copies me "(P37)* and "*I don't like hearing my voice* (P14)". Previous literature (Weisman & Peña, 2021) has shown users often feel uneasy seeing their personality and likeness embodied in interactive applications. With the advent of AI-clones of even-increasing fidelity, issues of impression management (Hogan, 2010; Zhang et al., 2021) and uncanny valley perceptions (Claude, 2011; Weisman & Peña, 2021; Welker et al., 2020) must be addressed to successfully develop applications that use AI-clones.

The ethical deployment of AI-clone in educational settings should also consider accessibility. AI-driven solutions often require advanced software, potentially excluding those who cannot afford or access these resources. For instance, based on our case, training AI requires high-quality audio, image, or sometimes video data, however, this becomes a barrier for those who cannot afford advanced devices. Additionally, the focus on high-quality audio-visual data for model training may marginalize specific groups, such as those who are blind or have speech impairments. Inclusivity means designing for the widest range of users, and this principle should be adhered to when deploying AI-clone in educational contexts. A more extensive focus on the ethical issues that might arise is necessary as a future direction.

6. Conclusion

This study leverages advanced AI technology to create AI-clones,

offering a novel approach to self-training for online presentations. These AI-clones address limitations in current feedback systems, such as negative reinforcement, while serving as self-improvement tools and culturally sensitive role models. By emphasizing the importance of personalization based on individual regulatory focus, the findings offer significant theoretical, practical, and ethical insights for future research and applications.

7. Limitation and future research

The study also comes with some limitations. We primarily recruited an international student sample, representing the younger generation, who are generally more receptive and trusting towards technology, as suggested by prior work (Wu et al., 2022). While the external validity of our findings should be examined in future studies, it is important to note that we do not claim that our findings are applicable to a wider population throughout this paper. Secondly, due to the cost of creating AI-clones with emerging technology, the sample size of this study is limited. However, we have carefully addressed and reported the effect size of our findings, allowing future studies to build on our current findings and associations to further validate the results. Thirdly, the focus of our research was on the immediate effects of interacting with AI-clones, particularly in the context of presentation delivery training. The study, as indicated by its title and content, does not explore the long-term sustainability of the observed behavioral changes. Given that mastering presentation skills requires time, subsequent longitudinal research could delve into whether initial perceptual improvements lead to enduring performance enhancements and whether improvements are maintained over time.

CRediT authorship contribution statement

Qingxiao Zheng: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhuoer Chen:** Writing – review & editing, Software, Resources. **Yun Huang:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2119589 and the AI Research Institutes program by the National Science Foundation and the Institute of Education Sciences, U.S. Department of Education through Award # 2229873 - AI Institute for Transforming Education for Children with Speech and Language Processing Challenges. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education. We sincerely thank the editor and the anonymous reviewers for their insightful feedback and constructive suggestions, which strengthened this work and improve the impact of our findings. We also extend our gratitude to Duo Wang, Jerry Ni, Yukang Xu, and Jiayu Liu for their supports in data collection and the coding process. Their assistance was instrumental to this research.

A Statistic Analysis Results

Table 3
Paired T-Test For AI-Clone and Self-Recording Groups

Variable	T (AI-Clone)	Significance	T (Self-Recording)	Significance
word_count_p1 vs word_count_p2	1.69	No	1.10	No
repetition_p1 vs repetition_p2	1.42	No	−0.17	No
repetition_percent_p1 vs repetition_percent_p2	0.00	No	0.21	No
filler_words_p1 vs filler_words_p2	3.10	Yes	0.54	No
filler_words_percent_p1 vs filler_words_percent_p2	3.03	Yes	0.10	No
weak_words_p1 vs weak_words_p2	3.07	Yes	0.86	No
weak_words_percent_p1 vs weak_words_percent_p2	2.78	Yes	0.59	No
eye_contact_percent_p1 vs eye_contact_percent_p2	−1.05	No	−1.38	No
pacing_w_per_m_p1 vs pacing_w_per_m_p2	0.68	No	−0.51	No
pacing_low_p1 vs pacing_low_p2	1.43	No	−1.23	No
pacing_high_p1 vs pacing_high_p2	2.07	No	0.41	No
pacing_variation_p1 vs pacing_variation_p2	0.78	No	1.01	No
pause_min_p1 vs pause_min_p2	−0.34	No	1.12	No
pause_max_p1 vs pause_max_p2	−0.77	No	0.29	No
pause_mean_p1 vs pause_mean_p2	−0.65	No	0.62	No
Confidence_Overall_p1 vs Confidence_Overall_p2	−5.91	Yes	−5.49	Yes
Confidence_Repetition_p1 vs Confidence_Repetiti …	−0.64	No	−1.23	No
Confidence_Fluency_p1 vs Confidence_F …	−1.15	No	−1.15	No
Confidence_Pace_p1 vs Confidence_Pa …	−2.17	Yes	−2.62	Yes
Confidence_Loudness_p1 vs Confidence_Loudness_p2	0.29	No	−2.88	Yes
Confidence_EyeConfidence_ontaConfidence_t_p1 vs …	−2.38	Yes	−4.76	Yes
Confidence_Pauses_p1 vs Confidence_Pauses_p2	−2.96	Yes	−1.73	No
Confidence_Smile_p1 vs Confidence_Smile_p2	−2.44	Yes	−1.47	No
Confidence_Pronunciation_p1 vs Confid …	−1.72	No	−3.16	Yes
Satisfaction_Overall_p1 vs Satisfaction_Overall_p2	−7.64	Yes	−3.37	Yes
Satisfaction_Repetition_p1 vs Satisfaction_Repe …	−2.75	Yes	−2.06	No
Satisfaction_Fluency_p1 vs Satisfaction_Fluency_p2	−2.61	Yes	−2.35	Yes

(*continued on next page*)

Table 3 (*continued*)

Variable	T (AI-Clone)	Significance	T (Self-Recording)	Significance
Satisfaction_Pacing_p1 vs Satisfaction_Pacing_p2	−3.76	Yes	−2.38	Yes
Satisfaction_Loudness_p …	−0.29	No	−0.71	No
Satisfaction_EyeContact_p1 vs Satisfaction_EyeC …	−2.44	Yes	−3.04	Yes
Satisfaction_Pause_p1 …	−4.43	Yes	−2.90	Yes
Satisfaction_Smile_p1 vs Satisfacti …	−1.42	No	−1.47	No
Satisfaction_Pronunciation_p1 vs Satisfaction_P …	−1.00	No	−4.32	Yes
Express_p1 vs Express_p2	−3.39	Yes	−4.37	Yes
Communication_p1 vs Communication_p2	−2.48	Yes	−0.76	No

Table 4
Paired T-Test For Promotion and Prevention Groups

	T(Promotion)	Significance	T(Prevention)	Significance
word_count_p1 vs word_count_p2	1.495603	No	2.069154	No
repetition_p1 vs repetition_p2	1.963961	No	2.566756	Yes
repetition_percent_p1 vs repetition_percent_p2	0.239229	No	1.857143	No
filler_words_p1 vs filler_words_p2	5.280648	Yes	3.572379	Yes
filler_words_percent_p1 vs filler_words_percent_p2	3.239338	Yes	3.126775	Yes
weak_words_p1 vs weak_words_p2	1.980208	No	4.060835	Yes
weak_words_percent_p1 vs weak_words_percent_p2	1.746130	No	3.536934	Yes
eye_contact_percent_p1 vs eye_contact_percent_p2	−1.666187	No	1.160104	No
pacing_w_per_m_p1 vs pacing_w_per_m_p2	0.017445	No	0.653331	No
pacing_low_p1 vs pacing_low_p2	−0.026353	No	0.148853	No
pacing_high_p1 vs pacing_high_p2	0.268635	No	1.270466	No
pacing_variation_p1 vs pacing_variation_p2	0.325423	No	0.740699	No
pause_min_p1 vs pause_min_p2	0.413899	No	0.437498	No
pause_max_p1 vs pause_max_p2	0.563535	No	−0.490012	No
pause_mean_p1 vs pause_mean_p2	0.715467	No	−0.286169	No
Confidence_Overall_p1 vs Confidence_Overall_p2	−5.870395	Yes	−1.718466	No
Confidence_Repetition_p1 vs Confidence_Repetiti …	−2.846974	Yes	−0.263589	No
Confidence_Flueny_p1 vs Confidence_Flueny_p2	−1.585356	No	−2.958040	Yes
Confidence_Pace_p1 vs Confidence_Pace_p2	−1.581139	No	−2.673651	Yes
Confidence_Loudness_p1 vs Confidence_Loudness_p2	−2.085549	No	−1.381699	No
Confidence_EyeContact_p1 vs Confidence_EyeConta …	−2.442292	Yes	−1.457256	No
Confidence_Pauses_p1 vs Confidence_Pauses_p2	−3.055050	Yes	−1.975658	No
Confidence_Smile_p1 vs Confidence_Smile_p2	−1.657850	No	−1.748646	No
Confidence_Pronunciation_p1 vs Confidence_Pronu …	−3.666494	Yes	−0.211154	No
Satisfaction_Overall_p1 vs Satisfaction_Overall_p2	−6.204837	Yes	−3.500000	Yes
Satisfaction_Repetition_p1 vs Satisfaction_Repe …	−3.089572	Yes	−2.442292	Yes
Satisfaction_Fluency_p1 vs Satisfaction_Fluency_p2	−4.063180	Yes	−1.870829	No
Satisfaction_Pace_p1 vs Satisfaction_Pace_p2	−3.095624	Yes	−3.378463	Yes
Satisfaction_Loudness_p1 vs Satisfaction_Loudne …	−2.085549	No	−1.701926	No
Satisfaction_EyeContact_p1 vs Satisfaction_EyeC …	−3.162278	Yes	−3.055050	Yes
Satisfaction_Pause_p1 vs Satisfaction_Pause_p2	−3.240370	Yes	−2.449490	Yes
Satisfaction_Smile_p1 vs Satisfaction_Smile_p2	−0.844616	No	−1.895158	No
Satisfaction_Pronunciation_p1 vs Satisfaction_P …	−2.954989	Yes	−0.564076	No
Express_p1 vs Express_p2	−2.322693	Yes	−3.674235	Yes
Communication_p1 vs Communication_p2	−0.487122	No	−2.256304	Yes

Table 5
ANCOVA Result For Self-Recording and AI-Clone Groups

index	F_value	p_value	df_group	df_error	eta_squared
word_count_p2	0.0152	0.9027	1	26	0.0003
repetition_p2	0.0667	0.7982	1	26	0.0025
repetition_percent_p2	0.1766	0.6778	1	26	0.0048
filler_words_p2	2.2560	0.1451	1	26	0.0563
filler_words_percent_p2	3.3421	0.0790	1	26	0.0763
weak_words_p2	1.1586	0.2916	1	26	0.0283
weak_words_percent_p2	1.0889	0.3063	1	26	0.0310
eye_contact_percent_p2	0.6808	0.4168	1	26	0.0073
pacing_w_per_m_p2	1.2174	0.2800	1	26	0.0067
pacing_low_p2	4.4982	0.0436	1	26	0.0390
pacing_high_p2	1.5016	0.2314	1	26	0.0173
pacing_variation_p2	0.3609	0.5532	1	26	0.0108
pause_min_p2	1.0610	0.3125	1	26	0.0061
pause_max_p2	0.5418	0.4683	1	26	0.0072
pause_mean_p2	0.7247	0.4024	1	26	0.0047
Confidence_Overall_p2	0.1133	0.7391	1	26	0.0025

(*continued on next page*)

Table 5 (*continued*)

index	F_value	p_value	df_group	df_error	eta_squared
Confidence_Repetition_p2	1.8943	0.1805	1	26	0.0670
Confidence_Fluency_p2	0.3840	0.5409	1	26	0.0130
Confidence_Pace_p2	0.1076	0.7456	1	26	0.0041
Confidence_Loudness_p2	2.6903	0.1130	1	26	0.0689
Confidence_EyeContact_p2	0.3810	0.5424	1	26	0.0087
Confidence_Pauses_p2	0.0325	0.8584	1	26	0.0010
Confidence_Smile_p2	0.2277	0.6372	1	26	0.0072
Confidence_Pronunciation_p2	2.6962	0.1126	1	26	0.0737
Satisfaction_Overall_p2	0.0121	0.9134	1	26	0.0003
Satisfaction_Repetition_p2	1.8033	0.1909	1	26	0.0551
Satisfaction_Fluency_p2	0.9456	0.3398	1	26	0.0322
Satisfaction_Pacing_p2	0.0877	0.7694	1	26	0.0027
Satisfaction_Loudness_p2	0.0186	0.8926	1	26	0.0003
Satisfaction_EyeContact_p2	1.4508	0.2393	1	26	0.0405
Satisfaction_Pause_p2	0.0087	0.9265	1	26	0.0003
Satisfaction_Smile_p2	3.5334	0.0714	1	26	0.1195
Satisfaction_Pronunciation_p2	10.4976	0.0033	1	26	0.2566
Express_p2	1.2524	0.2733	1	26	0.0295
Communication_p2	0.1626	0.6901	1	26	0.0046

Table 6
ANCOVA Result For Promotion and Prevention Groups

Dependent Variable	F Value	p Value	df Group	df Error	Eta Squared
word_count_p2	0.132	0.719	1.000	27.000	0.001
repetition_p2	0.132	0.720	1.000	27.000	0.004
repetition_percent_p2	0.239	0.629	1.000	27.000	0.006
filler_words_p2	1.353	0.255	1.000	27.000	0.010
filler_words_percent_p2	1.567	0.221	1.000	27.000	0.018
weak_words_p2	0.218	0.644	1.000	27.000	0.003
weak_words_percent_p2	0.088	0.768	1.000	27.000	0.002
eye_contact_percent_p2	1.942	0.175	1.000	27.000	0.031
pacing_w_per_m_p2	0.192	0.665	1.000	27.000	0.001
pacing_low_p2	0.005	0.942	1.000	27.000	0.000
pacing_high_p2	0.440	0.513	1.000	27.000	0.003
pacing_variation_p2	0.706	0.408	1.000	27.000	0.020
pause_min_p2	0.003	0.956	1.000	27.000	0.000
pause_max_p2	0.570	0.457	1.000	27.000	0.011
pause_mean_p2	0.492	0.489	1.000	27.000	0.004
Confidence_Overall_p2	0.918	0.346	1.000	27.000	0.022
Confidence_Repetition_p2	1.387	0.249	1.000	27.000	0.048
Confidence_Flueny_p2	0.159	0.693	1.000	27.000	0.005
Confidence_Pace_p2	0.278	0.602	1.000	27.000	0.010
Confidence_Loudness_p2	0.040	0.843	1.000	27.000	0.001
Confidence_EyeContact_p2	0.258	0.615	1.000	27.000	0.007
Confidence_Pauses_p2	0.053	0.820	1.000	27.000	0.001
Confidence_Smile_p2	0.109	0.744	1.000	27.000	0.003
Confidence_Pronunciation_p2	5.434	0.027	1.000	27.000	0.101
Satisfaction_Overall_p2	1.770	0.195	1.000	27.000	0.036
Satisfaction_Repetition_p2	0.061	0.807	1.000	27.000	0.001
Satisfaction_Fluency_p2	2.348	0.137	1.000	27.000	0.063
Satisfaction_Pace_p2	0.349	0.559	1.000	27.000	0.009
Satisfaction_Loudness_p2	0.358	0.555	1.000	27.000	0.006
Satisfaction_EyeContact_p2	0.060	0.808	1.000	27.000	0.001
Satisfaction_Pause_p2	0.052	0.821	1.000	27.000	0.002
Satisfaction_Smile_p2	0.140	0.711	1.000	27.000	0.004
Satisfaction_Pronunciation_p2	3.044	0.092	1.000	27.000	0.057
Express_p2	0.060	0.809	1.000	27.000	0.001
Communication_p2	0.964	0.335	1.000	27.000	0.020

B Technology Drawbacks and Ethical Concerns

Below we also report the concerns from participants' interviews.

Personal identity and self-hood are of utmost importance to make sure their safety use

The central concern related to "self" underscores the apprehension regarding AI-clones potentially "dehumanizing" the individuals they replicate. This fear is not unfounded, as participants explicitly voiced worries over their personal identity being misappropriated or used without their consent. For instance, one participant emphasized the need for autonomy and security by stating, *"I want to make sure that my AI-clone will not be used elsewhere*

without me knowing that (P29).”

Moreover, the desire for control over one’s digital representation was evident, as another participant expressed a need for a mechanism to distinguish their AI-clone from their actual self, highlighting a nuanced approach to identity management in the digital age. They remarked, “I want a a package myself. Although AI is mirroring me, but I want a way to deliberately manipulate those differences so no one can tell if this is my AI-clone besides myself (P4).” These responses reveal a deep-seated concern for personal agency and the integrity of one’s identity in the face of advancing AI technology. They reflect a collective desire among participants to navigate the delicate balance between leveraging AI for personal or professional gain and safeguarding their individuality against potential misuse or exploitation.

AI-clone’s advanced language enhances learning, but excessive complexity can reduce engagement

When comparing the AI-clones with the advanced vocabulary, participants noted the AI’s advanced version facilitated learning through its use of an expanded vocabulary. For instance, one participant observed, *“It used more expanded vocabulary. You know like this avatar is speaking, unscripted. For example, “soothing”. “I am aware of that word, but I didn’t really think of using that word during my presentation. I felt like that skill is really particular to a native speaker. You know just in importing the other words. I’m not sure if that’s the right expression. The AI was less academic, more soothing, and more caring, which felt contextually appropriate. I think I used this word in my second presentation (P7).”*

Another participant shared a different perspective: *“Hearing AI’s speech, I realize that there should be a story! I am like an audience listening to myself speaking. The AI is based on me, but it is so good that I feel it is another person. I feel like I can see my speech from a different person. This is very similar to me, and I think this is very important. If a person is different from me I will think this is not my style. Each person has different characteristics. We have to consider my true level of personality, and tones (P11).”* These insights highlight a delicate balance between utilizing advanced language for educational purposes and ensuring it doesn’t overwhelm or alienate learners by straying too far from their personal expression and comfort level.

Will it Represent Me? Controlling self-clone identities to foster alignments in human-AI interactions

The AI-clone is a standard version and should allow for scenario customization. *“I want to control the ambient of the environment. I want to adjust the tone, to a more enlightening, more serious tone. Sometimes there’s a serious presentation, it depends on the situation. Sometimes you need less smile, even with your mouth closed. Don’t show your teeth at all. Have some templates so we can choose, e.g., this is for a school interview, this is for a job interview, so you can professional advice (P3).”* This sentiment underscores a deeper concern among participants about the potential of AI-clones to accurately and consistently represent their human counterparts in various settings. The desire for scenario-specific customization reflects a need for authenticity and appropriate presentation, balancing professionalism with personal expression.

Moreover, the expectation that AI-clones should evolve over time to mirror changes in the individual’s appearance and preferences speaks to a broader anticipation of AI as dynamic entities capable of growth. For example, *“I think I want to make it prettier, wearing make-up because sometimes I don’t really want to do that in an online presentation. But you know, it can’t feel too good to be true. Also, if my hair style changed, it should change as well (P41).”* These insights reveal a complex relationship between individuals and their AI counterparts, highlighting a keen interest in ensuring these digital clones do not merely mimic but genuinely represent the person’s evolving identity. This approach not only addresses concerns of authenticity and trust but also reflects a nuanced understanding of the role AI can play in extending personal identity into digital spaces.

Navigating the complexities of AI-clones in AI-mediated interpersonal relationships

Participants raised concerns about the potential for AI-clones to disrupt interpersonal relationships, especially when nuances of personal interaction are overlooked or when individuals are unaware they’re engaging with an AI-clone. This issue is particularly pronounced in contexts where AI-clones are utilized for specific training purposes but may not fully encapsulate the broad spectrum of human communication. For example, the lack of capturing the subtleties in interpersonal dynamics by AI-clones can lead to misunderstandings. As some participants noted, *“An overly positive version of me (P3)”*, and *“Sometimes they don’t blink and they are always smiling in this official expression. I want myself to be viewed as a “well-educated” person, but I need a way to tell AI what do I perceive “well-educated” is (P6).”*

Moreover, participants worry about presenting a consistent image across different platforms and situations. *“I am worried if I only use it for some presentation, but not all aspects of my speech and language, will my presentation leave an impression to the audience that is not consistent with who I really am? Similarly, if my AI alone is shared with others, it will also have this issue (P36).”* These insights suggest that while AI-clones offer innovative opportunities for personal development and training, there’s a critical need for these technologies to more accurately reflect the individual’s personality and communication style. Ensuring that users and their audiences are aware of when they are interacting with AI-clones is vital for maintaining trust and authenticity in interpersonal relationships.

References

Albert Bandura. (1986). *Social foundations of thought and action* (Vol. 1986, pp. 23–28). NJ: Englewood Cliffs, 1986.

Alene, M. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press.

Ann, E. S. (2020). Self-disclosure versus self-presentation on social media. *Current opinion in psychology, 31*(2020), 1–6.

Asadi, R., Trinh, H., Fell, H. J., & Bickmore, T. W. (2017). IntelliPrompter: Speech-based dynamic note display interface for oral presentations. In *Proceedings of the 19th ACM international conference on multimodal interaction* (pp. 172–180).

Aylmer Fisher, R. (1970). Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution* (pp. 66–70). Springer.

Barmaki, R., & Hughes, C. E. (2018). Embodiment analytics of practicing teachers in a virtual immersive environment. *Journal of Computer Assisted Learning, 34*(4), 387–396, 2018.

Bartholomay, E. M., & Houlihan, D. D. (2016). Public speaking anxiety scale: Preliminary psychometric data and scale validation. *Personality and Individual Differences, 94* (2016), 211–215.

Bu, Y., Ma, T., Li, W., Zhou, H., Jia, J., Chen, S., et al. (2021). PTeacher: A computer-aided personalized pronunciation training system with exaggerated audio-visual corrective feedback. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–14).

Cecilia, P. (2018). Psychological well-being, improved self-confidence, and social capacity: Bibliotherapy from a user perspective. *Journal of Poetry Therapy, 31*(2), 124–134, 2018.

Claude, D. (2011). Intermediaries: Reflections on virtual humans, gender, and the uncanny valley. *AI & Society, 26*, 319–327, 2011.

Croft, R. J., Gonsalvez, C. J., Gander, J., Lechem, L., & Barry, R. J. (2004). Differential relations between heart rate and skin conductance, and public speaking anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 35*(3), 259–271, 2004.

Dale, H. S. (1987). Peer models and children’s behavioral change. *Review of Educational Research, 57*(2), 149–174, 1987.

Damian, I., Tan, C. S., Baur, T., Schöning, J., Luyten, K., & André, E. (2015). Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems* (pp. 565–574).
Davis, D., Chen, G., Jivet, I., Hauff, C., & Houben, G.-J. (2016). Encouraging metacognition & self-regulation in MOOCs through increased learner feedback. In *LAL@ lak* (pp. 17–22).
De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129–142, 2012.
Dermody, F., & Sutherland, A. (2019). Recommendations from a study of a multimodal positive computing system for public speaking. In *Computer vision, imaging and computer graphics theory and applications: 13th international joint conference, VISIGRAPP 2018 funchal–madeira, Portugal, january 27–29, 2018, revised selected papers 13* (pp. 81–92). Springer.
DeVito, M. A., Birnholtz, J., & Hancock, J. T. (2017). Platforms, people, and perception: Using affordances to understand self-presentation on social media. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 740–754).
Dijkstra, P., Kuyper, H., Van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*(4), 828–879. https://doi.org/10.3102/0034654308321210, 2008.
Du, H. (2015). American college students studying abroad in China: Language, identity, and self-presentation. *Foreign Language Annals, 48*(2), 250–266, 2015.
Edward, L. D., & Ryan, R. M. (2012). Self-determination theory. *Handbook of theories of social psychology, 1*(20), 416–436, 2012.
ElevenLabs. (2023). Language and voice generator. https://elevenlabs.io/about.
Emily, Q. R., & Miele, D. B. (2016). Do you have an opportunity or an obligation to score well? The influence of regulatory focus on academic test performance. *Learning and Individual Differences, 45*, 114–127.
Farrar, S., Stopa, L., & Turner, H. (2015). Self-imagery in individuals with high body dissatisfaction: The effect of positive and negative self-imagery on aspects of the self-concept. *Journal of Behavior Therapy and Experimental Psychiatry, 46*(2015), 8–13.
Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal, 70*(4), 455–471, 2018.
Fleur, D. S., van den Bos, W., & Bredeweg, B. (2023). Social comparison in learning analytics dashboard supporting motivation and academic achievement. *Computers and Education Open, 4*(2023), Article 100130.
Freeman, G., & Maloney, D. (2021). Body, avatar, and me: The presentation and perception of self in social virtual reality. In *Proceedings of the ACM on human-computer interaction 4, CSCW3* (pp. 1–27), 2021.
Gary, M. O., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227–252, 2010.
Giraud, T., Soury, M., Hua, J., Delaborde, A., Tahon, M., Jauregui, D. A. G., et al. (2013). Multimodal expressions of stress during a public speaking task: Collection, annotation and global analyses. In *2013 humaine association conference on affective computing and intelligent interaction* (pp. 417–422). IEEE.
Guerra, J., Hosseini, R., Somyurek, S., & Brusilovsky, P. (2016). An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement. In *Proceedings of the 21st international conference on intelligent user interfaces* (pp. 152–163).
Gunawan, G., Kristiawan, M., Risdianto, E., & Monicha, R. E. (2021). Application of the zoom meeting application in online learning during the pandemic. *Education Quarterly Review, 4*, 2, 2021.
Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & education, 80*(2015), 152–161.
Hazenberg, S., & Hulstun, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17*(2), 145–163, 1996.
HeyGen. (2023). AI video making. https://www.heygen.com/blog.
Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society, 30*(6), 377–386, 2010.
Holly, M. P. (2015). Conjuring the ideal self: An investigation of self-presentation in video game avatars. *Press Start, 2*(1), 1–20, 2015.
Horigome, T., Kurokawa, S., Sawada, K., Kudo, S., Shiga, K., Mimura, M., et al. (2020). Virtual reality exposure therapy for social anxiety disorder: A systematic review and meta-analysis. *Psychological Medicine, 50*(15), 2487–2497, 2020.
Iftekhar Tanveer, M., Lin, E., & Hoque, M. (2015). Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th international conference on intelligent user interfaces* (pp. 286–295).
Jacobson, J. (2020). You are a brand: Social media managers' personal branding and "the future audience". *The Journal of Product and Brand Management, 29*(6), 715–727, 2020.
Järvelä, S., Nguyen, A., Vuorenmaa, E., Malmberg, J., & Järvenoja, H. (2023). Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior, 144*(2023), Article 107737.
Jiang, C., & Papi, M. (2022). The motivation-anxiety interface in language learning: A regulatory focus perspective. *International Journal of Applied Linguistics, 32*(1), 25–40, 2022.
Jo, H., & Ippolito, K. (2018). A skill to be worked at: Using social learning theory to explore the process of learning from role models in clinical settings. *BMC Medical Education, 18*(1), 1–8, 2018.
Joel Whalen, D., & Drehmer, C. (2023). My favorite assignment: Selections from the abc 2022 annual international conference, tampa, Florida: A sunrise of classroom-tested pedagogy. *Business and Professional Communication Quarterly, 86*(2), 187–206, 2023.
Julius Meisel, C., & Blumberg, C. J. (1990). The social comparison choices of elementary and secondary school students: The influence of gender, race, and friendship. *Contemporary Educational Psychology, 15*(2), 170–182, 1990.
Kang, Y. S., Hong, S., & Lee, H. (2009). Exploring continued online service usage behavior: The roles of self-image congruity and regret. *Computers in Human Behavior, 25*(1), 111–122, 2009.
Kang, J., & Liu, B. (2019). A similarity mindset matters on social media: Using algorithm-generated similarity metrics to foster assimilation in upward social comparison. *Social Media+ Society, 5*(4), Article 2056305119890884, 2019.
Karl, K. A., & Kopf, J. M. (1994). Will individuals who need to improve their performance the most, volunteer to receive videotaped feedback? *Journal of Business Communication, 31*(3), 213–223, 1994.
Kimani, E., Bickmore, T., Trinh, H., & Pedrelli, P. (2019). You'll be great: Virtual agent-based cognitive restructuring to reduce public speaking anxiety. In *2019 8th international conference on affective computing and intelligent interaction (ACII)* (pp. 641–647). IEEE.
Kimani, E., Parmar, D., Murali, P., & Bickmore, T. (2021). Sharing the load online: Virtual presentations with virtual co-presenter agents. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1–7).
Ladd, G. W., & Mize, J. (1983). A cognitive–social learning model of social-skill training. *Psychological Review, 90*(2), 127, 1983.
Lamb, E., Bryan, B., & Alberti, H. (2022). The impact of role modelling on the future general practitioner workforce: A systematic review. *Education for Primary Care, 33*(5), 265–279, 2022.
Latu, I. M., Schmid Mast, M., Lammers, J., & Bombari, D. (2013). Successful female leaders empower women's behavior in leadership tasks. *Journal of Experimental Social Psychology, 49*(3), 444–448, 2013.
Lee, P. Y. K., Ma, N. F., Kim, I.-J., & Yoon, D. (2023). Speculating on risks of AI clones to selfhood and relationships: Doppelganger-phobia, identity fragmentation, and living memories. In *Proceedings of the ACM on human-computer interaction 7, CSCW1* (pp. 1–28), 2023.
Leitzelar, B. N., Blom, L. C., Guilkey, J., Bolin, J., & Mahon, A. (2020). Regulatory fit: Impact on anxiety, arousal, and performance in college-level soccer players. *International Journal of Exercise Science, 13*(5), 1430, 2020.
Leon, F. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140, 1954.
Li, B., Scott, O. K. M., Naraine, M. L., & Ruihley, B. J. (2021). Tell me a story: Exploring elite female athletes' self-presentation via an analysis of Instagram stories. *Journal of Interactive Advertising, 21*(2), 108–120, 2021.
Lin, S. T., & Zhang, L. J. (2022). Can self-regulation be transferred to second/foreign language learning and teaching? Current status, controversies, and future directions. *Applied Linguistics, 43*(3), 587–595, 2022.
Lockwood, P. (2006). "Someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly, 30*(1), 36–46, 2006.
Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology, 83*(4), 854, 2002.
Lockwood, P., & Pinkus, R. T. (2008). *The impact of social comparisons on motivation*, 2008.
Luong, K. T., Knobloch-Westerwick, S., & Niewiesk, S. (2020). Superstars within reach: The role of perceived attainability and role congruity in media role models on women's social comparisons. *Communication Monographs, 87*(1), 4–24, 2020.
Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861–878, 2003.
Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof, 23*(3), 255–262, 2019.
Matthias, J. S., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 195, 2010.
Micari, M., & Drane, D. (2011). Intimidation in small learning groups: The roles of social-comparison concern, comfort, and individual characteristics in student academic outcomes. *Active Learning in Higher Education, 12*(3), 175–187, 2011.
Michikyan, M., Dennis, J., & Subrahmanyam, K. (2015). Can you guess who I am? Real, ideal, and false self-presentation on Facebook among emerging adults. *Emerging Adulthood, 3*(1), 55–64, 2015.
Michinov, E., & Michinov, N. (2001). The similarity hypothesis: A test of the moderating role of social comparison orientation. *European Journal of Social Psychology, 31*(5), 549–555, 2001.
Mika, W. (2019). The emergence of deepfake technology: A review. *Technology innovation management review, 9*, 11, 2019.
Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine, 19*(2), 98–100, 2012.
Muir, C., Dörnyei, Z., & Adolphs, S. (2021). Role models in language learning: Results of a large-scale international survey. *Applied Linguistics, 42*(1), 1–23, 2021.
Murphey, T. (1998). *Motivating with near peer role models*. On JALT97: Trends & Transitions, 1998.
Myles, P. B., & Diehl, R. (1970). *Effect of videotape playback and teacher comment on anxiety during subsequent task performance*, 1970.
Naruse, K., Takamichi, S., Tanikawa, T., Yoshida, S., Narumi, T., & Hirose, M. (2019). Estimating confidence in voices using crowdsourcing for alleviating tension with altered auditory feedback. In *Proceedings of asian CHI symposium 2019: Emerging HCI research collection* (pp. 15–22).

Nash, G., Crimmins, G., & Oprescu, F. (2016). If first-year students are afraid of public speaking assessments what can teachers do to alleviate such anxiety? *Assessment & Evaluation in Higher Education, 41*(4), 586–600, 2016.

Neelkumar, P. P., Parikh, D. R., Patel, D. A., & Patel, R. R. (2019). AI and web-based human-like interactive university chatbot (UNIBOT). In *2019 3rd international conference on electronics, communication and aerospace technology (ICECA)* (pp. 148–150). IEEE.

Neff, K. D. (2022). Self-Compassion: Theory, method, research, and intervention. *Annual Review of Psychology, 74*(1), 193–218. https://doi.org/10.1146/annurev-psych-032420-031047

Neff, K. D., & Dahm, K. A. (2014). *Self-Compassion: What it is, what it does, and how it relates to mindfulness*. New York: Springer. https://doi.org/10.1007/978-1-4939-2263-5_10

Ngai, K. C. (2019). "Becoming an expert in driving for Uber": Uber driver/bloggers' performance of expertise and self-presentation on YouTube. *New Media & Society, 21*(9), 2048–2067, 2019.

Oliver, L. H., Bowser, A. E., Melcer, E. F., & Churchill, E. F. (2015). Online inspiration and exploration for identity reinvention. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3809–3818).

Omori-Miyake, M. (2021). The use of self-compassion techniques in elite footballers: Mistakes as opportunities to learn. *The Cognitive Behaviour Therapist, 15*(2021).

Orii, L., Ogawa, N., Hatada, Y., & Narumi, T. (2022). Designing for speech practice systems: How do user-controlled voice manipulation and model speakers impact self-perceptions of voice?. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–14).

Panke, S. (2023). Workshop 6: AI in education: Course design Institute. In *EdMedia+ innovate learning* (pp. 1362–1367). Association for the Advancement of Computing in Education (AACE).

Pasko, K., Arigo, D., et al. (2021). The roles of social comparison orientation and regulatory focus in college students' responses to fitspiration posts on social media: Cross-sectional study. *JMIR Mental Health, 8*(9), Article e26204, 2021.

Pasquini, L. A., & Eaton, P. W. (2021). Being/becoming professional online: Wayfinding through networked practices and digital experiences. *New Media & Society, 23*(5), 939–959, 2021.

Pataranutaporn, P., Leong, J., Danry, V., Lawson, A. P., Maes, P., & Sra, M. (2022). AI-generated virtual instructors based on liked or admired people can improve motivation and foster positive emotions for learning. In *2022 IEEE frontiers in education conference (FIE)* (pp. 1–9). IEEE.

Pérez-Álvarez, R., Maldonado-Mahauad, J. J., Sapunar-Opazo, D., & Pérez-Sanagustín, M. (2017). NoteMyProgress: A tool to support learners' self-regulated learning strategies in mooc environments. In *Data driven approaches in digital education: 12th European conference on technology enhanced learning, EC-TEL 2017, tallinn, Estonia, september 12–15, 2017, proceedings 12* (pp. 460–466). Springer.

Ratcliff, R., & McKoon, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 754, 1995.

Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching, 52*(1), 1–39, 2019.

Reinhold, F., Schons, C., Scheuerer, S., Gritzmann, P., Richter-Gebert, J., & Reiss, K. (2021). Students' coping with the self-regulatory demand of crisis-driven digitalization in university mathematics instruction: Do motivational and emotional orientations make a difference? *Computers in Human Behavior, 120*(2021), Article 106732.

Roberts, C., Moss, B., Wass, V., Sarangi, S., & Jones, R. (2005). Misunderstandings: A qualitative study of primary care consultations in multilingual settings, and educational implications. *Medical Education, 39*(5), 465–475, 2005.

Rochlin, M. (1982). Sexual orientation of the therapists and therapeutic effectiveness with gay clients. *Journal of Homosexuality, 7*(2–3), 21–29, 1982.

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7*(1), 31–44, 1973.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 352, 1989.

Schokker, M. C., Keers, J. C., Bouma, J., Links, T. P., Sanderman, R., Wolffenbuttel, B. H. R., et al. (2010). The impact of social comparison information on motivation in patients with diabetes as a function of regulatory focus and self-efficacy. *Health Psychology, 29*(4), 438, 2010.

Schouten, A. P., Valkenburg, P. M., & Peter, J. (2007). Precursors and underlying processes of adolescents' online self-disclosure: Developing and testing an "Internet-attribute-perception" model. *Media Psychology, 10*(2), 292–315, 2007.

Sheldon, S., Kathleen Pichora-Fuller, M., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America, 123*(1), 489–499, 2008.

Shepherd, D. A., & Cardon, M. S. (2009). Negative emotional reactions to project failure and the self-compassion to learn from the experience. *Journal of Management Studies, 46*(6), 923–949, 2009.

Sonnenschein, S., & Whitehurst, G. J. (1980). The development of communication: When a bad model makes a good teacher. *Journal of Experimental Child Psychology, 29*(3), 371–390, 1980.

Spaans, I., de Kleijn, R., Seeleman, C., & Dilaver, G. (2023). 'A role model is like a mosaic': Reimagining URiM students' role models in medical school. *BMC Medical Education, 23*(1), 396, 2023.

Stutzman, F., & Hartzog, W. (2012). Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 769–778).

Suls, J., & Wheeler, L. (2013). *Handbook of social comparison: Theory and research*. Springer Science & Business Media.

Szpyra-Kozłowska, J., & Szpyra, J. (2014). Pronunciation in EFL instruction. *Multilingual Matters, 82*.

Szyszka, M. (2017). *Pronunciation learning strategies and language anxiety* (Vol. 10). Switzerland: Springer (2017), 978–3.

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education, 89*, 53–74, 2015.

Teachey, D. T., Devidas, M., Wood, B. L., Chen, Z., Hayashi, R. J., Hermiston, M. L., et al. (2022). Children's oncology group trial AALL1231: A phase III clinical trial testing bortezomib in newly diagnosed T-cell acute lymphoblastic leukemia and lymphoma. *Journal of Clinical Oncology, 40*(19), 2106–2118, 2022.

Tory Higgins, E., Friedman, R. S., Harlow, R. E., Chen Idson, L., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology, 31*(1), 3–23, 2001.

Wan, F. (2021). Statistical analysis of two arm randomized pre-post designs with one post-treatment measurement. *BMC Medical Research Methodology, 21*(1), 1–16, 2021.

Wang, X., Zeng, H., Wang, Y., Wu, A., Sun, Z., Ma, X., et al. (2020). Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–12).

Watling, C., Driessen, E., van der Vleuten, C. P. M., Vanstone, M., & Lingard, L. (2012). Understanding responses to feedback: The potential and limitations of regulatory focus theory. *Medical Education, 46*(6), 593–603, 2012.

Weisman, W. D., & Peña, J. F. (2021). Face the uncanny: The effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 182–187, 2021.

Welker, C., France, D., Henty, A., & Wheatley, T. (2020). *Trading faces: Complete AI face doubles avoid the uncanny valley*, 2020.

Williams, A. M., & Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science, 24*(3), 283–307, 2005.

Wolff, L. S., Acevedo-Garcia, D., Subramanian, S. V., Weber, D., & Kawachi, I. (2010). Subjective social status, a new measure in health disparities research: Do race/ethnicity and choice of referent group matter? *Journal of Health Psychology, 15*(4), 560–574, 2010.

Wong, K. P., Lai, C. Y. Y., & Qin, J. (2023). Systematic review and meta-analysis of randomised controlled trials for evaluating the effectiveness of virtual reality therapy for social anxiety disorder. *Journal of Affective Disorders, 333*(2023), 353–364.

Wu, W., Zhang, B., Li, S., & Liu, H. (2022). Exploring factors of the willingness to accept AI-assisted learning environments: An empirical investigation based on the UTAUT model and perceived risk theory. *Frontiers in Psychology, 13*(2022), Article 870777.

Zhang, J. (2016). Can MOOCs be interesting to students? An experimental investigation from regulatory focus perspective. *Computers & Education, 95*(2016), 340–351.

Zhang, L., Jiang, L., Washington, N., Liu, A. A., Shao, J., Fourney, A., et al. (2021). Social media through voice: Synthesized voice qualities and self-presentation. In *Proceedings of the ACM on human-computer interaction 5, CSCW1 (2021)* (pp. 1–21).

Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S., & Cosley, D. (2013). The many faces of Facebook: Experiencing social media as performance, exhibition, and personal archive. *Proceedings of the SIGCHI conference on human factors in computing systems*, 1–10.

Zhao, T., Ye, L., Hu, Z., & Lian, X. (2024). Exploring the impact of positive reappraisal on self-regulated learning in MOOCs: The mediating roles of control and value appraisals and positive emotion. *Computers in Human Behavior, 152*(2024), Article 108070.

Zhou, H., Fujimoto, Y., Kanbara, M., & Kato, H. (2020). Effectiveness of virtual reality playback in public speaking training. In *Companion publication of the 2020 international conference on multimodal interaction* (pp. 462–466).