Suen, H. Y., & Hung, K. E. (2025). Enhancing learner affective engagement: The impact of instructor emotional expressions and vocal charisma in asynchronous video-based online learning. *Education and Information Technologies, 30*, 4033–4060. https://doi.org/10.1007/s10639-024-12956-w

Enhancing Learner Affective Engagement: The Impact of Instructor Emotional Expressions and Vocal Charisma in Asynchronous Video-Based Online Learning

Hung-Yue Suen[a]*, Kuo-En Hung[a],

[a] Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei City, Taiwan. R.O.C.

Abstract

In the rapidly evolving landscape of higher education and adult learning, asynchronous video-based online learning has not only become the new norm but has also emerged as the cornerstone of instructional delivery for Massive Open Online Courses (MOOCs). Despite its widespread adoption, this learning mode confronts a critical challenge: the inherent lack of social presence, posing a significant risk of diminishing learner affective engagement and, consequently, jeopardizing the efficacy of learning outcomes. Addressing this pressing issue, our study conducted a comprehensive analysis of 240 instructional videos from 240 distinct instructors on a MOOC platform, supplemented by 845 post-course learner feedback surveys from a diverse cohort of college students and adult learners. Using deep learning and statistical analysis, the research revealed that the on-screen presence of instructors does not inherently affect students' affective engagement. The study revealed that learners' affective engagement is affected by distinct combinations of the instructor's facial and paraverbal expressions, including happiness, surprise, and anger, which vary depending on whether the instructor is visible. The discovery that vocal attractiveness is a pivotal element in enhancing learners' affective engagement with instructional videos marks a paradigm shift in our understanding of digital andragogy and heutagogy. This study propels academic discourse by illuminating the critical role of instructor non-verbal cues in establishing social presence and facilitating emotional contagion within asynchronous video-based online learning but also provides

educators and content creators with empirically-backed techniques to revolutionize video instruction and amplify affective engagement.

Keywords: Affective computing; E-learning; Facial expressions; Nonverbal communication; Paraverbal communication; Voice quality

Suen, H. Y., & Hung, K. E. (2025). Enhancing learner affective engagement: The impact of instructor emotional expressions and vocal charisma in asynchronous video-based online learning. *Education and Information Technologies, 30*, 4033–4060. https://doi.org/10.1007/s10639-024-12956-w

1. Introduction

The global coronavirus disease (COVID-19) pandemic has necessitated a rapid and widespread shift to digital learning platforms for instructors and learners across public and private educational institutions. This transition from traditional face-to-face environments may not be a temporary adaptation but could represent a permanent shift to the new normal in the post-COVID era (Lockee, 2021). Although digital learning first emerged in the 1990s, the sudden and widespread transition to online modes has introduced challenges, including maintaining student engagement and adapting teaching methodologies for instructors to effectively engage online learners throughout the course (Martin et al., 2020a).

Online learning encompasses various formats, broadly classified into asynchronous and synchronous modes. Asynchronous learning allows students to engage with pre-recorded videos and materials at their convenience, without live interaction with instructors. In contrast, synchronous learning involves real-time, virtual interactions between students, instructors, and peers using digital platforms (Martin et al., 2020b).

Asynchronous video-based learning is perceived by learners as more focused on content delivery and less on learner-instructor interaction than synchronous learning. Despite this, it has become a dominant modality in education. It is valued for the flexibility and convenience it offers instructors and learners through pre-recorded lectures, a feature increasingly highlighted in recent literature (Hodges et al., 2020).

A comparative study by Garcia and Yousef (2023) suggests that asynchronous video-based online learning provides greater flexibility than synchronous methods, offering the advantage of learning 'anytime and anywhere' without constraints. This approach is also the mainstay of global Massive Open Online Courses (MOOCs) (Duan, 2022).

However, this approach can increase learner distractions and lack of dynamic interaction between instructors and students, presenting a significant educational challenge. Unlike

synchronous online conferencing or traditional classrooms, where instructors can immediately respond to learners' reactions and adapt their teaching methods, asynchronous video-based learning lacks such real-time interactivity (Yuyun, 2023). This absence of dynamic feedback may reduce learner engagement.

Engagement is crucial as it is directly associated with learners' ongoing participation (Zhao & Khan, 2022), learning achievements (Lin et al., 2019), and overall academic success (Özhan & Kocadere, 2020). Moreover, enhanced engagement improves motivation, satisfaction, and performance and helps mitigate the sense of social isolation often experienced in online learning environments (Martin & Bolliger, 2018).

Learning engagement is a multifaceted construct that includes three distinct yet interconnected dimensions. Affective engagement is centered on learners' emotional responses, encompassing their feelings, attitudes, and values, influencing their motivation and connection to the learning material. Behavioral engagement refers to observable actions, such as participation in class, completion of assignments, and general involvement in academic tasks. Meanwhile, cognitive engagement focuses on the mental effort learners invest, characterized by using deep learning strategies, critical thinking, and self-regulation (Walker & Koralesky, 2021). Affective engagement, in particular, poses a challenge in video-based learning due to the reduced learner-instructor interactions compared to face-to-face and synchronous settings (Daher et al., 2021; Garcia & Yousef, 2023). While instructors have been key to nurturing affective engagement in classrooms (Ainley, 2006), the potential for replicating this engagement in online video-based courses is a promising yet under-researched area (Walker & Koralesky, 2021).

Recognizing this gap, there is a pressing need for more comprehensive research on how instructors can maintain learners' affective engagement through pre-recorded video lectures, a significant topic within computer-mediated education (Henrie et al., 2015; Lytvyn et al., 2021). Research suggests that instructors can use non-verbal communication, such as facial

and vocal expressions, to affect learners' emotions and engagement in physical classrooms (Bambaeeroo & Shokrpour, 2017). However, the challenge intensifies in asynchronous video-based learning, where instructors cannot easily gauge and respond to learners' reactions (Mershad & Said, 2022).

In the evolving domain of online learning, the importance of instructors' social presence and emotional expressions is increasingly recognized by researchers and practitioners (Chiu, 2022). Through facial and paraverbal cues, these expressions significantly impact learners' affective engagement, which is crucial for positive learning outcomes (Fredricks et al., 2004). While the impact of an instructor's social presence and emotional expressions is well-documented in traditional classrooms and synchronous online environments (Keller & Eva, 2021), their relevance in asynchronous video-based learning remains an open question.

This question highlights a notable gap in current research, leaving the effects of an instructor's social presence and emotional expressions in video-based learning largely unexplored (Bhatti et al., 2021). The need for focused research in this area is echoed by recent studies that emphasize the urgency of this investigative path (Chiu, 2022; Collins et al., 2019).

Given the limited communication channels in asynchronous video-based learning, examining how instructors can effectively foster affective engagement is crucial. Beyond the on-screen presence, instructors' facial and paraverbal expressions are critical in influencing learners' affective engagement in asynchronous video-based online learning. This study explores the complex interplay of emotional expressions in video instruction materials, seeking to develop instructional techniques that enhance learners' affective engagement. It focuses on extrinsic emotions, which are observable by others and may not always correspond with an individual's intrinsic emotions (Nozaki & Mikolajczak, 2020). Therefore, this paper addresses the following research questions in higher education contexts:

Q1: Does an instructor's visibility influence learners' affective engagement in asynchronous

video-based online learning?

Q2: Do an instructor's non-verbal emotional expressions influence learners' engagement in asynchronous video-based online learning? Specifically, which has a more significant effect: facial expressions or paraverbal cues?

2. Literature Review

In the evolving domain of online education, the role of instructors' social presence and nonverbal communication has been established as a crucial determinant of learners' affective engagement, a finding robustly supported by extensive research (Dixson, 2010; Wang, 2022). This form of engagement is primarily influenced by instructors' expressive behaviors, serving as a medium for emotional conveyance and impacting learners' engagement, focus, memory retention, and intrinsic motivation (Mottet et al., 2006; Horan et al., 2012; Fanselow, 2018; Iskrenovic-Momcilovic, 2018).

Building upon these insights, the Community of Inquiry (CoI) framework posits that genuine emotional expressions by instructors in digital formats can significantly deepen learner engagement (Borup et al., 2012). Additionally, Emotional Contagion Theory explains how emotions are subconsciously transmitted between individuals, often leading to shared emotional states. This theory highlights the impact of such emotional exchanges on the learning experience, suggesting that the emotional states of instructors can directly influence the engagement and motivation of learners in an online setting (Hatfield et al., 1993; Liu et al., 2019; Robinson, 2013; Wang et al., 2019, 2022).

Despite these insights, a significant gap persists in the literature concerning asynchronous video-based online learning. Although increasingly prevalent, this context remains distinct from the synchronous interactions characteristic of traditional classrooms. This paper endeavors to bridge this gap by delving into the unique dynamics of asynchronous video-

based online learning. Through an exhaustive literature review, we aim to illuminate the potential impact of instructors' social presence and their capacity for emotional contagion on learners' affective engagement, thereby enriching the discourse on online andragogy and heutagogy.

2.1 Instructor Visibility and Affective Engagement in Asynchronous Video-Based Online Learning

Research indicates that the visibility of an instructor's face can enhance the social presence in online learning (Henderson & Schroeder, 2021). Social presence, the sense of interaction among class participants (Wise et al., 2004), is a key element of the CoI framework (Akyol & Garrison, 2008), which is instrumental in creating engaging online learning experiences. Instructors are advised to foster a sense of authenticity and presence, primarily through immediacy and intimacy, to strengthen their social presence as perceived by learners (Richardson & Lowenthal, 2017).

Chew (2022) discovered that in synchronous online learning, seeing the instructor's face can enhance learners' perceptions of the instructor's responsiveness and care, thus increasing affective engagement similar to in-person interactions. However, this impact in asynchronous video-based learning is not as well-defined. While an instructor's face on-screen can make the learning experience feel more personal, it also has the potential to divert attention from the main content (Mayer, 2014). Some studies indicate that learners may ignore the instructor's face when concentrating on the content, especially if it is not directly relevant to the learning task, to avoid the split-attention effect (Rop et al., 2018).

However, Henderson and Schroeder (2021) reviewed 395 studies and found minimal evidence supporting the split-attention effect. Their research convincingly demonstrated that an instructor's on-screen presence significantly strengthens social presence, vital for affective engagement. The on-screen presence of instructors makes them appear more immediate and

relatable, thus deepening the intimacy of the learning experience. This effect can lead to levels of engagement in asynchronous video-based online learning comparable to those in real-time interactions (Borup et al., 2012). Further research supports the advantages of visible instructors in asynchronous learning, showing that such visibility not only grabs learners' attention (Wermeskerken et al., 2018) but also boosts their satisfaction (Yuan et al., 2021; Wang et al., 2020) and perceived learning (Kizilcec et al., 2015), without negatively affecting learning outcomes. This likely accounts for the increasing trend of instructors incorporating their images into video content (Henderson & Schroeder, 2021). Therefore, we propose the first hypothesis for Q1:

H1: In asynchronous video-based online learning, learners show higher affective engagement when instructors' faces are visible than when they are not.

2.2 Instructor Facial Expression and Affective Engagement in Asynchronous Video-Based Online Learning

An empirical study demonstrated that instructors' positive emotions favor student engagement in both Chinese and U.S. cultures (Zhang & Zhang, 2013). However, the study did not investigate the specific sensing modalities and cues learners use to discern instructors' emotional states. Instructor visibility is pivotal in establishing a social presence, a concept anchored in the CoI framework, where facial expressions are crucial to conveying immediacy and intimacy (Richardson & Lowenthal, 2017). By leveraging appropriate social cues, an instructor can enhance a sense of social presence, thereby boosting learner engagement in online settings (Henderson & Schroeder, 2021).

However, establishing a social presence alone does not automatically translate to learner engagement. The nature, frequency, and quality of social cues during instructor-learner interactions significantly influence engagement (Wise et al., 2004; Richardson & Lowenthal, 2017). These cues include facial and vocal signals, such as expressions and voice modulation,

which foster engagement (Chew, 2022).

According to Social Cognitive Theory, individuals learn by observing others, with the process being heavily influenced by the emotional cues exhibited by the observed (Bandura, 1986). Therefore, when instructors display expressions that reflect understanding, enthusiasm, and positivity, they can instill similar feelings in learners, prompting deeper engagement with the content. Interpersonal Neurobiology Theory suggests that human relationships and social connections can affect brain function and, by extension, an individual's mental well-being (Siegel, 2020). Positive facial expressions from instructors, in this context, can nurture a supportive relationship with learners, fostering a conducive emotional state and deeper engagement with the material.

Theonas et al. (2007) highlighted that expressive facial cues, such as smiling, can boost learners' interest, motivation, and performance in asynchronous learning environments that use pre-recorded videos. Further, research by Wang (2022) and Wang et al. (2019) emphasized the significance of an instructor's facial expressions in conveying emotions, which can stimulate learners' motivation and satisfaction while reducing cognitive load in asynchronous video-based online learning.

While the universality of facial expressions of emotion is a subject of debate (Jack et al., 2012), seven primary expressions—anger, disgust, fear, happiness, sadness, surprise, and neutrality—are widely recognized across different cultures as representing distinct emotional states (Ekman, 1992; 1994). These emotions are commonly categorized within Russell's (1980) Circumplex Model of Affect, which differentiates them along two dimensions: valence (pleasant versus unpleasant feelings) and arousal (calm to excited). This dual-axis framework, applicable across diverse cultural contexts (Russell, 2003), suggests that emotions characterized by distinct valence and high arousal likely have a more substantial emotional impact.

Liu et al. (2022) identified six distinct facial expressions, each associated with specific levels

of arousal and valence. According to their research, happiness and surprise correspond to positive valence and high arousal, whereas anger is linked to negative valence and high arousal. Sadness is described as having negative valence and low arousal, while fear and disgust are characterized by negative valence but high arousal.

Juslin and Laukka (2003) discovered that emotional arousal can influence the accuracy with which these facial expressions and their corresponding valences are recognized. Stimuli that are both positive and negative with high arousal levels can activate learners' emotional states (Törmänen et al., 2021), indicating that emotions like happiness, surprise, and anger, which are characterized by distinct valence and high arousal, have a more substantial emotional impact.

Research by Sohriani et al. (2018) revealed that expressions of happiness and surprise can engage learners, elevating their morale and self-confidence, whereas expressions of anger can impede concentration. Given the significant impact of instructor facial expressions on learners' perceptions in synchronous online learning and traditional classrooms (Chew, 2022), it is plausible to expect a similar positive effect on learners' affective engagement in asynchronous video-based online learning. Therefore, we propose the following hypothesis for Q2:

H2: In asynchronous video-based online learning, instructors' frequent positive facial expressions with high arousal correlate positively with learners' affective engagement, whereas frequent negative expressions with high arousal correlate negatively.

2.3 Instructor Paraverbal Expression and Affective Engagement in Asynchronous Video-Based Online Learning

Kraus (2017) conducted five experiments and found that people perceive others' emotions more accurately through vocal cues than through facial cues alone or a combination of both, due to humans' limited cognitive capacity for decoding facial expressions and processing

multiple sensory modalities. Furthermore, individuals can discern emotions via paraverbal cues (i.e., 'how it is said') without relying on visual information (Kraus, 2017; Simon-Thomas et al., 2009).

Psychologists have identified that speakers communicate emotions through paraverbal cues such as pitch, loudness, speech rate, and voice quality (Laukka, 2017; Juslin & Laukka, 2003). Certain combinations of these paraverbal cues can elicit specific perceived emotions (Sauter et al., 2010). Gunes et al. (2011) suggest that paraverbal expressions correspond to seven universal emotions, akin to facial expressions. Among these, Happiness and Surprise are linked to high positive valence and arousal, while Anger is associated with high negative valence and arousal. Given that paraverbal emotions with distinct valence and heightened arousal have a greater impact, happiness, surprise, and anger exert a more significant emotional influence than other emotional categories.

Paralinguistic theory asserts that the non-verbal aspects of speech convey substantial information and shape listeners' perceptions and responses (Trager, 1958). This implies that an instructor's vocal nuances can create a more engaging learning atmosphere. Self-determination theory posits that a teacher's warm and understanding voice can foster relatedness, thereby enhancing students' intrinsic motivation and engagement (Ryan & Deci, 2000). Voice and Emotion Theory emphasizes that paraverbal cues carry emotional content that can affect listeners' emotional states (Juslin & Laukka, 2003). Therefore, an instructor’s paraverbal expressions, imbued with appropriate emotional tones, can cultivate a positive emotional environment, promoting affective engagement among learners.

Chew’s (2022) qualitative study highlighted the significance of an instructor’s paraverbal expressions as essential teaching skills for engaging learners in synchronous online learning. A recent empirical study showed that an instructor’s paraverbal expressions positively influenced learners' social presence, satisfaction, and enjoyment while reducing boredom in asynchronous video-based online learning (Yuan et al., 2021). Consequently, we propose the

following hypothesis for Q2:

H3: In asynchronous video-based online learning, instructors' frequent positive paraverbal expressions with high arousal positively correlate with learners' affective engagement, whereas frequent negative expressions with high arousal correlate negatively.

2.4 Instructor Paraverbal Versus Facial Expressions in Asynchronous Video-Based Online Learning

In asynchronous video-based online learning, instructors may not always be visible on screen, and learners' focus on visual content can vary. Consequently, the time learners devote to 'listening' often exceeds that spent on 'watching' the video content (Wang et al., 2022). Social Presence Theory suggests that when instructors are intermittently visible, learners rely more on auditory cues to establish a sense of presence (Short et al., 1976). Hence, instructors' paraverbal expressions become crucial in determining the social presence and emotional engagement.

According to the Dual Coding Theory, individuals process information using both verbal and non-verbal codes (Paivio, 1990). In asynchronous video-based online learning, where real-time interaction is absent and visual cues are limited, learners may depend more on auditory cues. In such scenarios, the paraverbal cues of instructors become essential in creating emotional connections and guiding learners. An instructor's voice can demystify complex concepts (Mayer et al., 2014), enhance engagement in an otherwise solitary learning environment (Bolliger & Martin, 2018), and affect learners' responses to feedback (Narciss, 2008). Therefore, paraverbal cues can influence learner engagement in these settings more than facial cues.

Cognitive Load Theory highlights the limitations of cognitive resources (Sweller, 1988). An engaging voice can prompt learners to allocate more cognitive resources to auditory information, thus enhancing the learning experience. In such contexts, an instructor's

paraverbal expressions may impact learners' affective engagement more than facial emotive expressions. Based on these considerations, we propose the following hypothesis for Q2:

H4: In asynchronous video-based online learning, an instructor's paraverbal expressions affect learners' affective engagement more than their facial expressions.

2.5 'Vocal Charisma Effect' in Asynchronous Video-based Online Learning

Nordström and colleagues (2017) found in a cross-cultural study that audiences from different countries tend to subjectively infer and experience similar emotions through specific paraverbal cues of performers, indicating a potential universality in these cues. In speech communication, 'voice quality' refers to these paraverbal cues that can elicit positive emotions. This quality is a blend of vocal characteristics—like timbre, pitch, and modulation—that create an affective perception that is greater than the sum of its parts and cannot be pinpointed to specific affective qualities (Gobl & Chasaide, 2003). Cavanagh et al. (2014) noted that the voice quality of preservice teachers significantly improved students' evaluations of their oral presentations in pre-recorded videos.

Xu et al. (2013) contend that pleasant voice quality can profoundly affect audiences' perceptions of 'vocal attractiveness.' A voice that triggers positive emotions is naturally perceived as more attractive, engaging listeners and creating an emotionally stimulating auditory experience.

According to the Social Information Processing Theory, vocal attractiveness conveyed through paraverbal cues, can communicate emotional subtleties and personality traits that resonate with audiences, enhancing connections and engagement in computer-mediated environments (Walther, 1992). A warm and appealing voice can generate a positive first impression and boost affective engagement by inducing positive emotions and promoting a receptive and cooperative response to the content.

This concept is supported by the 'halo effect' in psychology (Nisbett & Wilson, 1977), where a positive perception in one area can influence overall impressions, potentially leading to a favorable initial view of the instructor and a more engaging learning experience. Shang and Liu (2022) mark the crucial role of vocal attractiveness in social preferences. Their research indicates that speakers with attractive voices foster cooperative behaviors and increase emotional valence and investment decisions among listeners using headphones.

Given the significant influence of an instructor's 'vocal charisma' on learners' affective engagement—possibly exceeding that of other paraverbal and facial emotive expressions—we propose the following hypothesis for Q2:

H5: In asynchronous video-based online learning, an instructor's vocal attractiveness affects learners' affective engagement more than the other paraverbal and facial emotive expressions.

3. Method

3.1 Research Design

To explore the research questions, this study utilized a correlational design with the following elements: Firstly, the design enabled precise control over variables by using authentic instructional videos in Chinese from a renowned MOOC platform in Taiwan, "Hahow", secured through an industry-academia partnership, ensuring the study's relevance. Secondly, it allowed for the direct observation of cause-and-effect relationships, a crucial element in understanding the dynamics of asynchronous video-based online learning. The standardized approach of this design also ensured the comparability and reliability of the results among different participants. Additionally, this methodology facilitated the collection of measurable data on learners' affective engagement through self-assessment scores, which is essential for comprehensive statistical analysis. Finally, using de-identified pre-recorded course videos and learners' self-reported engagement scores from the MOOC platform fully complied with

Research Ethics Committee (REC) standards (Reference No: 202305HS089). This approach provided access to varied content and diverse participant demographics, crucial for a holistic understanding of the research questions. Therefore, this design was deemed ideal for assessing the influence of instructors' non-verbal cues on learner engagement in an asynchronous online video learning environment.

To ensure the smooth execution and feasibility of the research, a pilot test was carried out, separate from the main study, using participants and materials not included in the formal research. Five actual asynchronous online instructors and 15 graduate students were invited to participate in this preliminary phase. The instructors provided pre-recorded lectures, with some being on screen (visible) and others not on screen (invisible) in the videos, ranging from 8 to 28 minutes, appropriate for individuals with a high school education level. Following a randomized process, three students were selected to watch a video and complete the study's questionnaire, which included manipulation check questions. The feedback and results from this pilot test were instrumental in refining the research procedures and questionnaire, setting the stage for the formal field study.

As illustrated in Figure 1, the formal study's participants in asynchronous video-based online learning were divided into two groups based on instructors' visibility in the videos: the visible group and the invisible group. Deep learning (DL) techniques were used to identify the probability (0-100%) of six paraverbal emotive expressions (happiness, sadness, surprise, anger, fear, and neutral) and to assess vocal attractiveness (scored from 1 to 5) in both groups. These variables were then analyzed and compared to determine their impact on learners' self-reported affective engagement in post-class surveys. Additionally, DL techniques identified the probability (0-100%) for the visible group of seven facial emotive expressions (happy, sad, surprised, angry, fearful, disgusted, and neutral). The influence of these facial expressions on learners' affective engagement was further analyzed and compared. Learners in both the visible and invisible groups rated their self-reported affective engagement (scored from 1 to

5) during the post-class feedback for the video instructors, with each instructor being evaluated by at least three learners.

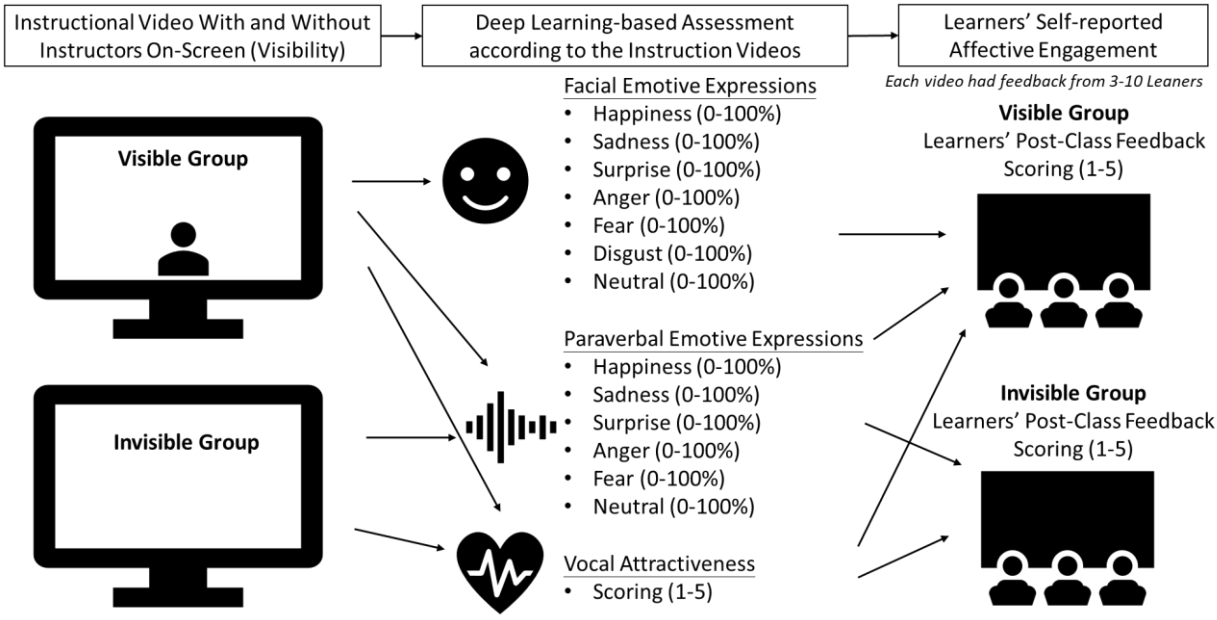

Figure 1. Research design

3.2 Research Participants

We gathered 240 valid videos from unique instructors. A valid video is one that DL techniques can analyze to identify the instructor's facial and paraverbal emotive expressions. These videos were all selected by the platform, featuring the latest uploads with prior consent from the instructors. The sample demographics included 174 male instructors (72%) and 66 female instructors (28%). Course content primarily focused on the social sciences (197 courses, 82%) and science (43 courses, 18%). The lectures varied in length, ranging from 8 to 28 minutes, with an average duration of 15 minutes. Of these, 112 instructors appeared on-screen, categorized as the visible group (47%), while the remaining 128 were in the invisible group (53%), not appearing on-screen.

For these 240 courses, the sample of learners was obtained from the platform. These learners had registered, completed the courses, and were invited to participate in a survey. They filled out feedback surveys and knew their responses would be used for this research, resulting in 1051 responses from at least three students per course. After discarding 206 responses that failed the manipulation check, 845 valid responses were analyzed. In the visible group, 393 valid responses were collected for 112 videos. In the invisible group, 452 valid responses were collected for 128 videos. Each video had feedback from 3-10 learners on average.

The respondents, aged 18 to 47, were mainly between 26-30 years (29%) and 31-35 years (27%). Course distribution among respondents was 63% in social sciences and 37% in science. Female students comprised 59% of the survey participants, with males accounting for 41%. Regarding educational background, most had a bachelor's degree (61%), followed by master's degree holders (37%), Ph.D. holders (1%), and less than 1% with only high school education.

3.3 Data Collection Tools

Working in collaboration with the platform, we employed advanced software utilizing DL and computer vision technologies to anonymize the instructional videos. This allowed us to determine instructors' visibility during the lessons and assess their facial and vocal emotional expressions.

The platform provided pre-recorded videos in 1280 x 720 resolution. In some of these, instructors were intermittently visible, with their faces appearing for about a quarter of the total video duration. In contrast, in others, the instructors' faces were not shown at all. The duration of the videos varied. Our analysis targeted facial and paraverbal expressions in the segments where the instructors' faces were visible. We focused solely on analyzing the paraverbal expressions for the video portions without visible faces.

In addition to directly extracting instructors' facial and paraverbal expression features from instructional videos, we aimed to deepen our understanding of learners' affective engagement

with the videos and their perception of the instructors' vocal attractiveness. To this end, we enriched the platform's existing learner feedback mechanism with additional survey questions. We introduced five questions focused on affective engagement and two on vocal attractiveness perception into the after-class opinion questionnaire, which is presented at the end of each video. Recognizing that completing the questionnaire was voluntary, we extended the duration of the learner feedback survey to two months. This extended timeframe was crucial to ensure that we could collect responses from a sufficient number of participants (at least three volunteers per video) and secure valid questionnaires for each video.

3.3.1 Instructor's Visibility

In this study, the visibility of the instructor's face in the video is treated as a categorical variable. It is coded as 2 when the instructor's face is shown and 1 when it is not. In our data, instructors intermittently show their faces throughout the video, with face-visible time accounting for approximately 30-50% of the total video duration.

3.3.2 Instructor's Facial Expression

Following Goodfellow et al. (2013) and Nguyen et al. (2022), we harnessed the FER-2013 dataset from Kaggle to train a model for emotion detection in images using a DL Convolutional Neural Network (CNN) approach. This dataset comprises 35,887 grayscale images (48 x 48 pixels), each labeled with seven emotional states: happy, sad, surprised, angry, fearful, disgusted, and neutral. Our model was developed using the TensorFlow framework and the Keras library. It comprises four Conv2D layers, two dense layers, and an output layer. The model was compiled using the Adam optimizer and a categorical cross-entropy loss function. To counteract overfitting, we integrated the EarlyStopping method and Dropout layers. Optimal model weights were preserved using the ModelCheckpoint method, and the ReduceLROnPlateau adjusted the learning rate if validation loss remained stagnant. After

training over 48 epochs, the model's performance metrics were gauged in accuracy and loss. The dataset was split into training, validation, and testing segments in an 80-10-10 ratio. The primary objective was to cultivate models adept at autonomously identifying the seven emotions in each image frame. Our proposed model achieved an average of 70% testing set accuracy, closely approaching the 74% benchmark for image-based emotion recognition (see Nguyen et al., 2022).

For this study, OpenCV was employed to capture one frame from each video at half-second intervals, given that a half-second facial macro-expression represents the discernible limit for the human eye (Warren et al., 2009). With the model trained earlier, we gauged the probability values (0-100%) for seven emotions across each frame. The mean probability values of the seven emotions across all frames in each video were computed, serving as a metric to gauge the perceived emotional expressions of instructors in the videos.

3.3.3 Instructor's Paraverbal Emotive Expression

To identify the paraverbal emotive expressions of instructors, we adopted the methodology proposed by Ye et al. (2022). This approach combined CNN with Long Short-Term Memory (LSTM) and utilized the CASIA dataset (Bao et al., 2014) to develop the recognition model. Audio features, including Mel-frequency cepstral coefficients (MFCCs), Zero Crossing Rate (ZCR), Chroma, Root Mean Square (RMS), and Spectrogram, were extracted from each video to serve as the voice input for emotion recognition. The spectrum covered 20Hz and 20kHz, the default range for human hearing. Each audio segment spanned five seconds with a two-second overlap to capture context from preceding utterances. Segments with a confidence level below 0.7 were excluded from the analysis due to potential interference from noise or music (Allison et al., 2022). The dataset, recorded by professional voice actors under specific scenarios, was divided into training, validation, and testing sets at an 80-10-10 ratio. It encompasses six distinct emotions listeners perceive: happiness, sadness, surprise, anger, fear,

and neutral. Each recording provides a perceived probability (ranging from 0-100%) for these six emotions. Our proposed model achieved an average testing set accuracy of 81%, closely approaching the 84% benchmark for video-based emotion recognition (see Ye et al., 2022).

3.3.4 Instructor's Vocal Attractiveness

Given the absence of a dataset for vocal attractiveness, we constructed one based on Yuan et al.'s (2021) 7-point Likert scale, which ranges from 1 (totally disagree) to 7 (totally agree), comprising two items. A sample question from this scale is "I find the instructor's voice mellow and attractive." After course completion in this study, all learners rated the instructor's voice, yielding a Cronbach's α of 0.86 and an average intraclass correlation coefficient (ICC) of 0.68, indicating a high level of consistency in the learners' evaluations of vocal attractiveness. Adapting the approach used for recognizing paraverbal emotive expressions (Ye et al., 2022), we employed regression head to create an automated model for voice attractiveness, with the output ranging from 1 to 7 points. The model's test results showed an R-value of 0.89 ($p < .01$) and an RMSE of 0.24.

3.3.5 Learners' Affective Engagement

We adopted the Online Student Engagement Scale (Dixson, 2010) for emotional engagement, which consists of five items rated on a 5-point scale. This scale was chosen for its suitability to the online learning environment and its superior reliability (Cronbach's $\alpha = 0.91$) compared to other similar scales (Henrie et al., 2015). A sample item from the scale is "Putting forth effort."

3.3.6 Potential Control Variables

In this study, we conducted analyses across various courses, each differing in subject area (either Social Science = 1 or Science =2) and duration (measured in minutes). These factors

might influence learners' affective engagement (Akiha et al., 2018; Buckley et al., 2004). Moreover, the sex (Female = 1; Male =2) of the instructors could affect how learners perceive emotive expressions (Leung et al., 2018). Therefore, if subject areas, class durations, and instructors' gender significantly correlate with affective engagement, they would be considered covariates.

3.4 Procedure

After obtaining 240 de-identified instructional videos from this MOOC platform, we immediately classified the videos into visible and invisible groups. We then utilized the aforementioned data collection tools to extract the facial and paraverbal emotive expression features of these 240 instructors. This included the probability of six major emotions in the voice, scores of vocal attractiveness, and, for the visible group, the probability of seven major facial emotive expressions. Concurrently, through this platform, we obtained questionnaires assessing affective engagement from learners who had recently watched these 240 instructional videos, all within the last two months. Each video received responses from at least three students. By calculating the average score of all responses for each video, we derived the affective engagement score attributed to each instructor by their learners. The demographic information of both instructors and learners was also provided through this platform, ensuring compliance with privacy and security laws.

3.5 Data Analysis

Our study employed an independent T-test to test the impact of the binary nominal variable 'visibility' on the dependent variable of the affective engagement scale. When the groups did not exhibit a normal distribution, we supplemented this with the Mann-Whitney U test. This methodology facilitated an effective analysis of the influence of instructor visibility on learner engagement for H1.

To ensure that the items on the scale are consistent and that different learners have consistent views regarding their self-reported affective engagement in the same course, we conducted reliability analyses using Cronbach's α and ICC, respectively. Additionally, we used Pearson correlation analysis to understand the relationships between all the variables in this study and to determine which variables should be included as control variables and which should not. For testing H2 to H5, we utilized multiple regression analysis to examine the effects of the continuous scores of facial and paraverbal emotive expressions and vocal attractiveness on the dependent variable and to compare the impact of these independent variables on the dependent variable. This statistical method was selected because it assesses how the probability of each type of expression and the scores of vocal attractiveness, treated as independent variables, impact the overall affective engagement of learners.

4. Findings

4.1 The Scale's Reliability and Variable Interconnections

In this study, the Cronbach's α value for the learners' self-reported affective engagement was 0.85, and the mean ICC was 0.51, indicating moderate consistency. For the independent variables, all were assessed using outputs from DL models, except for the objective conditions related to instructor visibility. Based on the correlation analysis in Table 1, this study found that asynchronous instructional videos feature more female instructors in the social sciences, while the sciences have more male instructors. Instructors from the social sciences, as well as female instructors, are more inclined to show their faces in the videos. In contrast, courses with longer durations have a higher presence of female instructors, but they tend to appear on screen less frequently. Moreover, instructors exhibited more paraverbal emotions of happiness and surprise, with fewer instances of anger, leading to notably higher vocal attractiveness. Vocal attractiveness, coupled with the paraverbal emotions of happiness and surprise,

demonstrated a significant positive correlation with learners' affective engagement. When instructors conveyed paraverbal emotions of anger, their facial expressions also prominently reflected anger, and both exhibited a marked negative correlation with learners' affective engagement. Although there was no significant correlation between instructors' facial emotions of happiness and surprise and their corresponding paraverbal emotions, both facial emotive expressions were positively associated with learners' affective engagement.

Table 1. Pearson's Correlations

Variable	$\bar{x}$	s	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Subject[a]	1.82	0.38	1	2																		
2. Durations	15.26	3.07	8.00	28.00	-.10																	
3. Sex[b]	1.73	0.45	1	2	-.14*	.16*																
4. Visibility[c]	1.47	0.50	1	2	.20**	-.20**	-.02															
5. Facial-Happy	0.49	0.13	0.20	0.76	.05	.09	.05	—														
6. Facial-Surprise	0.12	0.05	0.10	0.41	.12	.00	-.04	—	.06													
7. Facial-Anger	0.28	0.07	0.00	0.20	.03	-.02	.05	—	-.14	-.19												
8. Facial-Sad	0.06	0.04	0.02	0.20	.01	.06	.09	—	.12	-.02	-.08											
9. Facial-Fear	0.16	0.08	0.00	0.29	-.06	-.07	-.01	—	.09	.04	-.17	-.04										
10. Facial-disgust	0.06	0.04	0.00	0.16	.14	.09	.06	—	-.12	-.11	.14	-.04	.08									
11. Facial-Neutral	0.73	0.14	0.48	0.97	.02	-.16	-.13	—	-.08	-.16	.15	-.02	-.08	-.04								
12. Paraverbal-Happy	0.16	0.05	0.06	0.29	.05	.04	.10	-.08	-.10	-.00	.08	-.01	-.08	.06	.06							
13. Paraverbal-Sad	0.13	0.08	0.02	0.28	-.04	-.09	-.03	.05	.14	-.05	.06	.65***	-.05	-.09	-.00	-.05						
14. Paraverbal-Surprise	0.20	0.10	0.05	0.43	.12	.01	.00	-.02	.07	.12	-.06	-.09	-.11	-.01	.07	.09	.01					
15. Paraverbal-Anger	0.07	0.03	0.02	0.15	-.02	.03	-.02	-.04	-.08	.02	.72***	-.11	-.11	.05	.12	-.01	-.05	-.06				
16. Paraverbal-Fear	0.14	0.07	0.03	0.26	.01	-.03	-.01	.01	.01	.08	.04	.02	.63***	.12	.07	-.11	.09	-.03	.01			
17. Paraverbal-Neutral	0.63	0.14	0.45	0.88	-.04	.06	.01	-.02	.06	-.18	.01	.12	.02	.06	-.08	-.08	.12	-.03	.01	.01		
18. Vocal Attractiveness	4.06	0.40	3.00	5.00	.08	.02	-.04	.11	-.08	.05	-.12	-.01	-.03	.13	-.02	.25***	.07	.34***	-.24***	-.01	-.02	
19. Affective Engagement	4.08	0.35	2.53	5.00	.08	.06	-.05	-.02	.33***	.22*	-.30**	-.05	.05	-.03	-.01	.33***	.04	.38***	-.32***	-.04	-.08	.74***

* $p < .05$, ** $p < .01$, *** $p < .001$
a Social Science = 2, Science = 1
b Male = 2, Female = 1
c On screen = 2, Off screen = 1; The variance in visibility is equal to 0 for facial emotive expressions denoted by dashes (—).

Using DL-based computer vision, we analyzed instances of instructor visibility in videos (n = 112), measuring their on-screen time against total playback ($\bar{x}$ = 0.43; s = 0.17). Correlational analysis showed no significant link between on-screen time proportion and variables of instructor expressions or learners' affective engagement.

4.2 The Visibility of Instructor and Affective Engagement

To determine whether the visibility of the instructor affected learners' affective engagement for H1, an Independent Samples T-Test was conducted, comparing groups where the instructor was visible with those where the instructor was invisible. The results, $t(238) = 0.31$, $p = 0.62$, showed no significant difference. The Visible Group had a mean score ($\bar{x}$) = 4.07 ($s = 0.41$, $n = 112$), similar to the Invisible Group's $\bar{x} = 4.08$ ($s = 0.29$, $n = 128$). Given the significant Shapiro-Wilk Test of Normality for both groups, indicating non-normal distribution, a Mann-Whitney U test was conducted, yielding $W = 6477.50$, $p = 0.10$. This suggests that the on-screen or off-screen of the instructor in the recorded video did not significantly influence learners' affective engagement, indicating that H1 was not supported.

4.3 The Facial and Paraverbal Expressions of Instructor and Affective Engagement

To test hypotheses H2-H5, we utilized two separate stepwise regression models. Model 1 was used to evaluate the impact of instructors' facial and paraverbal expressions on learners' affective engagement within the group where instructors were visible. For the group where instructors were invisible, we applied Model 2 to investigate the influence of various paraverbal expressions on learners' affective engagement. In both models, we initially conducted stepwise regressions to pinpoint the specific facial and paraverbal expression factors that significantly affected learners' affective engagement. The findings from these analyses are detailed in Table 2:

Table 2. Stepwise Regression Analysis Predicting Learners' Affective Engagement

Predictors	Model 1 (Visible) Adjusted R^2 = 0.81 (n = 112)	Model 2 (Invisible) Adjusted R^2 = 0.59 (n = 128)	Adjusted R^2 Change		95% CI (Lower / Upper)	
	Standardized Coefficients		Visible	Invisible	Visible	Invisible
Facial-Happy	0.37***	--	0.15***	--	0.94 / 1.49	--
Facial-Suprise	0.14**	--	0.03***	--	0.32 / 1.32	
Facial-Anger	-0.14**	--	0.02**	--	-2.20 / -0.53	--
Paraverbal-Happy	0.15**	0.16**	0.02**	0.03**	0.49 / 1.89	0.22 / 1.44
Paraverbal-Surprise	--	0.13*	--	0.01*	--	0.02 / 0.80
Paraverbal-Anger	--	-0.17**	--	0.05***	--	-2.88 / -0.50
Vocal Attractiveness	0.74***	0.59***	0.60***	0.51***	0.64 / 0.82	0.36 / 0.55

*p < .05, **p < .01, ***p < .001

Model 1 analyzed 112 instructional videos featuring visible instructors. Stepwise regression identified that facial emotive expressions significantly influenced learners' affective engagement. Results showed that instructors' frequent expressions of happiness and surprise, and infrequent expressions of anger, correlated with increased affective engagement. Of the seven assessed facial emotive expressions, happiness, surprise, and anger—the three with the highest arousal levels—significantly impacted learners' affective engagement, accounting for 20% of the variance, thus confirming H2.

The analysis suggests instructors can enhance learners' affective engagement by expressing paraverbal emotions like happiness, particularly when paired with vocal attractiveness. These paraverbal expressions alone accounted for 62% of the variance, much higher than the 20% from facial expressions. Vocal attractiveness was the most influential factor in the model, explaining 60% of the variance, with facial expressions of happiness contributing 15%. Facial expressions of surprise and anger, and paraverbal expressions of happiness each had a marginal contribution of about 2-3%. This indicates that paraverbal expressions, especially those involving an attractive voice and positive high-arousal expressions of happiness, significantly enhance learners' affective engagement. The other six paraverbal emotive expressions, including high-arousal emotions like anger and surprise, had no significant impact. Therefore, H3 was partially

supported, while H4 and H5 were fully supported when instructors were visible.

Model 2 analyzed 128 instructional videos without visible instructors. Stepwise regression identified the most significant paraverbal expressions for learners' affective engagement, in order of influence: vocal attractiveness (51%), paraverbal anger (5%), paraverbal happiness (3%), and paraverbal surprise (1%), totaling around 59% when rounded. This indicates that in scenarios without the instructor's visible presence, those with a more attractive voice, with higher frequencies of positive and high-arousal paraverbal emotions like happiness and surprise, and lower frequencies of negative emotions like anger, can significantly enhance learners' affective engagement. Thus, H3 was fully supported in scenarios where the instructor was invisible. Consistent with the results from Model 1, H4 and H5 were also fully validated.

5. Discussion

5.1 Instructor Visibility Exerts Minimal Influence

The field study provides robust evidence indicating that instructor visibility in asynchronous video-based online learning does not significantly affect the affective engagement of diverse adult learners. This finding is consistent with a meta-analysis reporting that instructor presence in videos does not significantly impact the social presence of online learners (Alemdag, 2022). However, this contrasts with synchronous online learning environments, where an instructor's social presence can notably enhance affective engagement through immediacy and intimacy (Chew, 2022; Henderson & Schroeder, 2021). Furthermore, our study did not find evidence to suggest that instructors' on-screen presence reduces learners' affective engagement, potentially due to the split-attention effect (Mayer, 2014). In asynchronous settings, learners appear to prioritize the core content of the learning material, both visual and auditory, over the instructor's on-

screen presence. This shift in focus suggests a fundamental change in instructor-learner interactions within asynchronous learning environments and underscores the need to reassess how engagement is facilitated in these contexts.

5.2 Instructors' High Arousal Facial Expressions Can Be Beneficial

Contrary to past studies which assert that the valence of an instructor's facial emotions influences students' emotions (Wang, 2022; Wang et al., 2019), our study found that facial expressions with high-intensity arousal are required to effectively stimulate learners' emotional responses, regardless of whether the valence is positive or negative (Törmänen et al., 2021). We observed that instructors frequently displaying happiness, surprise, and restrained anger can moderately enhance learners' affective engagement in asynchronous video-based learning when instructors were on screen. In contrast, negative, low-arousal emotions like sadness, fear, and disgust did not affect learners' engagement. This finding aligns with Sohriani et al.'s (2018) results, where high-arousal facial expressions of happiness, surprise, and anger significantly affected learners' engagement in synchronous online learning.

5.3 Instructors' High Arousal Paraverbal Expressions Matter

This study discovered that an instructor's paraverbal expressions significantly impact asynchronous online learners' emotions, similar to their impact on synchronous online learners, as Chew (2022) observed. Furthermore, it was revealed that instructors' paraverbal expressions have a more pronounced influence on learners' affective engagement than their facial emotive expressions. This is attributed to learners, when interacting with multimedia instructional materials, finding it more efficient to discern and respond to an instructor's voice rather than facial expressions due to cognitive load limitations (Kraus, 2017; Wang et al., 2022).

When instructors were not visible on screen, their frequent paraverbal expressions of happiness and surprise and less frequent expressions of anger resonated the most with learners. These paraverbal cues fostered a sense of social presence and emotional contagion in asynchronous video-based online learning. Like facial expressions, paraverbal expressions with high arousal demonstrated a more substantial emotional impact (Törmänen et al., 2021).

In cases where instructors were visible, their vocal attractiveness and the happiness conveyed through paraverbal expressions significantly influenced learners' affective engagement. However, vocal attractiveness accounted for 60% of the influence, while the paraverbal emotive expression of happiness only had a 2% influence, which is comparable to facial anger (2%) and lower than some specific facial emotive expressions, such as happiness (15%) and surprise (3%). The findings suggest that in the context of asynchronous video-based online learning where instructors are visually present, the overall vocal attractiveness of the instructors' paraverbal expressions emerges as the salient factor influencing learners' affective engagement, rather than specific paraverbal emotive expressions. In the same context, instructors' high-arousal facial expressions, especially those of happiness and surprise, have a significant impact on learners' affective engagement. These expressions are more influential than paraverbal expressions of happiness and facial expressions of anger.

5.4 Instructors' Vocal Charisma Effect is Profound

This study confirmed that a pleasant voice quality, as the audience perceives, significantly influences their affective engagement during listening. Such voice quality is challenging to categorize by specific emotional classifications or distinct acoustic features (Xu et al., 2013). Vocal attractiveness influences not only learners' evaluations of an instructor's presentation (Cavanagh et al., 2014) but also, as revealed by the study, significantly

enhances learners' affective engagement in video materials. This impact of vocal attractiveness exceeds that of other paraverbal and facial emotive cues, regardless of the instructors' on-screen visibility.

These findings suggest that learners in asynchronous video learning environments are exceptionally responsive to vocal charisma, significantly surpassing the impact of specific facial or paraverbal emotive expressions. The study indicated that vocal charisma alone could amplify affective engagement in video-based learning by 51% when instructors were not visible and 60% when they were visible on-screen.

6. Limitation

This study acknowledges certain limitations that may affect the generalizability of its results. Firstly, while our reliance on DL techniques and validated datasets facilitated the recognition of instructors' explicit facial and paraverbal expressions, we could not directly measure learners' subjective emotional perception of these cues due to privacy concerns. This limitation suggests an avenue for future research to validate and potentially expand our findings, as proposed by Mejbri et al. (2022).

Secondly, our model for recognizing instructors' vocal attractiveness is based on learners' subjective assessments of vocal charm, incorporating the halo effect. Identifying specific vocal features that contribute to attractiveness is challenging, making it difficult to optimize vocal charisma through feedback. Future studies could explore vocal training methods, like breathing and singing exercises, combined with technological tools, as suggested by Wang (2023).

Thirdly, while literature indicates that an instructor's appearance can influence learner engagement in online environments, we could not assess this aspect due to concerns about discrimination in the study's locale and with our partners. Future research should consider

the impact of instructor appearance on online learner engagement, considering ethical and cultural contexts and employing technology to mitigate bias, as noted by Jabali et al. (2023).

Fourthly, this study encounters a limitation stemming from a gender imbalance in the instructor sample, which was predominantly male, comprising 72%. Although this imbalance did not significantly impact most variables in our study, the skewed gender distribution may limit the generalizability of our findings. Future research should consider including more female instructors to investigate potential sex-specific influences in online video instruction settings.

Lastly, the instructional videos analyzed primarily feature popular science content and target college students and the general adult population. The applicability of our findings to more specialized or diverse content areas, as discussed by Du (2023), remains to be explored in future studies.

7. Conclusion

Our research sheds light on the unique dynamics of asynchronous video-based online learning, distinguishing it from classroom settings and synchronous online environments. By integrating DL technology into our field study, we discovered that instructors' facial and paraverbal expressions exert varying degrees of influence on learner affective engagement, depending on whether the instructors were visible or invisible. This finding highlights the crucial role of instructor vocal charisma in video instructional materials, confirming its vital contribution to learner affective engagement. Our research contributes to academic discourse by providing a detailed understanding of the paraverbal aspects of social presence in asynchronous video-based online learning. It also offers practical guidelines for educators and content creators, aiming to engage and sustain learner interest in the evolving digital educational landscape.

8. Suggestions & Implications

Our study builds on social presence and emotional contagion theories, suggesting that instructors can transfer emotions to learners through non-verbal communication. However, research predominantly focused on classroom and synchronous online learning does not entirely capture the nuances of asynchronous video-based online learning, characterized by a lack of real-time interaction and a higher propensity for learning distractions and affective disengagement.

Addressing this gap, we applied the Circumplex Model of Affect to demonstrate how instructors' facial emotive expressions of happiness, surprise, and anger in videos can influence learners' affective engagement. Furthermore, we emphasize the auditory focus in video learning by drawing on Cognitive Load Theory. Through Voice and Emotion Theory, we propose that instructors' paraverbal cues of happiness and surprise and minimized expressions of anger are essential for engaging learners. Our research demonstrates that elevated levels of emotional arousal, whether conveyed through instructors' facial or paraverbal emotive expressions, have a significant and profound impact on learners' emotional states.

Our findings augment the CoI framework's understanding of the instructor's voice, revealing its underrated role in fostering affective engagement through social presence and emotional connectedness. We introduce the concept of the 'Vocal Charisma Effect,' highlighting the significant influence of vocal attractiveness in asynchronous learning environments. This insight is crucial for the discourse on social presence, emphasizing the need to reevaluate paraverbal expressions in asynchronous video-based online learning.

This study also offers practical recommendations for video content creators and educators, incorporating insights into instructors' social presence and the effects of emotional contagion. First, while instructors need not always be visible in pre-recorded videos if they

do appear on-screen, it's crucial to focus on displaying positive emotions, especially happiness and surprise, in both facial and vocal expressions while minimizing expressions of anger. This approach leverages the principles of emotional contagion, as positive emotional expressions by instructors can subconsciously influence learners' affective states (Hatfield et al., 1993).

Second, in cases where instructors instruct off-screen, their vocal expressions should predominantly convey happiness and surprise, avoiding anger to capitalize on the auditory aspects of social presence and emotional transmission (Robinson, 2013).

Third, training instructors in facial and paraverbal expressions can be beneficial. Such training should aim at eliciting positive emotions in learners, employing explicit emotion recognition techniques based on this research for feedback and personal development. As Silvia and Duval (2001) observed, reflecting on explicit emotional states, even without constructive feedback, can foster self-awareness and motivate adjustments in emotional expression.

Fourth, the study underscores the significance of vocal charisma in emotionally engaging learners, consistent with findings on social presence in online learning environments (Borup et al., 2012). Instructors should refine their vocal delivery to enhance its attractiveness, using this study's vocal attractiveness recognition model as a guide.

Finally, to further enhance the vocal attractiveness of video materials, the use of voice synthesis software or AI-generated human voices is recommended, aligning with the latest advancements in online educational tools and technologies.

References

Akyol, Z., & Garrison, D. R. (2008). The development of a Community of Inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. *Journal of Asynchronous Learning Networks*, *12*, 3–22. https://doi.org/10.24059/olj.v12i3.72

Alemdag, E. (2022). Effects of instructor-present videos on learning, cognitive load, motivation, and social presence: A meta-analysis. *Education and Information Technologies*, *27*, 12713–12742. https://doi.org/10.1007/s10639-022-11154-w

Allison, T. H., Warnick, B. J., Davis, B. C., & Cardon, M. S. (2022). Can you hear me now? Engendering passion and preparedness perceptions with vocal expressions in crowdfunding pitches. *Journal of Business Venturing*, *37*, 106193. https://doi.org/10.1016/j.jbusvent.2022.106193

Ainley, M. (2006). Connecting with Learning: Motivation, Affect and Cognition in Interest Processes. *Educational Psychology Review*, *18*, 391–405. https://doi.org/10.1007/s10648-006-9033-0

Akiha, K., Brigham, E., Couch, B. A., Lewin, J., Stains, M., Stetzer, M. R., Vinson, E. L., & Smith, M. K. (2018). What Types of Instructional Shifts Do Students Experience? Investigating Active Learning in Science, Technology, Engineering, and Math Classes across Key Transition Points from Middle School to the University Level. *Frontiers in Education*, *2*, 68. https://doi.org/10.3389/feduc.2017.00068

Bambaeeroo, F., & Shokrpour, N. (2017). The impact of the teachers' non-verbal communication on success in teaching. *Journal of Advances in Medical Education & Professionalism*, *5*, 51–59. https://pubmed.ncbi.nlm.nih.gov/28367460/

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bao, W., Li, Y., Gu, M., Yang, M., Li, H., Chao, L., & Tao, J. (2014). Building a Chinese Natural Emotional Audio-Visual Database. *In 2014 12th International Conference on Signal Processing (ICSP)* (pp. 583-587). IEEE. https://doi.org/10.1109/ICOSP.2014.7015071

Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., & Velastin, S. A. (2021). Facial Expression Recognition of Instructor Using Deep Features and Extreme Learning Machine. *Computational Intelligence and Neuroscience*, 2021, 5570870. https://doi.org/10.1155/2021/5570870

Bolliger, D. U., & Martin, F. (2018). Instructor and student perceptions of online student engagement strategies. *Distance Education*, *39*(4), 568-583. https://doi.org/10.1080/01587919.2018.1520041

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

Buckley, S., Hasen, G., & Ainley, M. (2004). Affective engagement: A person-centred approach to understanding the structure of subjective learning experiences. Paper presented at *the Annual Conference of the Australian Association for Research in Education*, Melbourne, Australia. https://works.bepress.com/sarah_buckley/17/

Cavanagh, M., Bower, M., Moloney, R., & Sweller, N. (2014). The effect over time of a video-based reflection system on preservice teachers' oral presentations. *Australian Journal of Teacher Education*, *39*(6), Article 1. https://doi.org/10.14221/ajte.2014v39n6.3

Chew, Y. W. (2022). Performing Presence with the Teaching-Body via Videoconferencing: A Postdigital Study of the Teacher's Face and Voice. *Postdigital Science and Education*, *4*, 394–42. https://doi.org/10.1007/s42438-022-00288-2

Chiu, T. K. F. (2022). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *Journal of Research on Technology in Education*, 54, 14–30. https://doi.org/10.1080/15391523.2021.1891998

Collins, K., Groff, S., Mathena, C., & Kupczynski, L. (2019). Asynchronous video and the development of instructor social presence and student engagement. *Turkish Online Journal of Distance Education*, *20*(1), 53–70. https://doi.org/10.17718/tojde.522378

Daher, W., Sabbah, K., & Abuzant, M. (2021). Affective Engagement of Higher Education Students in an Online Course. *Emerging Science Journal*, *5*, 545–558. https://doi.org/10.28991/esj-2021-01296

Dixson, M. D. (2010). Creating effective student engagement in online courses: What do students find engaging? *Journal of the Scholarship of Teaching and Learning*, *10*, 1–13. https://scholarworks.iu.edu/journals/index.php/josotl/article/view/1744

Du, B. (2023). Research on the factors influencing the learner satisfaction of MOOCs. *Education and Information Technologies*, *28*, 1935–1955. https://doi.org/10.1007/s10639-022-11269-0

Duan, T. (2022). A new idea for the optimization of MOOC-based teaching. *Education and Information Technologies*, *27*, 3623–3650. https://doi.org/10.1007/s10639-021-10755-1

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*, 268–287. https://doi.org/10.1037/0033-2909.115.2.268

Ekman, P. (1992). Are there basic emotions? *Psychology Review*, *99*, 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Fanselow, M. S. (2018). Emotion, motivation and function. *Current Opinion in Behavioral Sciences*, *19*, 105–109. https://doi.org/10.1016/j.cobeha.2017.12.013

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Garcia, M., & Yousef, A. (2023). Cognitive and affective effects of teachers' annotations and talking heads on asynchronous video lectures in a web development course. *Research and Practice in Technology Enhanced Learning*, *18*(20), 1–23. https://doi.org/10.58459/rptel.2023.18020.

Gobl, C., & Chasaide, A. N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1-2), 189–212. https://doi.org/10.1016/S0167-6393(02)00082-1

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., ... Bengio, Y. (2013). Challenges in Representation Learning: A Report on Three Machine Learning Contests. In M. Lee, A. Hirose, Z.-G. Hou, & R. M. Kil (Eds.), *Neural Information Processing*. *ICONIP 2013*. *Lecture Notes in Computer Science* (Vol. 8228, pp. 117–124). Springer. https://doi.org/10.1007/978-3-642-42051-1_16

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis, and synthesis in continuous space: A survey. In *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 827-834). Santa Barbara, CA, USA. https://doi.org/10.1109/FG.2011.5771357

Hahow. (n.d.). *Hahow*. https://hahow.in/

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current directions in psychological science*, *2*, 96–100. https://doi.org/10.1111/1467-8721.ep10770953

Henderson, M. L., & Schroeder, N. L. (2021). A Systematic review of instructor presence in instructional videos: Effects on learning and affect. *Computers and Education Open*, *2*, 100059. https://doi.org/10.1016/j.caeo.2021.100059.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Horan, S. M., Martin, M. M., & Weber, K. (2012). Understanding Emotional Response Theory: The Role of Instructor Power and Justice Messages. *Communication Quarterly*, *60*, 210–233. https://doi.org/10.1080/01463373.2012.669323

Iskrenovic-Momcilovic, O. (2018). Learning a programming language. *International Journal of Electrical Engineering Education*, *55*, 324–333. https://doi.org/10.1177/0020720918773975

Jabali, O., Saeedi, M., Rabayaa, M., & Othman, N. (2023). Zoom dysmorphia in e-teaching: Shifting the value from attributes to appearance. *Education and Information Technologies*, *28*, 6781–6799. https://doi.org/10.1007/s10639-022-11470-1

Juslin, P. N., & Laukka, P. (2003). Communication of emotion in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Keller, M. K., & Eva, S. B. (2021). Teachers' emotions and emotional authenticity: do they matter to students' emotional responses in the classroom? *Teachers and Teaching*, *27*, 404–422. https://doi.org/10.1080/13540602.2020.1834380

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

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*, 644–654. https://doi.org/10.1037/amp0000147

Laukka, P. (2017). Vocal Communication of Emotion. In: V. Zeigler-Hill & T. Shackelford (eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-319-28099-8_562-1

Liu, M., Duan, Y., Ince, R. A. A., Chen, C., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2022). Facial expressions elicit multiplexed perceptions of emotion categories and dimensions. *Current Biology*, *32*, 200–209. https://doi.org/10.1016/j.cub.2021.10.035.

Liu, X.-Y., Chi, N.-W., & Gremler, D. D. (2019). Emotion Cycles in Services: Emotional Contagion and Emotional Labor Effects. *Journal of Service Research*, *22*, 285–300. https://doi.org/10.1177/1094670519835309

Lockee, B. B. (2021). Online education in the post-COVID era. *Nature Electronic*, *4*, 5–6. https://doi.org/10.1038/s41928-020-00534-0

Martin, F., Sung, T., & Westine, C. D. (2020a). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education*. *159*, 104009. https://doi.org/ 10.1016/j.compedu.2020.104009

Martin, F., Drew, P., Ritzhaupt, A. D. (2020b). Synchronous online learning: Blending asynchronous and synchronous online learning. *Educause Review*.

https://er.educause.edu/articles/2020/9/bichronous-online-learning-blending-asynchronous-and-synchronous-online-learning

Martin, F., & Bolliger, D. U. (2018). Engagement Matters: Student Perceptions on the Importance of Engagement Strategies in the Online Learning Environment. *Online Learning*, *22*, 205–222. https://doi.org/10.24059/olj.v22i1.1092

Mayer, R. E. (Ed.). (2014). T*he Cambridge handbook of Multimedia Learning* (2nd, rev. ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.

Mejbri, N., Essalmi, F., Jemni, M., & Alyoubi, B. A. (2022). Trends in the use of affective computing in e-learning environments. *Education and Information Technologies*, *27*(5), 3867–3889. https://doi.org/10.1007/s10639-021-10769-9

Mershad, K., & Said, B. (2022). DIAMOND: A tool for monitoring the participation of students in online lectures. *Education and Information Technologies*, *12*, 1–31. https://doi.org/10.1007/s10639-021-10801-y

Mottet, T. P., Richmond, V. P., & McCroskey, J. C. (2006). *Handbook of instructional communication: Rhetorical and relational perspectives*. Allyn & Bacon. https://doi.org/10.4324/9781315189864

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In *Handbook of research on educational communications and technology (pp. 125-144)*. Lawrence Erlbaum Associates.

Nguyen, H.-D., Kim, S.-H., Lee, G.-S., Yang, H.-J., Na, I.-S., & Kim, S.-H. (2022). Facial Expression Recognition Using a Temporal Ensemble of Multi-Level Convolutional Neural Networks. *IEEE Transactions on Affective Computing*, *13*(1), 226–237. https://doi.org/10.1109/TAFFC.2019.2946540.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Nordström, H., Laukka, P., Thingujam, N., Schubert, E., & Elfenbein, H. (2017). Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: a comparison between Australia and India. *Royal Society Open Science*, *4*, 170912. https://doi.org/10.1098/rsos.170912.

Nozaki, Y., & Mikolajczak, M. (2020). Extrinsic emotion regulation. *Emotion*, *20*(1), 10–15. https://doi.org/10.1037/emo0000636

Özhan, Ş. Ç., & Kocadere, S. A. (2020). The effects of flow, emotional engagement, and motivation on success in a gamified online learning environment. *Journal of Educational Computing Research*, *57*, 2006–2031. https://doi.org/10.1177/0735633118823159.

Paivio, A. (1990). Mental representations: A dual coding approach. Oxford Psychology Series. Oxford Academic. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Richardson, J. C., & Lowenthal, P. (2017). Instructor social presence: Learners' needs and a neglected component of the community of inquiry framework. In A. Whiteside, A. Garrett Dikkers, & K. Swan, (Eds.), Social presence in online learning: Multiple perspectives on practice and research (pp. 86-98). Stylus.

Robinson, K. (2013). The interrelationship of emotion and cognition when students undertake collaborative group work online: an interdisciplinary approach. *Computers & Education*, *62*, 298–307. https://doi.org/10.1016/j. compedu.2012.11.003

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, *63*, 2251–2272. https://doi.org/10.1080/17470211003721642

Shang, J., & Liu, Z. (2022). Vocal attractiveness matters: Social preferences in cooperative behavior. *Frontiers in Psychology*, *13*, 877530. https://doi.org/10.3389/fpsyg.2022.877530

Siegel, D. J. (2020). *The developing mind: How relationships and the brain interact to shape who we are* (3rd ed.). Guilford Press. https://drdansiegel.com/book/the-developing-mind/

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, *9*, 838–846. http://dx.doi.org/10.1037/a0017810

Silvia, P. J., & Duval, T. S. (2001). Objective Self-Awareness Theory: Recent Progress and Enduring Problems. *Personality and Social Psychology Review*, *5*(3), 230–241. https://doi.org/10.1207/S15327957PSPR0503_4

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. John Wiley & Sons. https://dx.doi.org/10.1177/009365028000700307

Sohriani, E., Djafar, M. B., & Salija, K. (2018). *Lecturer's facial expression in EFL classroom at the University of Cokroaminoto Palopo* [Unpublished thesis]. Universitas Negeri Makassar. http://eprints.unm.ac.id/id/eprint/11172

Sweller, J. (1988). Cognitive load during problem solving: *Effects on learning. Cognitive Science*, *12*(2), 257-285. https://doi.org/10.1207/S15516709COG1202_4

Theonas, G., Hobbs, D., and Rigas, D. (2007). The effect of facial expressions on students in virtual educational environments. *International Journal of Educational and Pedagogical Sciences*, *11*, 625–632. https://doi.org/10.1186/2193-1801-2-455

Törmänen, T., Järvenoja, H., & Mänty, K. (2021). All for one and one for all – How are students' affective states and group-level emotion regulation interconnected in collaborative learning? *International Journal of Educational Research*, *109*, 101861. https://doi.org/10.1016/j.ijer.2021.101861

Trager, G. L. (1958). Paralanguage: A first approximation. *Studies in Linguistics*, *13*, 1-12.

Walker, K. A., & Koralesky, K. E. (2021). Student and instructor perceptions of engagement after the rapid online transition of teaching due to COVID-19. *Natural Sciences Education*, 50, e20038. https://doi.org/10.1002/nse2.20038

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*(1), 52–90. https://doi.org/10.1177/009365092019001003

Wang, Y. (2023). The effectiveness of innovative technologies to manage vocal training: The knowledge of breathing physiology and conscious control in singing. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12108-6

Wang, Y. (2022). To Be Expressive or Not: The Role of Teachers' Emotions in Students' Learning. *Frontiers in Psychology*, *12*, 737310. https://doi.org/10.3389/fpsyg.2021.737310

Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education*, *146*, 103779. https://doi.org/10.1016/j.compedu.2019.103779

Wang, J.-Y., Yang, C.-H., Liao, W.-C., Yang, K.-C., Chang, I.-W., Sheu, B.-C., & Ni, Y.-H. (2022). Highly engaged video-watching pattern in asynchronous online pharmacology course in pre-clinical 4th-year medical students was associated with a good self-expectation, understanding, and performance. *Frontiers in Medicine*, *8*, 799412. https://doi.org/10.3389/fmed.2021.799412

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

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, *33*(1), 59–69. https://doi.org/10.1007/s10919-008-0057-7.

Wermeskerken, M. V., Ravensbergen, S., & Gog, T.V. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior*, *89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Wise, A., Chang, J., Duffy, T., & Del Valle, R. (2004). The Effects of Teacher Social Presence On Student Satisfaction, Engagement, and Learning. *Journal of Educational Computing Research*, *31*, 247–271. https://doi.org/10.2190/V0LB-1M37-RNR8-Y2U1

Ye, B., Yuan, X., Peng, G., & Zeng, W. (2022). A Novel Speech Emotion Model Based on CNN and LSTM Networks. In *2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT)*, 1-4. https://doi.org/10.1109/ACAIT56212.2022.10137926

Yuyun, I. (2023). Investigating university student engagement in online learning: A case study in EFL classroom. *Indonesian Journal of Applied Linguistics*, *12*(3). https://doi.org/10.17509/ijal.v12i3.46035

Yuan, M., Zeng, J., Wang, A., & Shang, J. (2021). Would It Be Better if Instructors Technically Adjust Their Image or Voice in Online Courses? Impact of the Way of Instructor Presence on Online Learning. *Frontiers in Psychology*, *12*, 746857. https://doi.org/10.3389/fpsyg.2021.746857

Zhang, Q., & Zhang, J. (2013). Instructors' Positive Emotions: Effects on Student Engagement and Critical Thinking in U.S. and Chinese Classrooms. *Communication Education*, *62*, 395–411, https://doi.org/10.1080/03634523.2013.828842

Zhao, H. & Khan, A. (2022). The Students' Flow Experience With the Continuous Intention of Using Online English Platforms. *Frontiers in Psychology*, *12*, 807084. https;//doi.org/10.3389/fpsyg.2021.807084