European Journal of Psychology of Education
https://doi.org/10.1007/s10212-023-00715-3

How do native and non-native speakers recognize emotions in the instructor's voice in educational videos? Exploring the first step of the cognitive-affective model of e-learning for international learners

Nežka Sajinčič[1] · Anna Sandak[1,2,3] · Amy Simmons[1,3] · Andreja Istenič[4]

Received: 26 January 2023 / Revised: 18 May 2023 / Accepted: 5 June 2023

Abstract

The emotional stance of the instructor in an educational video can influence the learning process. For this reason, we checked the first link of the cognitive-affective model of e-learning, namely, whether learners can recognize emotions that an instructor expresses only with their voice. Since English is not the native language for many learners and most instructional videos are produced in English, we tested for possible differences in emotion recognition between native and non-native speakers. We focused on positive emotions typically conveyed in such videos — enthusiasm and calmness. Native and non-native English speakers watched 12 short video clips about wood as a building material spoken by an instructor in different emotional tones — five videos expressed enthusiasm, five calmness, one boredom and one frustration. Participants rated the extent to which they thought the narrator expressed a specific emotion, the valence and activation level of the narration and solved an English vocabulary test. Both native and non-native speakers recognized the correct emotions (except for frustration), demonstrating the power of voice prosody to convey emotion in a multimedia learning scenario. Native speakers rated the enthusiastic videos more positively than non-native speakers, indicating a subtle difference in the way the two groups perceive emotions expressed through voice.

Keywords Emotional design · Emotion recognition · Voice prosody · Non-native speakers · Instructional video · Multimedia learning

✉ Nežka Sajinčič
nezka.sajincic@innorenew.eu

[1] InnoRenew CoE, Izola, Slovenia

[2] Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia

[3] Andrej Marušič Institute, University of Primorska, Koper, Slovenia

[4] Faculty of Education, University of Primorska, Koper, Slovenia

Published online: 01 July 2023

Introduction

Learning by watching online videos has become ubiquitous all over the world, and creating educational videos, such as narrated PowerPoint presentations, has never been easier, so research on how to make them more effective is vital from both the learners' and educators' perspectives. Over the past two decades, extensive research has been conducted based on cognitive multimedia learning theory (Mayer, 2014) and cognitive load theory (Sweller et al., 2011) to determine how to design learning videos in accordance with the way the human cognition works. Since learning is also affected by non-cognitive factors such as affective processes (Tyng et al., 2017; Wu et al., 2016), the focus has now shifted to expanding the theory of learning with multimedia.

Using an approach called emotional design, researchers are now investigating whether certain features of learning materials can affect learners' emotions, which in turn could enhance learning (Plass & Kaplan, 2016). To date, the focus has been on visual features such as shapes, colours (Javora et al., 2018; Wong & Adesope, 2020), decorative images (Schneider et al., 2016) and onscreen pedagogical agents (Lawson et al., 2021c), but the auditory portion of videos needs further attention, particularly the emotional tone of the video instructor.

Many educational videos contain narration without a visual representation of the instructor, so it is important to examine the effect of emotions conveyed only through voice. The first question is whether learners can differentiate and recognize the emotions that the video instructor expresses only through vocal cues. Voice plays a crucial role in communication because it can convey emotional information independent of verbal content by using nonverbal emotional vocalizations and emotional prosody (Wilson & Wharton, 2006). Emotional or vocal prosody refers to the changes in pitch, loudness, rhythm and voice quality present in speech and is a key part of a video lecture that needs further investigation.

Another gap in the literature on the emotional design of multimedia learning is that the vast majority of research has been conducted with educational materials in the learners' native language. Comparative studies only involved participants who were native speakers of the language of the instructional video. There are some documented differences in learning processes when learning from multimedia presentations in our native language or a foreign language, with some interventions being ineffective or detrimental for native speakers while benefiting learners who view videos in their non-native language (Davis & Vincent, 2019; Lee & Mayer, 2018; Mayer & Fiorella, 2014) or even just in another dialect (Rey & Steib, 2013; Schneider et al., 2015). A similar pattern emerged in the field of emotion recognition (Laukka & Elfenbein, 2021).

In a globalized world with open online education, many people who consume English learning materials are non-native speakers with different levels of language proficiency, so research on such a ubiquitous learning technology needs to better represent the context in which many people now learn. This paper, therefore, presents a study that included both native and non-native English speakers who watched educational clips with an English-speaking narrator.

Recognizing the emotional tone of video instructors from their voice

In this article, we describe emotional states as a combination of two bipolar and orthogonal dimensions of core affect: valence (the degree of positivity or negativity) and activation (the degree of physiological alertness or attentiveness; Russell, 1980).

Enthusiasm is an example of a pleasant and activated emotional state, and calmness is an example of a pleasant but deactivated state (Fig. 1). Early research on the role of emotions in educational contexts has focused almost exclusively on negative emotions such as anxiety, but research on technology-based learning has shown how positive emotions can improve learning and affective outcomes (Loderer et al., 2020).

While learning through video watching lacks the direct interaction that is part of face-to-face instruction, video lectures can still be designed to convey social cues. Social cues are verbal and non-verbal signals that can trigger a feeling of social presence in learners (Atkinson et al., 2005) and can lead to better processing and learning outcomes (Moreno et al., 2001; Schneider et al., 2021). Based on cognitive-affective theories of multimedia learning like the cognitive-affective theory of learning with media (Moreno, 2006), the integrated model of cognitive-affective learning with media (Plass & Kaplan, 2016), the cognitive-affective-social theory of learning in digital environments (CASTLE; Schneider et al., 2021) and the cognitive-affective model of e-learning (Mayer, 2020), the emotional stance of an instructor can provide such social cues and influence learners and their learning process even through the use of learning technologies such as video instruction.

The CASTLE theory (Schneider et al., 2021) proposes that the cognitive processing of the learning material is mediated by emotional and social processes that are triggered by the social cues in an instruction. However, the effect can vary in people with different cultural and/or language background (Brom et al., 2017; Schneider et al., 2015), so more research is needed to determine not only the actual process, but the boundary conditions as well. Recently, the authors of the cognitive-affective model of e-learning (Lawson et al., 2021b; Mayer, 2020) described a five-step process of how the emotional stance of a video instructor can affect learners. First, (1) the instructor displays an emotional state (e.g. enthusiasm) in an e-learning episode, (2) which is detected and recognized by learners. This triggers (3) an emotional response in learners (e.g. feeling a positive social partnership with the instructor), which in turn (4) affects their cognitive processing (e.g. promotes motivation to try to learn harder) and finally (5) their learning outcomes.

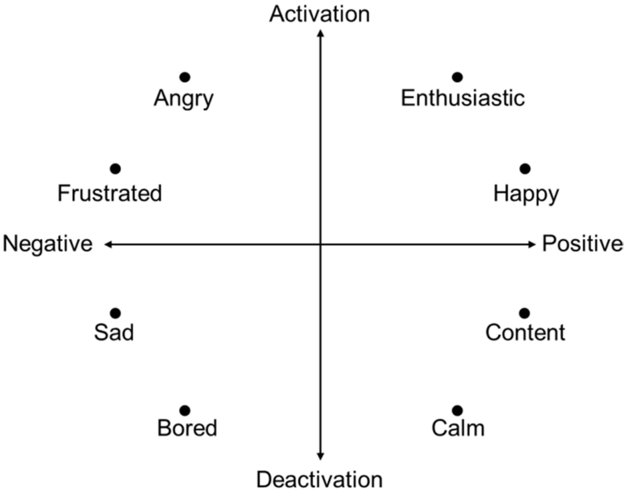


Fig. 1 Model of core affect (Russell, 1980)

Some studies have already examined how the emotional state of an onscreen agent affects learners and their learning outcomes. However, in these cases, instructors conveyed their emotions not only through the prosody of their voice, but also with the help of other non-verbal cues such as facial expressions, gestures, body posture or anthropomorphisms (e.g. Davis & Vincent, 2019; Lawson et al., 2021a, 2021c; Schneider et al., 2022; Um et al., 2012). For instance, recent research has shown that learners can successfully recognize emotions displayed by human or virtual instructors (Horovitz & Mayer, 2021; Lawson et al., 2021b).

Research on the effects of cues given only by voice, particularly emotional prosody, however, is sparse, but promising (Beege et al., 2020; Liew et al., 2020). A study comparing learners watching videos with an instructor present or absent found that the ability to recognize emotion did not decrease when only the agent's voice was used (Lawson & Mayer, 2021). However, the authors noted that while participants were able to successfully distinguish between positive and negative emotions, they had greater difficulty distinguishing between emotions of the same emotional valence.

Emotion recognition of native and non-native speakers

Previous studies of emotion recognition across languages and cultures have focused primarily on facial expressions, while voice has only recently attracted attention. Based on 37 cross-cultural studies, a meta-analysis provided evidence that a wide array of emotions expressed using only speech prosody can be detected with above-chance accuracy by non-native speakers or even people who do not understand the language (Laukka & Elfenbein, 2021). However, there was also a clear in-group advantage, as people recognized emotions more accurately in their native language than in a foreign language.

For example, a recent study showed that native speakers are much better at recognizing emotion from voice only than non-native speakers and that those with higher language proficiency outperform those with lower proficiency (Lorette & Dewaele, 2018). However, in this study, emotions were not only expressed through vocal prosody, but verbally as well, providing additional cues for participants. To eliminate the confounding variable of verbal cues, a study used pseudo-utterances — nonsense speech that sounds like a language — and found similar results: native speakers were both more accurate and faster at recognizing emotional expressions compared to non-native speakers, and non-native speakers who were more proficient in the foreign language in question were also faster and more accurate than participants with lower proficiency (Jiang et al., 2015). This is consistent with the dialect theory, which states that while the communication of emotions is largely universal, there are subtle cultural differences in how we express emotions (Elfenbein, 2013). When people interpret other people's non-verbal signals, they do so based on their own cultural style of producing signals, which can lead to subtle misunderstandings when recognizing emotions in a foreign language, manifested in lower accuracy and efficiency of vocal emotion processing, even when individuals are highly proficient in the target language (Jiang et al., 2015).

However, another study showed that non-native speakers with higher English proficiency were less accurate in recognizing positive emotions than those with lower English proficiency (Bhatara et al., 2016), suggesting that the more proficient participants paid more attention to sentence content than the prosody. In terms of valence, the authors argued that recognizing negative emotions has greater evolutionary value both within and between groups because it is more important for survival, whereas recognizing positive emotions is only important within the same culture to strengthen social ties, making recognizing negative emotions more universal and recognizing positive emotions more culturally specific (Ekman, 1992). On the other hand, research on emotion communication has been heavily biased towards negative emotions, distinguishing between different negative emotions (e.g. sadness, disgust, anger, fear) but only using happiness as a positive emotion, making it necessary to study different positive emotions as well (Sauter, 2010). The scientific community has also been calling for more socially relevant and naturally occurring emotional stimuli in emotional recognition research that is free of other cues (such as verbal meaning or facial expressions) (Morningstar et al., 2021). Recorded lectures conveying the same content but expressing different emotions by vocal prosody only are therefore a suitable stimulus to verify whether there are differences in emotion recognition between native and non-native speakers.

The present study

Before addressing the question of what emotional tone teachers should convey in educational videos to support learning, we need to establish the first link of the cognitive-affective model of e-learning and define whether voice prosody provides enough cues for learners to recognize the intended emotion. The purpose of this study was therefore to investigate whether learners can recognize the emotional tone of an instructor in educational videos solely based on their voice, and whether there are differences in emotion recognition between native and non-native English speakers. Since most educational videos are communicated in a positive or neutral emotional tone that aims to promote positive emotional states like motivation (Liew et al., 2017), the present study focused on two positive emotions — enthusiasm and calmness. Participants were shown multiple educational video clips in which a disembodied (non-visually present) instructor narrated content in English in either an enthusiastic or calm voice and one video clip narrated in a frustrated and one in a bored voice. Native and non-native English speakers rated the extent to which they believed the instructor expressed certain emotions.

According to the cognitive-affective model of e-learning, participants should correctly recognize the emotion portrayed by the narrator (Hypothesis 1). Specifically, participants will rate enthusiastic videos as significantly more enthusiastic than the other emotions (Hypothesis 1a) and calm videos as significantly calmer than other emotions (Hypothesis 1b). We also predict that enthusiastic videos will be rated as significantly higher in activation level than calm videos (Hypothesis 2). Based on the dialect theory, we predict that native speakers will rate the videos more accurately (Hypothesis 3) — they will rate enthusiastic clips as significantly more enthusiastic (Hypothesis 3a) and calm clips as significantly calmer (Hypothesis 3b) than non-native speakers.

Methods

Design

The online experiment used a 2×2(4) mixed factorial design, in which the between-subjects factor was being a native or non-native English speaker, and the within-subjects factor was the type of emotion portrayed by the video narrator (enthusiasm and calmness, and in the case of one video, frustration and boredom). All participants viewed and rated 12 clips.

Participants

A convenience sample of 207 people (132 women, 69 men, 2 non-binary, 4 undisclosed) participated in the study, of which 196 completed the survey in full. Their demographic information is presented in Table 1. One hundred sixty-two completed the survey in English and 47 in Slovene (with the videos still being in English). Eighty-seven reported being native English speakers; for 99, English is their second language; for 18, their third and for three, their fourth. Most native English speakers originated from the UK (49) and the USA (16), and most participants who were not native speakers were originally from Slovenia (49), Poland (12) and Germany (11). Participation in the study was voluntary, and subjects received no compensation for their participation.

Materials and procedure

Participants were recruited via social media and emails from the authors. All materials were computer-based and presented on the online platform 1 ka.si (Faculty of Social

Table 1 Participants demographics split between native and non-native English speakers

	Native English speakers ($n = 87$)		Non-native English speakers ($n = 120$)	
	n	*f%*	*n*	*f%*
Gender				
Female	62	71.26%	70	58.33%
Male	23	26.44%	46	38.33%
Non-binary	1	1.15%	1	0.83%
Undisclosed	1	1.15%	3	2.50%
Age group				
16–25	56	64.37%	38	31.67%
26–35	11	12.64%	50	41.67%
36–45	11	12.64%	12	10.00%
46–55	7	8.05%	14	11.67%
56 <	2	2.30%	6	5.00%
Education				
Primary education	1	1.15%	2	1.67%
Secondary education	26	29.89%	11	9.17%
Bachelor's degree (first Bologna cycle or equivalent)	42	48.28%	36	30.00%
Master's degree (second Bologna cycle or equivalent)	12	13.79%	43	35.83%
Doctorate degree or equivalent	5	5.75%	27	22.50%
Undisclosed	1	1.15%	1	0.83%
Status				
High school student	3	3.45%	3	2.50%
University student	55	63.22%	40	33.33%
(Self-)employed	22	25.29%	68	56.67%
Employed and student	4	4.60%	0	0.00%
Unemployed	2	2.30%	6	5.00%
Retired	1	1.15%	0	0.00%
Undisclosed	0	0.00%	3	2.50%

Sciences, University of Ljubljana, 2022). They included 12 short video clips, demographic questions, rating surveys, a questionnaire and an English vocabulary test.

Subjects viewed 12 narrated PowerPoint presentations, ranging in duration from 32 to 65 s, taken from a video presentation on wood as a building material (see Supporting information for video clips). In terms of content, there were five different videos (introduction, protective design measures, durability classes, degradation control and types of coatings) — five were narrated in an enthusiastic voice, five in a calm voice, one in a bored voice and one in a frustrated voice. Although comparing the recognition of negative and positive emotions was not the main goal of the study, we added two videos in which a negative emotion (one activated and one deactivated) was portrayed to cover a broader set of emotional stimuli and help the participants make more accurate ratings of the clips. This also allowed us to verify previous findings that participants are better able to discriminate emotions according to their valence rather than their activation level (Lawson et al., 2021b). The narrations were recorded by a woman with a Standard American English accent reading a script. She portrayed different emotions in a realistic and non-exaggerated way based on our pointers and feedback. For example, for the "enthusiasm" condition, the narrator was instructed to have a varied and uplifting intonation and to make regular changes in tone and pitch (Collins, 1978).

The videos were shown in random order. After viewing each video, participants rated each video on seven items adapted from similar studies (e.g. Lawson & Mayer, 2021; Lawson et al., 2021a, 2021b, 2021c). First, they were asked to indicate on a 7-point rating scale the extent to which they thought the narrator expressed five emotions: enthusiastic, calm, frustrated, happy, and bored. The rating for "happy" was added as a positive emotion with a level of activation that is between enthusiasm and calmness, to help us understand how positive emotions are perceived. The method of rating the presence of several emotions was chosen as it conveys more information compared to a forced choice question (e.g. whether more emotions are perceived in the voice and the intensity of the perceived emotion). Next, participants rated the activation level and pleasantness of the narrator video on a 9-point scale (extremely passive/unpleasant to extremely active/pleasant).

Before the videos, participants evaluated their knowledge ($M=2.90$, $SD=1.64$) and interest ($M=3.73$, $SD=1.91$) in the topic covered in the videos. After watching the clips, they were also asked to rate on a 7-point rating scale (from very low/very not interested to very high/very interested) how interesting the presented material was ($M=4.12$, $SD=1.73$) and how well they understand English. Participants were also asked about their language background, gender, age, and education.

At the end, we assessed participants' English proficiency using LexTALE, a vocabulary test that requires participants to indicate whether or not the presented 60 items are existing English words (Lemhöfer & Broersma, 2012). The resulting score has been shown to give a good indication of the English proficiency of people with varying language backgrounds and was highly reliable in our study ($\omega=0.93$).

Control of extraneous variables was limited, as this was an online study with non-random sampling. However, measures were taken to ensure participants had a similar and appropriate environment by standardizing the experimental procedure and instructing each participant to allocate sufficient time for the experiment and to minimize distractions in their surroundings before starting the study. Additionally, the video clips were presented randomly to minimize any potential order effects. While situational variables and individual differences are less problematic for the within-subject part of our experiment, these factors pose a greater threat to findings concerning the comparison of results between native and non-native English speakers. Therefore, we checked for differences in several factors between groups and included significant differences as covariates in further analyses.

Statistical analysis

Data were processed and analysed using the open-source software R (R Core Team, 2020) and jamovi (The Jamovi Project, 2021).

We conducted repeated-measures ANOVAs with a Greenhouse–Geisser correction for lack of sphericity and post hoc pairwise *t* tests with a Bonferroni correction to compare ratings on the averaged calm and enthusiastic clips (average made from ratings on all five clips per emotion) and the frustrated and bored clip. Separate ANOVAs were conducted for each dependent variable. However, comparisons involving negative emotions should be considered carefully as they are based on ratings for only one clip per emotion.

To check whether there are differences in recognizing emotion between native and non-native English speakers, we first tested for possible differences in basic characteristics between the two groups with independent *t* tests with a Bonferroni-adjusted α level of 0.010 (0.05/5). The variables that were shown to differ significantly between the native and non-native speakers were included as covariates in the following ANCOVAs to control for their possible effects. Comparisons were based on estimated marginal means.

Results

Recognising the portrayed emotion

Ratings of videos with different portrayed emotions are displayed in Table 2, where the results can be seen combined or split between native and non-native speakers. In general, participants recognized the emotion portrayed in the video clips, except for the frustrated one. Below, we present details for each emotion. We report only the results for both groups together unless there is a difference between the groups. In addition, frustration and boredom were only expressed in one video each, so results related to these clips should be interpreted with caution.

Enthusiastic videos

An ANOVA on the averaged ratings of enthusiastic videos produced a significant main effect, $F(2.25, 468.60) = 613.00$, $p < 0.001$, $\eta^2_p = 0.75$. Pairwise comparisons showed that the enthusiastic rating was significantly higher than the calm ($t(208) = 8.03$, $p < 0.001$, mean difference = 0.71, 95% *CI* [0.54–0.88], $d = 0.56$, 95% *CI* [0.41–0.70]), happy ($t(208) = 7.04$, $p < 0.001$, mean difference = 0.30, 95% *CI* [0.21–0.38], $d = 0.49$, 95% *CI* [0.34–0.63]), frustrated ($t(208) = 33.78$, $p < 0.001$, mean difference = 3.37, 95% *CI* [3.17–3.57], $d = 2.34$, 95% *CI* [2.07–2.60]) and bored rating ($t(208) = 27.68$, $p < 0.001$, mean difference = 3.05, 95% *CI* [2.83–3.27], $d = 1.91$, 95% *CI* [1.69–2.14]). These results are consistent with Hypothesis 1a predicting that participants will rate enthusiastic videos as significantly more enthusiastic than the other emotions.

Calm videos

An ANOVA on the averaged ratings of the calm videos produced a significant main effect, $F(2.89, 601.34) = 378.41$, $p < 0.001$, $\eta^2_p = 0.65$. Paired samples *t* tests showed that the calm rating was significantly higher than the enthusiastic ($t(208) = 33.21$, $p < 0.001$, mean difference = 3.16, 95% CI [2.97 – 3.35], $d = 2.30$, 95% CI [2.04–2.56]), happy ($t(208) = 31.81$,

Table 2 Descriptive statistics of emotional tone ratings for the video clips expressing different emotions

Clips	Emotion	Combined		Native speakers ($n=87$)		Non-native speakers ($n=120$)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Enthusiastic (averaged)	**Enthusiastic**	**5.08**	**0.97**	**5.26**	**0.85**	**4.95**	**1.02**
	Calm	4.37*	0.98	4.62*	0.87	4.17*	1.01
	Frustrated	1.71*	0.86	1.68*	0.91	1.74*	0.83
	Happy	4.78*	1.02	4.97*	0.96	4.64*	1.04
	Bored	2.03*	0.93	2.08*	0.90	2.00*	0.96
Calm (averaged)	Enthusiastic	2.17*	0.88	2.19*	0.95	2.17*	0.83
	Calm	**5.33**	**0.98**	**5.31**	**0.94**	**5.35**	**1.01**
	Frustrated	2.60*	1.37	2.31*	1.26	2.82*	1.41
	Happy	2.28*	0.93	2.47*	0.94	2.13*	0.89
	Bored	4.75*	1.31	4.92	1.32	4.63*	1.31
Frustrated	Enthusiastic	3.89	1.57	3.85	1.71	3.88	1.46
	Calm	3.96	1.52	3.70	1.41	4.12	1.57
	Frustrated	**3.33**	**2.06**	**3.84**	**2.11**	**3.01**	**1.95**
	Happy	3.06	1.61	2.92*	1.73	3.15	1.50
	Bored	3.58	1.77	4.01	1.88	3.29	1.62
Bored	Enthusiastic	1.34*	0.80	1.37*	0.86	1.33*	0.76
	Calm	4.16*	1.77	4.14*	1.69	4.17*	1.83
	Frustrated	4.32*	2.10	4.16*	2.16	4.47*	2.05
	Happy	1.40*	0.80	1.41*	0.80	1.39*	0.80
	Bored	**5.85**	**1.71**	**6.06**	**1.55**	**5.70**	**1.80**

*Ratings that are significantly lower than ratings for the target emotion. The emotional tone rating item that matches the emotion expressed in each video is bolded

$p<0.001$, mean difference = 3.05, 95% CI [2.86–3.24], $d=2.20$, 95% CI [1.95–2.45]), frustrated ($t(208)=21.24$, $p<0.001$, mean difference = 2.73, 95% CI [2.47–2.98], $d=1.47$, 95% CI [1.27–1.66]), and bored rating ($t(208)=4.93$, $p<0.001$, mean difference = 0.58, 95% CI [0.35–0.81], $d=0.34$, 95% CI [0.20–0.48]), supporting Hypothesis 1b positing that the calm ratings of calm videos will be significantly higher compared to ratings of other emotions.

A separate comparison of ratings in the native English speakers' group revealed that their calm rating was significantly higher than all other emotions, except for bored ($t(86)=2.17$, $p=0.033$, mean difference = 0.38, 95% CI [0.03–0.74], $d=0.23$, 95% CI [0.02–0.44]). Non-native speakers, on the other hand, had significantly lower ratings of bored than calm ($t(119)=4.56$, $p<0.001$, mean difference = 0.71, 95% *CI* [0.40–1.03], $d=0.42$, 95% *CI* [0.23–0.60]).

Frustrated video

An ANOVA on the frustrated video produced a significant, but small main effect, $F(2.43, 505.59)=9.40$, $p<0.001$, $\eta^2_p=0.04$. Post hoc tests revealed that the frustrated rating was not significantly higher than any of the other ratings. In fact, it was rated significantly lower than the enthusiastic (t(208) = −2.7, p = 0.007, mean difference = −0.55, 95% CI [−0.95 to −0.15], d = −0.19, 95% CI [−0.32 to −0.05]) and calm rating (t(208) = −2.96, p = 0.003, mean difference = −0.62, 95% CI [−1.04 to −0.21], d = −0.20, 95% CI [−0.34 to −0.07]), and there were no significant differences with the happy (t(208) = 1.32, p = 0.187, mean difference = 0.28, 95% CI [−0.14 to 0.69], d = 0.09, 95% CI [−0.04 to 0.23]), and bored rating (t(208) = −1.63, p = 0.104, mean difference = −0.25, 95% CI [−0.55 to 0.05], d = −0.11, 95% CI [−0.25 to 0.02]).

Native English speakers rated the clip as significantly less happy than frustrated (t(86) = 2.58, p = 0.011, mean difference = 0.92, 95% CI [0.21–1.63], d = 0.28, 95% CI [0.06–0.49]), which was not repeated by their non-native peers (t(119) = −0.57, p = 0.569, mean difference = −0.14, 95% CI [−0.63 to 0.35], d = −0.05, 95% CI [−0.23 to 0.13]).

Bored video

An ANOVA on the bored video produced a significant main effect, F(2.74, 570,50) = 343.16, $p < 0.001$, $\eta^2_p = 0.42$. t tests showed that the bored rating was significantly higher than the enthusiastic (t(208) = 31.53, $p < 0.001$, mean difference = 4.50, 95% CI [4.22–4.78], d = 2.18, 95% CI [1.93–2.43]), calm (t(208) = 10.43, $p < 0.001$, mean difference = 1.68, 95% CI [1.37–2.00], d = 0.72, 95% CI [0.57–0.87]), happy (t(208) = 31.83, $p < 0.001$, mean difference = 4.44, 95% CI [4.17–4.72], d = 2.20, 95% CI [1.95–2.45]) and frustrated rating (t(208) = 9.21, $p < 0.001$, mean difference = 1.53, 95% CI [1.20–1.85], d = 0.64, 95% CI [0.49–0.79]).

Comparing emotions based on activation level and valence

Additionally, participants rated the clips not only by means of discrete emotions, but also based on their activation level and valence (Table 3). We compared ratings of activation level and valence between the enthusiastic and calm clips, but also between enthusiastic and frustrated clips and between calm and bored clips.

Both activation (F(2.47, 617.37) = 330.70, $p < 0.001$, $\eta^2_p = 0.61$) and valence (F(2.41, 501.06) = 282.02, $p < 0.001$, $\eta^2_p = 0.58$) had a significant main effect. Enthusiastic clips were rated as significantly more activated than the calm clips (t(208) = 30.62, $p < 0.001$, mean difference = 2.72, 95% *CI* [2.55–2.90], d = 2.12, 95% *CI* [1.87–2.36]), confirming Hypothesis 2, and the frustrated clip as significantly more activated than the bored clip (t(208) = 16.10, $p < 0.001$, mean difference = 2.32, 95% CI [2.04–2.60], d = 1.11, 95% CI [0.94 – 1.29]). Considering valence, enthusiastic clips were rated significantly more positively than the frustrated one (t(208) = 12.72, $p < 0.001$, mean difference = 1.47, 95% CI [1.24–1.70], d = 0.88, 95% CI [0.72–1.04]) and the calm ones significantly more positively than the bored one (t(208) = 16.54, $p < 0.001$, mean difference = 1.54, 95% CI [1.35–1.72], d = 1.14, 95% CI [0.97–1.32]). There was, however, also a significant difference in valence between the enthusiastic and calm clips (t(208) = 21.89, $p < 0.001$, mean difference = 1.76, 95% CI [1.60–1.92], d = 1.51, 95% CI [1.31–1.71]), but smaller compared to the difference in activation level.

Comparison of native and non-native English-speaking participants

As expected, there were significant differences in both self-evaluated English proficiency (t(205) = 4.15, $p < 0.001$, mean difference = 0.61, 95% *CI* [0.32–0.90], d = 0.58, 95% *CI*

Table 3 Descriptive statistics of activation and valence ratings for the video clips expressing different emotions

Clips	Core affect dimension	Combined		Native speakers ($n=87$)		Non-native speakers ($n=120$)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Enthusiastic (averaged)	Activation	6.17	0.94	6.12	0.92	6.20	0.96
	Valence	6.23	0.96	6.25	0.97	6.21	0.95
Calm (averaged)	Activation	3.45*	0.99	3.38*	0.97	3.50*	1.00
	Valence	4.46*	1.02	4.61*	1.08	4.36*	0.97
Frustrated	Activation	5.03	1.59	4.89	1.79	5.12	1.43
	Valence	4.76*	1.62	4.48*	1.79	4.93*	1.47
Bored	Activation	2.71*	1.56	2.74*	1.51	2.70*	1.61
	Valence	2.93*	1.43	3.09*	1.52	2.78*	1.35

*Ratings that are significantly lower than ratings for the comparative activation/valence rating

[0.29–0.87]) and English vocabulary test scores ($t(196)=4.67$, $p<0.001$, mean difference $=9.84$, 95% *CI* [5.68–13.99], $d=0.67$, 95% *CI* [0.37–0.97]) between groups, indicating that the two groups are different to the point that there may also be differences in how they perceive emotions from the English videos. While native speakers had higher self-ratings ($M=6.49$, $SD=0.97$) and proficiency scores ($M=86.88$, $SD=14.26$) than non-native speakers, non-native speakers still had relatively high results in both cases ($M=5.88$, $SD=1.09$ for self-evaluation and $M=77.04$, $SD=14.91$ for test scores), representing the population that is likely to watch online educational videos in English. However, English proficiency was not related to the ability to recognize emotions in the non-native speakers' group, as there were no significant correlations between the English proficiency score and the ability to recognize either enthusiasm ($r=0.06$, $p=0.537$) or calmness ($r=0.12$, $p=0.197$).

In addition to English skills, the groups differed significantly in prior interest in the topic ($t(205)=5.89$, $p<0.001$, mean difference $=1.47$, 95% *CI* [0.98–1.96], $d=0.83$, 95% *CI* [0.52–1.13]) and finding the instructional materials interesting ($t(205)=5.57$, $p<0.001$, mean difference $=1.27$, 95% *CI* [0.82–1.72], $d=0.78$, 95% *CI* [0.48–1.08]), with non-native speakers being more interested in the content ($M=4.34$, $SD=1.88$) and finding the videos more interesting ($M=4.65$, $SD=1.58$) than their native speaking peers ($M=2.87$, $SD=1.61$ for prior interest and $M=3.38$, $SD=1.68$ for interest in the instructional material). While there were no significant differences in gender ($\chi^2(2, N=204)=3.55$, $p=0.169$), age ($t(205)=-2.22$, $p=0.028$), or self-evaluated prior knowledge of the content ($t(205)=-1.88$, $p=0.061$) between the groups, there were significant differences in educational level ($U=2871.50$, $p=<0.001$), with non-native speakers having higher education ($M=3.69$, $SD=0.98$, $Mdn=4$, $IQR=3$–4) than native speakers ($M=2.93$, $SD=0.85$, $Mdn=3$, $IQR=2$–3).

Based on these results, prior interest in the topic, interest in the videos and education are included as covariates in the ANCOVAs used to compare native and non-native speakers' ratings of video clips.

Figure 2 shows native English speakers have consistently rated enthusiastic videos more positively, both when rating discrete emotions and valence. There was a main effect of group affiliation, $F(4, 200)=7.80$, $p<0.001$, with a post hoc comparison ($t(200)=2.91$, $p=0.004$, mean difference $=0.43$, $d=0.47$, 95% CI [0.15–0.79]) showing that on average, native English speakers rated the enthusiastic videos as significantly more enthusiastic as their non-native peers, supporting Hypothesis 3a. The effect was also significant when rating enthusiastic videos as happy ($F(4, 200) = 9.32$, $p < 0.001$), but also as calm ($F(4, 200) = 3.94$, $p = 0.004$), as native speakers also rated the enthusiastic videos as more happy ($t(200) = 2.31$, $p = 0.022$, mean difference = 0.36, $d = 0.37$, 95% CI [0.05–0.69]) and calm ($t(200) = 2.61$, $p = 0.010$, mean difference = 0.40, $d = 0.42$, 95% CI [0.10–0.74]) than non-native English speakers. After controlling for covariates, there were no significant differences in the case of activation level ($t(200) = 0.05$, $p = 0.959$) and valence ratings ($t(200) = 1.39$, $p = 0.165$).

Regarding calm videos (Fig. 3), differences were less pronounced, but followed a similar pattern. In contrast to Hypothesis 3b, there was no main effect in the case of the calm rating, $F(4, 200) = 0.37$, $p = 0.830$. However, there was a significant main effect in rating calm videos as happy, $F(4, 200) = 4.38$, $p = 0.002$, with native speakers rating them as more happy than non-native speakers ($t(200) = 2.25$, $p = 0.026$, mean difference = 0.32, $d = 0.36$, 95% CI [0.04–0.68]). There were no significant differences in the case of activation level ($t(200) = -0.14$, $p = 0.885$), but in terms of valence, native speakers rated calm videos as significantly more positive than non-native speakers, $F(4, 200) = 2.48$, $p = 0.045$, $t(200) = 2.20$, $p = 0.029$, mean difference = 0.36, $d = 0.35$, 95% CI [0.03–0.68].

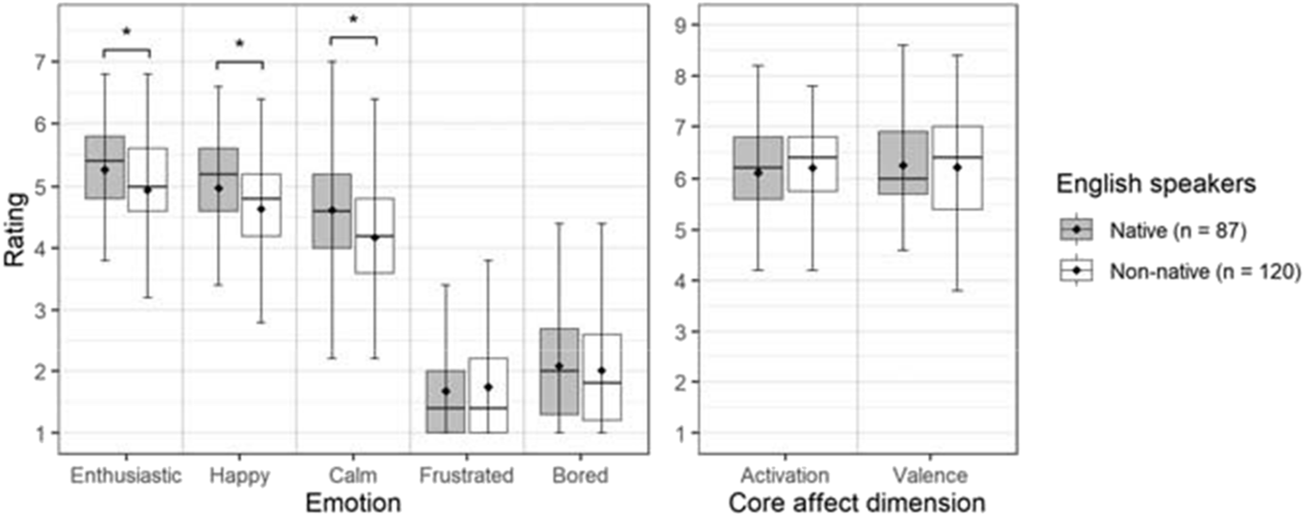
When looking at the general emotion recognition ability within the two groups, there were no significant differences in the native English speakers' ability to recognise enthusiasm and calmness ($t(86) = -0.39$, $p = 0.696$), but non-native speakers were better at recognising calmness than enthusiasm ($t(119) = 3.46$, $p < 0.001$, mean difference = 0.40, 95% CI [0.17–0.63], $d = 0.32$, 95% CI [0.13–0.50]).

The bored ratings of the bored clip were similar between the two groups, $F(4, 200) = 1.59$, $p = 0.179$, while there was a main effect in the frustration rating, $F(4, 200) = 2.66$, $p = 0.034$, but not a significant difference after controlling for interest in the video content. No main effects were found for activation ($F(4, 200) = 0.57$, $p = 0.684$) or valence ratings ($F(4, 200) = 0.96$, $p = 0.430$). Lastly, we found a main effect in the frustrated clip for frustration, $F(4, 200) = 5.26$, $p < 0.001$, but it was not attributable to belonging to the native or non-native group. Neither activation ($F(4, 200) = 0.85$, $p = 0.497$) nor valence ($F(4, 200) = 1.38$, $p = 0.241$) showed a main effect.

Discussion

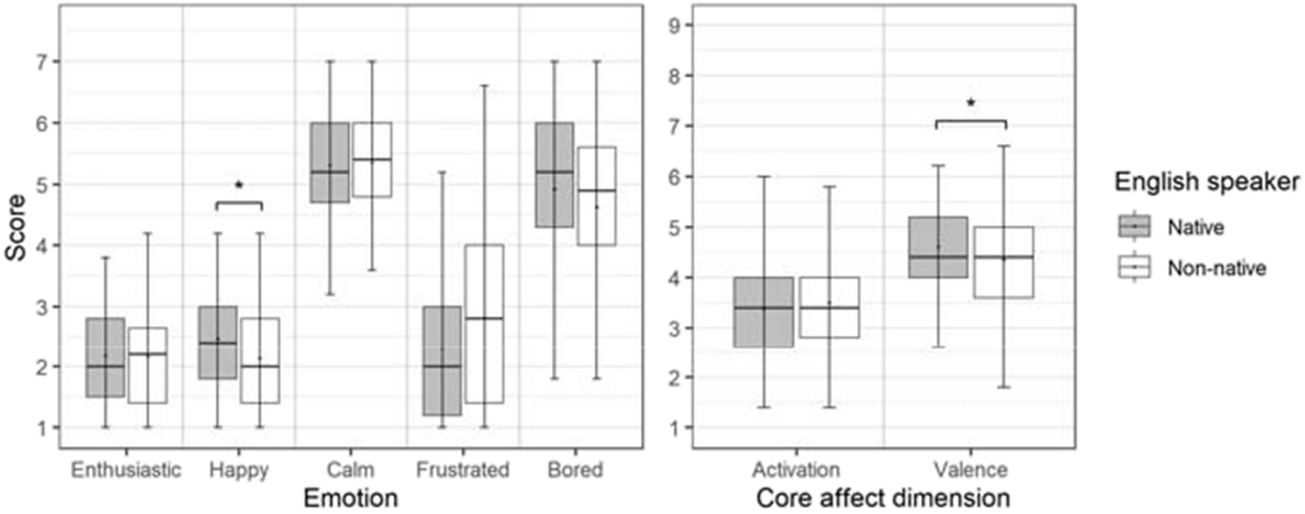
Consistent with the cognitive-affective model of e-learning (Lawson et al., 2021b; Mayer, 2020), this study demonstrated that learners could recognize the emotional tone of an instructor in an educational video from their voice alone, which solidifies the first link of the theoretical model. Because conveying positive affective states appears to be a beneficial instruction design strategy (Liew et al., 2020; Loderer et al., 2020), we used five videos of calm and enthusiastic narrations to specifically test whether learners could distinguish between positive emotions with varying levels of activation, and added two videos depicting a negative emotion (boredom and frustration) for comparison. Similar to previous research (Lawson & Mayer, 2021; Lawson et al., 2021b), participants were better able to differentiate between emotions with a different valence but the same activation level (e.g. enthusiasm from frustration) than between emotions with the same valence but a different activation level (enthusiasm from happiness and calmness).

However, participants were still sensitive to the differences between vocal expressions of positive emotions with a high activation level (enthusiasm), a low activation level



Note. Asterix indicates significant differences between groups; point indicates the mean.

Fig. 2 Ratings on the items assessing emotion and core affect dimensions for the enthusiastic videos split between native and non-native English speakers. Asterisks indicate significant differences between groups; point indicates the mean



Note. Asterix indicates significant differences between groups; point indicates the mean.

Fig. 3 Ratings on the items assessing emotion for the calm videos split between native and non-native English speakers. Asterisks indicate significant differences between groups; point indicates the mean

(calmness) and a medium activation level (happiness). Although there were no videos that specifically expressed happiness, learners rated enthusiastic videos as significantly more enthusiastic than calm and happy and calm videos as significantly more calm than happy or enthusiastic. In general, these differences were greater for the calm videos.

Results were different only for the frustrated video, where participants did not recognize that it expressed frustration. This was not necessarily due to their inability to do so, but possibly due to using only one clip for frustration or our narrator's portrayal of the emotion.

Most previous studies have examined the effects of the portrayed affective state of the video instructor on learning processes and outcomes by using an onscreen agent that conveys social cues through their facial expressions and body language in addition to the voice (e.g. Horovitz & Mayer, 2021; Schneider et al., 2022). In contrast, this study adds to the small but growing literature showing that voice is a powerful source of social and emotional information on its own even in the context of educational videos (Schneider et al., 2021) and that adding onscreen agents may not be necessary if the goal is to convey positive emotions (Lawson & Mayer, 2021).

Another important contribution of this work is the addition of non-native language speakers and their comparison to native speakers in recognizing portrayed emotions. Most online instructional videos are produced in English, which means that many learners around the world watch them in their foreign language, with varying levels of English proficiency. Yet, most studies of multimedia learning and emotional design use instructional videos spoken in learners' native languages. Testing instructional design interventions on more international audiences is critical, as the lack of verbal listening skills, or even just processing in a foreign language, can interfere with the learning process, meaning that learners watching educational videos in a foreign language might need additional aids that are unnecessary or even detrimental to those who learn in their native language (Davis & Vincent, 2019). In fact, many studies have shown the effect language background can have on learning from digital media (e.g. Lee & Mayer, 2018; Mayer & Fiorella, 2014; Schneider et al., 2015).

Consistent with previous research, both native and non-native English speakers were successful in recognizing the target emotion from the narrator's voice (Laukka & Elfenbein, 2021), and their ratings were generally very similar. However, compared to the native speakers, the non-native speaker group rated the enthusiastic videos as significantly less enthusiastic, calm and happy and the calm videos as significantly less happy and pleasant. These results are in line with the dialect theory (Elfenbein, 2013) in that they show a consistent difference in emotion recognition between native and non-native speakers. In our case, non-native speakers perceived the educational narrations as less positive than those who listened to the clips in their native language, which builds on the knowledge that not only accuracy but also other variables related to emotion recognition may be affected (Jiang et al., 2015). Non-native speakers were also better at recognising calmness than enthusiasm, indicating a possible misunderstanding when interpreting activated positive vocal cues in a different language. However, English proficiency did not play a significant role in recognizing the affective tone of the instructor, which differs from previous findings (Bhatara et al., 2016; Jiang et al., 2015). Because past research on emotion recognition has focused almost exclusively on negative emotions (Sauter, 2010), we do not know whether this is specific to positive emotions or to our sample.

Limitations

It should be noted that the results could be highly dependent on how our (young American female) narrator portrayed the emotions and on our non-random sample. For example, most native English-speaking participants were from the UK, not the USA, so differences in dialect may have affected the results. Our non-native speakers' sample was also highly proficient in English, lowering the generalizability of our results to only a part of those who understand the foreign language. Future research should use a more varied sample and clips with different topics and types of instructors, so that results can be generalized to different individuals, cultures, languages and content areas. Little research has been done on the effects of individual differences between narrators — gender and age of the instructor could play a role in how emotional tone is perceived through voice. In addition, although we used five videos to represent enthusiasm and calmness, we used only one for boredom and frustration, making the results for these two emotions less reliable. The frustration clip in particular was problematic because participants did not perceive it as such, which could be due to the complexity of the emotion itself or the result of our portrayal.

Conclusion

This study is a first step in providing evidence for the cognitive-affective model of e-learning. Specifically, it was shown that voice alone is enough to convey the instructor's emotions to learners all around the world who watch instructional videos primarily in English. While there were some subtle differences, both native and non-native English speakers recognized the portrayed emotion, and the emotion recognition ability was not dependent on the learners' English language skills.

The implication of the study is that instructors need to be aware of how their emotional stance is conveyed through their voice when creating educational videos. However, it is still unclear how emotional cues in the voice alone can affect learning from disembodied educational videos. Future studies should test the theoretical model further and check whether emotions expressed through voice prosody alone influence the cognitive, affective and learning processes of those who learn by watching videos. Recognizing emotion does not necessarily mean that learners feel the emotion in question or that it affects their learning, so further research on the cognitive-affective model of e-learning is needed. Subsequent studies should include measures of cognitive load to better understand the reasons for the differences between native and non-native speakers, and questions measuring knowledge retention and transfer to investigate whether and how different emotional tones expressed through voice prosody affect learning from multimedia materials in a language that is not our native one. In a globalized world, instructional videos in English are a popular learning tool for learning about various topics, so research about this educational technology should continue to test design guidelines that are inclusive of people all around the world.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10212-023-00715-3.

Acknowledgements The authors would like to thank Dr. Dean Lipovac for his help with the figures and feedback on the manuscript.

Funding This work was supported by the European Commission by funding the InnoRenew project (Grant Agreement #739574) under the H2020 Widespread-2-Teaming programme and by the Republic of Slovenia (investment funding from the Republic of Slovenia and the European Regional Development Fund). Part of this work was conducted during the WoodLCC project, which is supported under the umbrella of ERA-NET Cofund ForestValue.

Data availability Supplementary materials and the data underlying this article are available in Zenodo at https://zenodo.org/record/6719862

Declarations

Ethics statement Ethics approval was not required for this study. Participants were provided with an informed consent informing them about the purpose of the research, voluntary participation and data usage.

Conflict of interest The authors declare no competing interests.

References

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

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2020). Does the effect of enthusiasm in a pedagogical agent's voice depend on mental load in the learner's working memory? *Computers in Human Behavior, 112*, 106483. https://doi.org/10.1016/j.chb.2020.106483

Bhatara, A., Laukka, P., Boll-Avetisyan, N., Granjon, L., Elfenbein, H. A., & Bänziger, T. (2016). Second language ability and emotional prosody perception. *PLoS ONE, 11*(6), 1–13. https://doi.org/10.1371/journal.pone.0156855

Brom, C., Hannemann, T., Stárková, T., Bromová, E., & Děchtěrenko, F. (2017). The role of cultural background in the personalization principle: Five experiments with Czech learners. *Computers & Education, 112*, 37–68. https://doi.org/10.1016/j.compedu.2017.01.001

Collins, M. L. (1978). Effects of enthusiasm training on preservice elementary teachers. *Journal of Teacher Education, 29*(1), 53–57. https://doi.org/10.1177/002248717802900120

Davis, R. O., & Vincent, J. (2019). Sometimes more is better: Agent gestures, procedural knowledge and the foreign language learner. *British Journal of Educational Technology, 50*(6), 3252–3263. https://doi.org/10.1111/bjet.12732

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review, 5*(1), 90–96. https://doi.org/10.1177/1754073912451332

Faculty of Social Sciences, U. of L. (2022). *1KA* (21.11.16). https://www.1ka.si

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

Javora, O., Hannemann, T., Stárková, T., Volná, K., & Brom, C. (2018). Children like it more but don't learn more: effects of esthetic visual design in educational games. *British Journal of Educational Technology*, *50*(4), 1942–1960. https://doi.org/10.1111/bjet.12701

Jiang, X., Paulmann, S., Robin, J., & Pell, M. D. (2015). More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 597–612. https://doi.org/10.1037/xhp0000043

Laukka, P., & Elfenbein, H. A. (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review, 13*(1), 3–11. https://doi.org/10.1177/1754073919897295

Lawson, A. P., & Mayer, R. E. (2021). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education, 32*, 971–990. https://doi.org/10.1007/s40593-021-00282-y

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021a). The positivity principle: Do positive instructors improve learning from video lectures? *Educational Technology Research and Development, 69*, 3101–3129. https://doi.org/10.1007/s11423-021-10057-w

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior, 114*, 106554. https://doi.org/10.1016/j.chb.2020.106554

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021c). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education, 31*, 134–153. https://doi.org/10.1007/s40593-021-00238-2

Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology, 32*(5), 648–654. https://doi.org/10.1002/acp.3436

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE : A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Liew, T. W., Mat Zin, N. A., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-Centric Computing and Information Sciences, 7*(1), 9. https://doi.org/10.1186/s13673-017-0089-2

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

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: a systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction, 70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Lorette, P., & Dewaele, J. (2018). Emotion recognition ability across different modalities: The role of language status (L1/LX), proficiency and cultural background. *Applied Linguistics Review, 11*(1), 1–26. https://doi.org/10.1515/applirev-2017-0015

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). In Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R. E. (2020). Searching for the role of emotions in e-learning. *Learning and Instruction, 70*, 101213. https://doi.org/10.1016/j.learninstruc.2019.05.010

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158. https://doi.org/10.1111/j.1365-2729.2006.00170.x

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177–213. https://doi.org/10.1207/S1532690XCI1902_02

Morningstar, M., Gilbert, A. C., Burdo, J., Leis, M., & Dirks, M. A. (2021). Recognition of vocal socioemotional expressions at varying levels of emotional intensity. *Emotion, 21*(7), 1570–1575. https://doi.org/10.1037/emo0001024

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In *Emotions, Technology, Design, and Learning* (pp. 131–161). Elsevier Academic Press. https://doi.org/10.1016/b978-0-12-801856-9.00007-4

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0). Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org

Rey, G. D., & Steib, N. (2013). The personalization effect in multimedia learning: The influence of dialect. *Computers in Human Behavior, 29*(5), 2022–2028. https://doi.org/10.1016/j.chb.2013.04.003

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

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science, 19*(1), 36–40. https://doi.org/10.1177/0963721409359290

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015). Introducing the familiarity mechanism: A unified explanatory approach for the personalization effect and the examination of youth slang in multimedia learning. *Computers in Human Behavior, 43*, 129–138. https://doi.org/10.1016/j.chb.2014.10.052

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction, 44*, 65–73. https://doi.org/10.1016/j.learninstruc.2016.03.002

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2021). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educational Psychology Review, 34*, 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schneider, S., Krieglstein, F., Beege, M., & Daniel, G. (2022). The impact of video lecturers' nonverbal communication on learning – an experiment on gestures and facial expressions of pedagogical agents. *Computers & Education, 176*, 104350. https://doi.org/10.1016/j.compedu.2021.104350

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

The jamovi project. (2021). *jamovi* (1.6). https://www.jamovi.org

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology, 8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498. https://doi.org/10.1037/a0026609

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*(10), 1559–1579. https://doi.org/10.1016/j.pragma.2005.04.012

Wong, R. M., & Adesope, O. O. (2020). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review, 33*, 1–29. https://doi.org/10.1007/s10648-020-09545-x

Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology, 47*(6), 1304–1323. https://doi.org/10.1111/bjet.12324

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Nežka Sajinčič. InnoRenew CoE, Izola, Slovenia. nezka.sajincic@innorenew.eu.

Current themes of research:

Multimedia learning. Emotional design. Instructional design. Educational technology. Gamification of learning.

Most relevant publications in the field of Psychology of Education:

Sajinčič, N., Sandak, A., & Istenič, A. (2022, July 10–15). Making knowledge about renewable materials accessible and engaging with educational videos based on instructional design. 65th SWST International Convention, Kingscliff, NSW, Australia. https://doi.org/10.5281/zenodo.7568615 .

Sajinčič, N., Istenič, A., & Sandak, A. (2022, May 10–12). Learning about NDSS through video - Evidence-based guidelines for effective instructional videos for a smooth transition into industry. 1st sensorFINT International Conference: Non-Destructive Spectral Sensors advances and future trends, Izola, Slovenia. https://doi.org/10.5281/zenodo.6554047 .

Sajinčič, N., Sandak, A., & Istenič, A. (2022). How do Slovenian educators feel about gamification? Interested to know more. Education and Self Development, 17(1), 99–109. https://doi.org/10.26907/esd.17.1.09 .

Sajinčič, N., Sandak, A., & Istenič, A. (2022). Pre-service and in-service teacher's view on gamification. International Journal of Emerging Technologies in Learning, 17(3), 83–103. https://doi.org/10.3991/ijet.v17i03.26761 .

Anna Sandak. InnoRenew CoE, Izola, Slovenia; Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia. anna.sandak@innorenew.eu.

Current themes of research:

Multi-aspect characterization of ligno-cellulosic materials. non-destructive testing. evaluation of degradation level of wood and wooden based products and characterization of bio-based materials with spectroscopic techniques.

Most relevant publications in the field of Psychology of Education:

Sajinčič, N., Sandak, A., & Istenič, A. (2022, July 10–15). Making knowledge about renewable materials accessible and engaging with educational videos based on instructional design. 65th SWST International Convention, Kingscliff, NSW, Australia. https://doi.org/10.5281/zenodo.7568615.

Sajinčič, N., Istenič, A., & Sandak, A. (2022, May 10–12). Learning about NDSS through video - evidence-based guidelines for effective instructional videos for a smooth transition into industry.

1st sensorFINT International Conference: Non-Destructive Spectral Sensors advances and future trends, Izola, Slovenia. https://doi.org/10.5281/zenodo.6554047.

Sajinčič, N., Sandak, A., & Istenič, A. (2022). How do Slovenian educators feel about gamification? Interested to Know More. Education and Self Development, 17(1), 99–109. https://doi.org/10.26907/esd.17.1.09.

Sajinčič, N., Sandak, A., & Istenič, A. (2022). Pre-service and in-service teacher's view on gamification. International Journal of Emerging Technologies in Learning, 17(3), 83–103. https://doi.org/10.3991/ijet.v17i03.26761.

Amy Simmons. InnoRenew CoE, Izola, Slovenia; Andrej Marušič Institute, University of Primorska, Koper, Slovenia. amy.simmons@innorenew.eu;.

Current themes of research

Effective science communication and outreach. Gender inequality in science, technology, engineering and mathematics.

Most relevant publications in the field of Psychology of Education:

Sajinčič, N., Gordobil, O., Simmons, A., & Sandak, A. (2021). An exploratory study of consumers' knowledge and attitudes about lignin-based sunscreens and bio-based skincare products. Cosmetics, 8(3), 1-20, https://doi.org/10.3390/cosmetics8030078.

Ice, G., Hale, C., Light, J., Muldoon, A., Simmons, A. (2021). Understanding dissolved oxygen concentrations in a discontinuously perennial stream within a managed forest. Forest Ecology and Management, 479, 1-14.

Simmons, A. (2019, October 20–25). The forest for the trees: understanding the experiences of female PhDs in forestry-related academia. Proceedings of the 62nd International Convention of Society of Wood Science and Technology, Yosemite, California, USA.

Andreja Istenič. Faculty of Education, Koper, Slovenia. andreja.starcic@pef.upr.si.

Current themes of research:

Educational technology. Media and communication. Teacher education. Higher education. Research evaluation. Interdisciplinary research .

Most relevant publications in the field of Psychology of Education:

Sajinčič, N., Sandak, A., & Istenič, A. (2022). How do Slovenian educators feel about gamification? Interested to know more. Education and Self Development, 17(1), 99–109. https://doi.org/10.26907/esd.17.1.09.

Sajinčič, N., Sandak, A., & Istenič, A. (2022). Pre-service and in-service teacher's view on gamification. International Journal of Emerging Technologies in Learning, 17(3), 83–103. https://doi.org/10.3991/ijet.v17i03.26761.

Istenič, A., Rosanda, V., Volk, M., & Gačnik, M. (2023). Parental perceptions of child's play in the post-digital era: parents' dilemma with digital formats Informing the kindergarten curriculum. Children, 10(1), 1–22. https://doi.org/10.3390/children10010101.

Rosanda, V., Kavčič, T., & Istenič, A. (2022). Digital devices in early childhood play: digital technology in the first two years of Slovene toddlers' lives. Education and self-development, 3(17), 83–99. https://doi.org/10.26907/esd.17.3.08.

Istenič, A. & Lebeničnik, M. (2022). How communion and agentic beliefs predict technology-supported formal and informal learning: the implications for educational technology. International Journal of Emerging Technologies in Learning, 17(4), 171–193. https://doi.org/10.3991/ijet.v17i04.27249.

Istenič, A., Bratko, I., & Rosanda, V. (2021). Are pre-service teachers disinclined to utilise embodied humanoid social robots in the classroom? British journal of educational technology, 52(6), 2340–2358. https://doi.org/10.1111/bjet.13144.

Kukanja-Gabrijelčič, M., Antolin, U., & Istenič, A. (2021). Teacher's social and emotional competences: a study among student teachers and students in education science in Slovenia. European journal of educational research, 10(4), 2033–2044. https://doi.org/10.12973/eu-jer.10.4.2033.