Full length article

Recognizing the emotional state of human and virtual instructors

Alyssa P. Lawson [a,*], Richard E. Mayer [a], Nicoletta Adamo-Villani [b], Bedrich Benes [b], Xingyu Lei [b], Justin Cheng [b]

[a] Department of Psychological and Brain Sciences, University of California, Santa Barbara, USA
[b] Computer Science Department, Purdue University, USA

ARTICLE INFO

ABSTRACT

Students' learning from an instructional video could be affected by the instructor's emotional stance during a lesson. A first step in investigating this emotional design hypothesis is to determine whether students perceive the emotions displayed by an instructor during an instructional video. Building on Russell's (1980, 2003) model of core affect and the media equation theory (Reeves & Nass, 1996) this study investigated how well participants were able to perceive different emotions portrayed by a human and virtual instructor (i.e., animated pedagogical agent) in a video lecture on statistics. Participants were shown short video clips of either a human instructor or virtual instructor displaying four different emotions: happy, content, bored, and frustrated. The participants were asked to rate how well each video clip displayed each of those four emotions. Participants were able to recognize each of the emotions displayed by the instructor but were much better at distinguishing between positive (happy and content) and negative (bored and frustrated) emotions than between active (happy and frustrated) and passive (content and bored) emotions. Furthermore, participants were able to recognize the emotions of the instructor for both the human instructor and the animated agent. However, emotions that involved higher activity (happy and frustrated) were more easily recognized in a human instructor than an animated agent. This research shows that learners are aware of the emotions being portrayed by an instructor, both human and animated agent, and establishes the first link in the chain between how the emotional tone displayed by an instructor affects learning outcomes.

1. Introduction

1.1. Objective and rationale

Consider a learning scenario in which a student views an instructional video showing an instructor standing next to a slide as she lectures, such as exemplified in Fig. 1. An interesting issue in affective science and affective computing involves the degree to which the emotional state of the instructor affects student learning, but before we can address that issue, a preliminary question concerns whether students are even able to recognize the emotional state of the instructor. The primary goal of the present study is to determine whether people who view a short video lecture on a statistical procedure are able to detect the degree to which the instructor exhibits a happy, content, frustrated, or bored emotional state. To address this goal, we created four versions of a lecture on statistics–involving the same instructor, script, and slides–in which the instructor (an actor) exhibited either a

happy, content, frustrated, or bored emotional tone through her body stance, gestures, facial expression, and voice. We asked adult participants to view two clips from each version and rate the degree to which the instructor appeared to be happy, content, frustrated, or bored. If participants are able to recognize the instructor's emotional state, this should be reflected in their ratings: the happy instructor should be rated higher on the happy scale than on each of the other scales; the content instructor should be rated higher on the content scale than on each of the other scales; the frustrated instructor should be rated higher on the frustrated scale than on the other scales; and the bored instructor should be rated high on the bored scale than on the other scales.

A secondary goal of the present study is to determine whether participants are equally able to recognize the emotional state of human instructors in instructional videos and virtual instructors in animated lessons, when they say the same things and refer to the same slides. To address this issue we created animated pedagogical agents who mimicked the facial expressions, body stance, and gestures of the human

instructor and used the same voice. An example frame is shown in Fig. 2. We asked a different group of adult participants to view the same two clips from each version and rate the degree to which the instructor appeared to be happy, content, frustrated, or bored. If participants treat virtual instructors like human instructors, we expect them to show the same pattern of ratings as was reported for human instructors and we expect the level of ratings to be equivalent for human and virtual instructors (i.e., the happy rating for happy instructor, the content rating for the content instructor, the frustrated rating for the frustrated instructor, and the bored rating for the bored instructor should be indistinguishable for human and virtual instructors). Alternatively, if people see the virtual instructors as somewhat less human-like, their emotional state ratings of the instructor's emotional state may be lower than for human instructors.

1.2. Research and theory on the Instructor's emotional state

Although there are several systems for classifying emotional expression, especially facial expression (Ekman & Friesen, 2003; Ekman et al., 2013), we focus on Russell's model of core affect (Russell, 1980, 2003) because it has been useful in classification of achievement emotions (Harley et al., in press; Pekrun, 2016; Pekrun & Stephens, 2010) and in the instructional design of onscreen characters (Loderer et al., in press; Plass & Kaplan, 2016) and thus has potential relevance for categorizing the perceived emotional states of instructors. Fig. 3 shows an adapted version of Russell's (1980, 2003) model of core affect, which is based on two orthogonal dimensions: valence, running from displeasure on the left to pleasure on the right (or more simply, from negative valence to positive valence); and arousal, running from activation on the top and deactivation on the bottom (or more simply, from active to passive). These dimensions generate four quadrants from which we abstracted four emotional states that could be relevant to an instructor: happy (which represents positive valence and active arousal), content (which represents positive valence and passive arousal), frustrated (which represents negative valence and active arousal), and bored (which represents negative valence and passive arousal). Other descriptors also apply to each quadrant (Loderer et al., 2019; Pekrun, 2006, 2016; Pekrun & Perry, 2014; Russell, 1980, 2003), but we use these four as representatives in our study.

Although classic theories of e-learning such as Cognitive Load Theory (Paas & Sweller, 2014; Sweller et al., 2011) and the Cognitive Theory of Multimedia Learning (Mayer, 2014, in press-a) focused mainly on the cognitive processes during learning, some investigators have also attempted to incorporate affective processes during learning such as

Plass and Kaplan's (2016) Integrated Cognitive Affective Model of Learning with Media and Moreno and Mayer's (2007) Cognitive-Affective Model of Multimedia Learning. These attempts are in line with calls to incorporate emotion into theory of multimedia learning (Mayer, in press-b; Plass & Kalyuga, 2019). As an example, Fig. 4 presents an example of an adaptation of Mayer's (in press-b) cognitive-affective model of e-learning that consists of five basic components: (1) an e-learning episode (such an online multimedia lesson with an instructor displaying positive emotion) causes (2) the learner to recognize the instructor's emotional stance, (3) which primes an affective response in the learner towards the instructor (such feeling positive about the instructor), (4) which affects cognitive processing during learning (such as the degree to which the learner is motivated to engage in deep processing), (5) which, in turn, affects the learning outcome (as measured by posttest performance).

As research on the role of emotion in e-learning is now becoming a larger field of investigation, the research has produced some ambiguous findings, making more research necessary (i.e., Knörzer et al., 2016; Loderer et al., in press; Loderer et al., 2019; Mayer, in press-b; Plass & Kalyuga, 2019; Schneider et al., 2016). One encouraging strand of research on emotion in e-learning involves the emotional design of online learning material (Brom et al., 2018; Mayer & Estrella, 2014; Plass et al., in press; Plass et al., 2014; Plass & Kaplan, 2016; Um et al., 2012; Wong & Adesope, in press). For example, in multimedia lessons on how viral infection works, students gave more positive affective ratings and scored higher on posttests when the characters in the lesson (such as a virus or host cell) were portrayed in warm colors with facial expressions (Mayer & Estrella, 2014; Plass et al., 2014; Um et al., 2012). In this case, the emotional stance of the instructor in e-learning materials (corresponding to the first box in Fig. 4) affects the learner's perception of the instructor's emotional stance (corresponding to the second box in Fig. 4) improves the learner's feelings about the instructor as measured by ratings (corresponding to the third box in Fig. 4) and motivation to engage in deep cognitive processing as measured by ratings (corresponding to the fourth box in Fig. 4) which, in turn, improves learning outcomes (corresponding to the fifth box in Fig. 4).

A recent review of emotional design of multimedia lessons confirmed that adding emotional design features intended to portray positive emotional tone had a positive effect on improving learning, with an effect size of $d = 0.33$ for transfer test performance (Brom et al., 2018). An updated and broader meta-analysis also found a positive effect of emotional design on transfer test performance with $g = 0.27$ based on 38 comparisons (Wong & Adesope, in press).

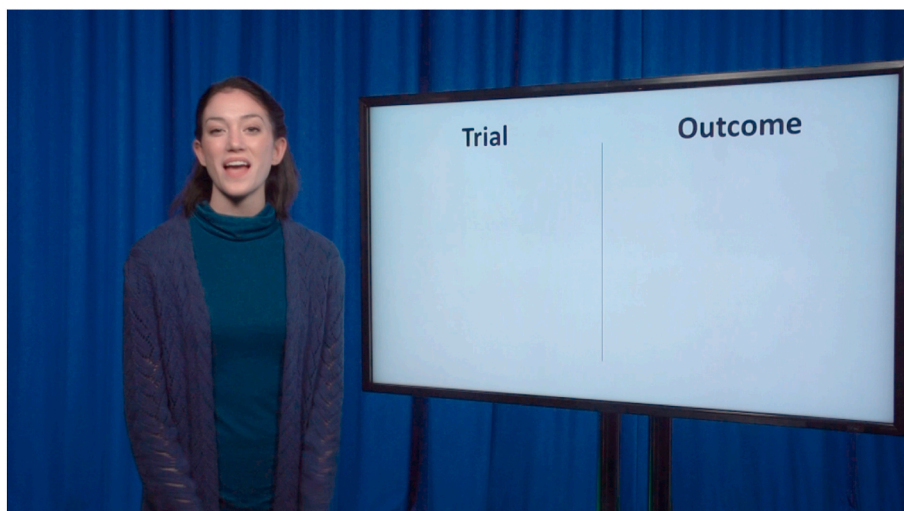In a study that focuses mainly on the first link in cognitive-affective

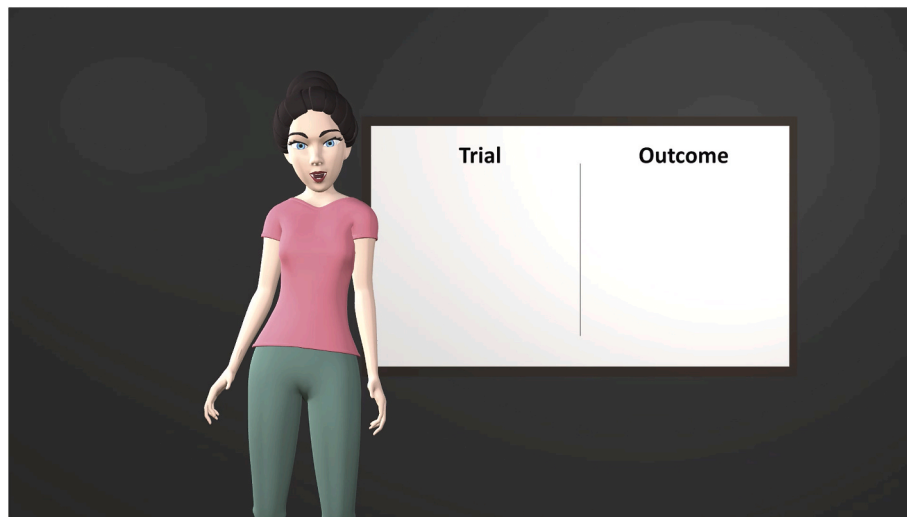**Fig. 1.** *Image from Human Instructor Video.*

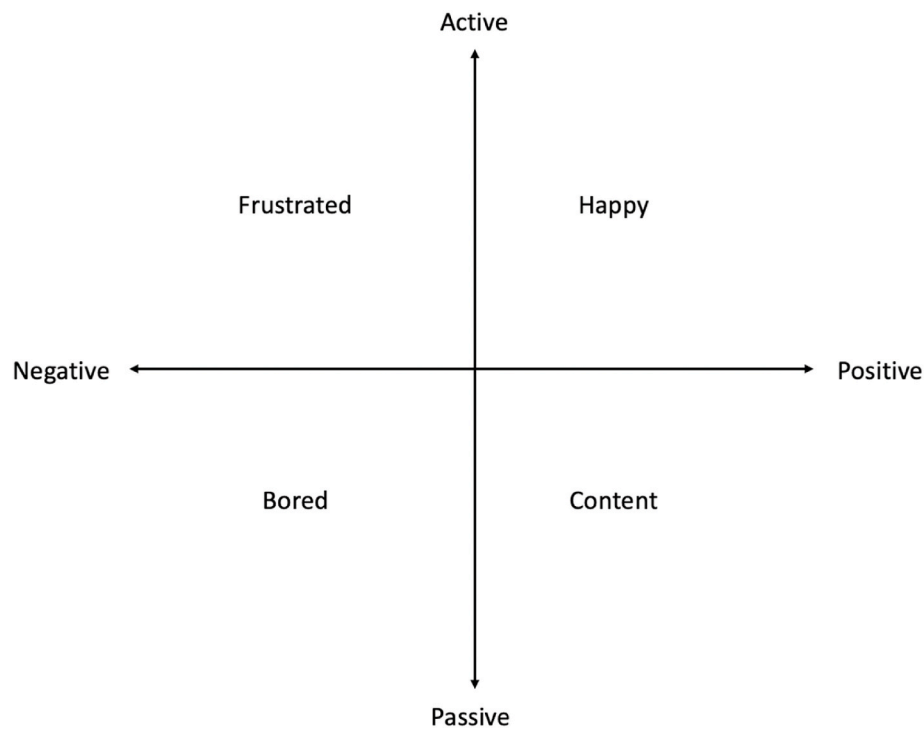**Fig. 2.** *Image from Virtual Instructor Video.*

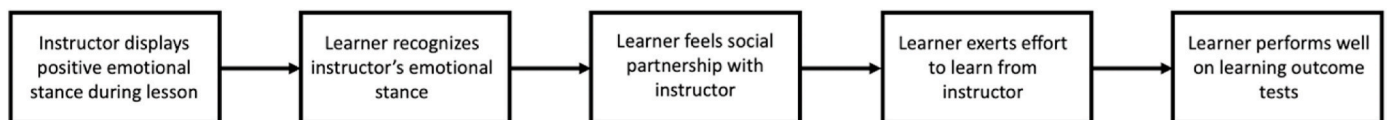**Fig. 3.** *Model of Core Affect Adapted from Russell (2003).*

Fig. 4. An Example of the Cognitive-affective Model of e-Learning, When the Instructor Displays Positive Emotion Such as Being Happy.

model of e-learning, Plass et al. (in press) varied the emotional tone of game characters for an online computer game by varying their facial expression and color. Students reported happy emotions for characters with happy facial expressions and warm colors, whereas students reported sad emotions for characters with sad or neutral facial expressions and cold colors. These findings encourage the proposal that people can recognize the emotional tone of onscreen characters.

Exemplary evidence concerning the first link involving student recognition of the emotional tone of a lesson comes from a study by Uzum and Yildirim (2018) in which students displayed stronger positive emotional arousal via biometric measures for multimedia lessons containing onscreen characters who displayed positive rather than neutral facial expression. Another piece of evidence concerning the first link is that students who receive multimedia lessons spend more time looking

at onscreen agents that display positive emotion than those that display neutral emotion (Park et al., 2015). Finally, Kramer et al. (2013) reported that people who engaged in an 8-min conversation with an agent who smiled (thereby indicating positive emotional stance) spent more time smiling themselves than people who had a communication with an agent who did not smile. These studies provide encouraging preliminary evidence that people are able to recognize and respond to the perceived emotional state of the instructor in a multimedia lesson. However, this previous research has focused on understanding how different pedagogical stimuli can elicit certain emotions in the learners. The current research expands on this by attempting to understand more deeply if and how learners recognize emotions of pedagogical stimuli.

1.3. Research and theory on leaning with human and virtual instructors

People can easily form a social relationship with a computer and treat a computer as if it is human. This is the thesis underpinning the media equation hypothesis concerning communication (Reeves & Nass, 1996) as well as social agency concerning multimedia learning (Mayer, 2014, in press-a). For example, Nass and Brave (2005) show how machine interfaces are capable of expressing emotions such as happiness just as well humans can. When the focus is on learning from multimedia lectures, research shows that people learn better from an animated pedagogical agent that engages in human-like gesturing, uses conversational language, and speaks in an appealing human voice while lecturing (Mayer, 2014, in press-a). When onscreen agents have these features, learners tend to report liking the agent and feeling they have a social connection (Mayer, 2014, in press-a). In the present study, we are interested in a direct comparison between how participants relate to human and virtual instructors who are teaching the same content with the same emotional tone. These findings encourage the prediction that people will show equivalent ratings of emotional tone for video lectures with equivalent virtual and human instructors.

1.4. Hypotheses

In the present study, students are shown a set of clips from a video lesson on the statistical concept of binomial probability; the clips contain the same script being rendered by the same instructor but displaying a happy, frustrated, content, and bored emotional tone, respectively. In addition, there are versions of each of these clips in which the instructor is a human and in which the instructor is an onscreen agent mimicking the same gestures, body stance, facial expression, and voice as the human instructor. This project involves an initial step in determining the role of the instructor's emotional tone in video lessons by examining the extent to which learners are aware of the instructor's emotional tone.

According to the emotional awareness hypothesis, based on the first link in the cognitive-affective theory of e-learning, learners recognize the emotional tone of instructors in video lectures. This hypothesis leads to the prediction that participants will give higher ratings to the emotion displayed by the instructor than each of the other three emotions (hypothesis 1). In particular, for the happy instructor, participants will give higher ratings of happy than each of the other three emotions (predication 1a); for the content instructor, participants will give higher ratings of content than of each of the other three emotions (hypothesis 1b); for the frustrated instructor, participants will give higher ratings of frustrated than of each of the other three emotions (hypothesis 1c); and for the bored instructor, participants will give higher ratings of bored than of each of the three emotions (hypothesis 1d).

According to the media equation hypothesis, people accept computers as social partners as if they were human. This leads to the prediction that participants' ratings for the emotional tone of human instructors will be equivalent to participants' ratings for the emotional tone of corresponding virtual instructors (hypothesis 2). In particular, participants will rate the happy virtual instructor as just as happy as the

happy human instructor (hypothesis 2a); participants will rate the content virtual instructor as just as content as the content human instructor (hypothesis 2b); participants will rate the frustrated virtual instructor as just as frustrated as the frustrated human instructor (hypothesis 2c); and participants will rate the bored virtual instructor as just as bored as the bored human instructor (hypothesis 2d).

2. Method

2.1. Participants and design

The participants were 202 adults recruited from Amazon Mechanical Turk (Mturk). Mturk has been shown to be a viable and reliable way to conduct studies online (Paolacci et al., 2010). Only participants currently in the United States were recruited from this study and 196 of them were born in the United States. Additionally, participants were only allowed to participate if their HIT approval rate was greater than 93%, which is the proportion of completed tasks that were approved by previous requesters. To remove bots from participating, there was also a reCAPTCHA item added to the survey.

The mean age of the participants was 36.29 years ($SD = 10.77$) and 77 of them were women. Of all the participants, 142 classified themselves as "White/Caucasian," 28 as "Black/African/African-American," 11 as "Hispanic/Latinx," 9 as "Asian/Asian-American," 6 as multiple races/ethnicities, 5 as "Native American," and 1 as "Other."

The experiment used a 2 (between-subjects) x 4 (within-subjects) mixed factorial design, with the between-subjects factor being the type of instructor (human or animated pedagogical agent) and the within-subjects factor being the emotional tone of the instructor (happy, content, bored, or frustrated). There were 99 participants in the human instructor group (who saw 8 videos with a human instructor) and 103 participants in the virtual instructor group (who saw 8 video clips with an animated pedagogical agent (APA). All participants in each group saw the same eight clips, consisting of two clips of each of four emotions with either a human instructor or an animated pedagogical agent instructor.

2.2. Materials

The materials were all computer-based on Qualtrics and included 8 video clips with a human instructor or 8 video clips with an APA instructor as well as rating surveys for each video clip and a postquestionnaire.

2.2.1. Video clips

There were 8 video clips from a lesson on binomial probability taught by a human instructor, and 8 video clips involving the same script and graphics taught by an APA instructor. Within each set of video clips, four involved a 31–42 s segment (depending on emotion condition) covering an example of when one could use binomial probability, and four involved a 30–49 s segment (depending in emotion condition) covering the definition and an example of a sequence. The script for the first segment was: "Hi everyone. Imagine that you are trying to impress your friends with your ability to predict what will happen if you roll a die a certain number of times. For example, suppose you win if you roll a 5 or 6 and you lose if you roll a 1, 2, 3, or 4. Let's say you roll the die 5 times and you win 2 times and lose 3 times. What exactly is the probability of that happening? Today, I will help you understand how to answer questions like this one. This is called binomial probability." The script for the second segment was: "The next concept you need to know is sequence. A sequence is what happens when you conduct several trials, one after another, like rolling a die 5 times in a row. For each trial, we have either a success or a failure, so the sequence reports what occurred. For example, say we rolled a die 5 times in a row and rolled a 2, then a 4, then a 6, then a 2, then a 5. The sequence would be failure, failure, success, failure, success."

For each instructor type (human or virtual) there were four versions of the first clip and four versions of the second clip: happy, content, frustrated, and bored. To create the human videos, a 21-year old female actor from a university's Theater Department was recorded displaying these four different emotions while reading a script from a teleprompter in front of the camera. She was told to vary her gestures, facial expression, body stance, and voice in accord with each of the emotions. The lessons were recorded in a university's TV production studio. The experimenters monitored each segment of the lecture during the filming and asked for a retake when the script was not followed accurately or the emotional tone did not seem appropriate.

Once these videos with a human instructor were complete, they were turned into corresponding videos that had a virtual instructor teaching the lesson. The virtual instructor resembled the human instructor, had the same voice as the human instructor, and exhibited some of the same gestures, facial expressions, and body stance as the human instructor. To turn the human instructor videos into the virtual instructor videos, a custom Unity 3D platform was developed to generate the virtual instructor videos. The agents are commercially available 3D character rigs whose joint structure was modified in order to be compatible with Unity's character animator feature. The agents' gestures were motion captured and manually blended together; the agents' lip-sync was generated with a Unity script. Facial deformations were produced using joint deformers and the agents' facial animations were manually keyframed. Camera angle, background, and lighting were kept the same in every clip. The links to clips of the 8 video lessons are provided in Appendix A.

How can we validate the success of the human instructor and virtual instructor in portraying the desired emotions in the instructional videos? This task is the central goal of the present study. We provided the actress with guidelines for how to display each emotion, and we provided feedback to her during rehearsal based on our judgment, but the ultimate test involves what users think about the emotions displayed by the human instructor. We built the virtual instructor to mimic the human instructor (using the original human voice), and adjusted the resulting animations based on our judgement, but again, the ultimate test involves what users think about the emotions displayed by the virtual instructor. This study is designed to determine whether learners recognize the emotions we asked the instructor to display. Thus, this study is a step towards validation.

2.2.2. Video clip ratings

After each video clip, participants were asked to make six ratings concerning the emotion being displayed. All the ratings had a 5-point Likert scale. First, they were asked to "Please slide the bar to the number associated with the level at which you think the actor in the clip displayed these emotions:" with a sliding scale for each "Happy," "Content," "Frustrated," and "Bored." The numbers on the sliders were "1 – Not at All," "2" "3 – Average," "4" and "5 – Very." These ratings constitute the primary data used in this study.

After rating each emotion, they were asked two additional questions. First, "Please slide the bar to the number associated with the level at which you think the actor in the clip was active/passive" with the numbers on the sliding scale being "1 – Passive," "2" "3 – Neither Active nor Passive," "4" or "5 – Active." The second question was "Please slide the bar to the number associated with the level at which you think the actor in the clip was pleasant/unpleasant" with the numbers on the sliding scale being "1 – Unpleasant," "2" "3 – Neither Pleasant nor Unpleasant," "4" or "5 – Pleasant." We did not use these ratings in our analysis because our predictions concerned how well participants could recognize specific emotions.

2.2.3. Postquestionnaire

In the postquestionnaire, additional information was collected to understand each participants' ratings. First, to assess each participant's individual interest in the presented material, they were asked "How interesting was the presented material" and had to rate on a sliding scale consisting of "1 – Not at all interesting", "2" "3 – Somewhat interesting", "4" and "5 – Very interesting." The average response to this question was 3.03 (SD = 1.14). Next, to assess each participant's prior knowledge of the material covered in the lesson, participants were also asked, "How much knowledge did you have about binomial probability prior to this study" and had to rate on a sliding scale from "1 – None" to "2 – Minimal" to "3 – Moderate" to "4 – Extensive." The average response to this question was 2.10 (SD = 0.94). Then, participants were given the instruction, "If you have any comments about how this actor could have portrayed the emotions (happy, content, frustrated, and bored) better, please write them below." A textbox was provided for a typed response. Lastly, participants were asked to report demographic information, including age, gender, and ethnicity, by typing in their response to corresponding prompts.

2.3. Procedure

Participants were recruited through Amazon Mechanical Turk and were randomly assigned to the human instructor group or the virtual instructor group. Participants were required to be located in the United States. The participants went through the survey at their own pace. When they opened the survey, they first saw a consent page. Then, once they agreed to continue, they read about their task and what they needed to do to complete the survey. Then, each video clip was presented in random order for participants. After each video clip, participants rated the emotions of the instructor by responding to the six rating items in a fixed order (i.e., happy, content, frustrated, bored, activeness, pleasantness). Once the participant had seen and rated each of the eight video clips, the postquestionnaire was presented. When the participants finished the postquestionnaire, they were thanked and were given $3 as compensation for participating. We obtained IRB approval for the study and followed guidelines for research with human subjects.

3. Results

3.1. Are the two between-subjects groups equivalent on basic characteristics?

An initial step is to determine whether the two between-subjects groups (i.e., human instructor versus virtual instructor) were equivalent on basic characteristics. Concerning prior knowledge of binomial probability, which was assessed in the postquestionnaire with the question, "How much knowledge did you have about binomial probability prior to this study?", there was no statistically significant difference between the groups, $F(1, 200) = 0.18, p = .675$, partial $\eta^2 = 0.001$. There also was no statistically significant difference between the groups based on interest in the material, based on the postquestionnaire question, "How interesting was the presented material?", $F(1, 200) = 0.05, p = .944$, partial $\eta^2 < 0.001$. Concerning gender, there was no statistically significant difference between the groups, $\chi^2 (1, N = 202) = 0.13, p = .714$. Concerning being born in the United States, there was no statistically significant difference between the groups, $\chi^2 (1, N = 202) = 0.77, p = .380$. Concerning race and/or ethnicity, there was no statistically significant difference between the groups, $\chi^2 (6, N = 202) = 7.60, p = .269$. Lastly, concerning age, there was a statistically significant difference between the groups, $F(1, 200) = 6.19, p = .014$, partial $\eta^2 = 0.03$, with those who saw the human instructor ($M = 34.48, SD = 9.33$) being younger than those who saw the virtual instructor ($M = 38.20, SD = 11.72$). For all further analyses comparing the two between-subject groups (i.e., human versus virtual instructors), age was included as a covariate to control for the difference between the groups.

3.2. Can People Recognize the emotion being portrayed by the instructor?

A primary goal of this study was to determine whether people can

recognize the target emotion portrayed by the instructor. Hypothesis 1 is that participants will give higher ratings to the emotion displayed by the instructor than each of the other three emotions. To understand how well participants were able to recognize the emotion of the instructor, we ran multiple repeated-measures ANOVAs comparing the emotion mean rating (averaged across two clips) for each participant on happy, content, frustrated, and bored items, with follow-up pairwise tests to compare ratings on the target emotion against each of the others, and a follow-up 2 × 2 ANOVA with valence and arousal as factors.

3.2.1. Recognizing the emotion of the happy instructor

The first columns of Table 1 show the mean rating (and standard deviation) for each of the 4 emotions based on the happy videos. An ANOVA on the ratings for the happy videos produced a significant main effect, $F(3, 597) = 413.86$, $p < .001$, partial $\eta^2 = 0.68$. To conduct follow-up tests comparing each descriptor to happy, multiple pairwise t-tests were used. Due to the multiple comparisons, a Bonferroni correction was used setting alpha at $p = .017$. The t-test showed there was no difference in the happy rating and the content rating, $t(210) = 2.07$, $p = .039$. However, the happy rating was significantly higher than both the frustrated rating, $t(199) = 20.98$, $p < .001$, $d = 1.48$, and the bored rating, $t(199) = 21.22$, $p < .001$, $d = 1.88$.

The second and third columns of Table 1 show the mean rating (and standard deviation) for each emotion based on the happy videos split by the instructor type (i.e., virtual or human). An ANOVA based on only the data for the virtual instructor produced a significant main effect, $F(3, 306) = 200.64$, $p < .001$, partial $\eta^2 = 0.66$. T-tests showed that for the virtual instructor, there was no significant difference between ratings of happy and content, $t(102) = 0.33$, $p = .740$, but the happy rating was significantly higher than both the frustrated rating, $t(102) = 14.68$, $p < .001$, $d = 1.44$, the bored rating, $t(102) = 14.47$, $p < .001$, $d = 1.42$. An ANOVA based on only data for the human instructor also produced a significant effect, $F(3, 288) = 216.21$, $p < .001$, partial $\eta^2 = 0.69$. For the human instructor, the happy rating was significantly higher than the content rating, $t(98) = 2.68$, $p = .009$, $d = 0.27$, the frustrated rating, $t(96) = 15.07$, $p < .001$, $d = 1.54$, and the bored rating, $t(96) = 15.76$, $p < .001$, $d = 1.63$.

Partially consistent with hypothesis 1a, the combined data show and the virtual instructor data show that people rated the happy instructor higher on the happy scale than on the bored scale or the frustrated scale but no different on the content scale. We interpret this pattern to indicate that participants recognized that the happy instructor was displaying a positive emotion rather than a negative emotion. Entirely consistent with hypothesis 1, the human instructor data show that people rated the happy instructor higher on the happy scale than on any of the other scales.

3.2.2. Recognizing the emotion of the content instructor

The first columns of Table 2 show the mean rating (and standard deviation) for each emotion based on the content videos. An ANOVA on the ratings for the content videos produced a significant main effect, $F(3, 600) = 161.43$, $p < .001$, partial $\eta^2 = 0.45$. Follow-up t-tests (with alpha set at 0.017 as a Bonferroni correction) showed that there was a significantly higher rating for content than for and each of the other emotions ratings, including the happy rating, $t(200) = 6.01$, $p < .001$, d

Table 1
Means and SDs of ratings for the happy instructors.

	Combined		Virtual		Human	
	M	SD	M	SD	M	SD
Happy	**3.94**	**0.77**	**3.75**	**0.78**	**4.15**	**0.72**
Content	3.84	0.79	3.73	0.77	3.96*	0.79
Frustrated	1.68*	1.15	1.65*	1.13	1.72*	1.17
Bored	1.10*	1.10	1.69*	1.09	1.66*	1.07

Note. Asterisk (*) indicates significantly lower rating than for happy.

Table 2
Means and SDs of Ratings for the Content Instructors.

	Combined		Virtual		Human	
	M	SD	M	SD	M	SD
Happy	3.01*	0.73	3.02*	0.75	3.01*	0.72
Content	**3.31**	**0.79**	**3.35**	**0.71**	**3.27**	**0.87**
Frustrated	1.85*	1.12	1.79*	1.13	1.91*	1.10
Bored	2.18*	1.07	2.14*	1.08	2.22*	1.06

Note. Asterisk (*) indicates significantly lower rating than for content.

$= 0.42$, the frustrated rating, $t(200) = 16.23$, $p < .001$, $d = 1.14$, and the bored rating, $t(201) = 12.16$, $p < .001$, $d = 0.85$.

The second and third columns of Table 2 show the mean rating (and standard deviation) for each emotion based on the content videos split by type of instructor. An ANOVA based on only the data for the virtual instructor produced a significant main effect, $F(3, 306) = 96.26$, $p < .001$, partial $\eta^2 = 0.49$. Follow-up t-tests (with Bonferroni correction) showed that for the virtual instructor, the content rating was significantly greater than the happy rating, $t(102) = 4.62$, $p < .001$, $d = 0.45$, the frustrated rating, $t(102) = 12.90$, $p < .001$, $d = 1.27$, and the bored rating, $t(102) = 10.06$, $p < .001$, $d = 0.99$. An ANOVA based on only the data for the human instructor produced a significant main effect, $F(3, 291) = 66.42$, $p < .001$, partial $\eta^2 = 0.41$. Similar to other results, for the human instructor, the content rating was significantly higher than the happy rating, $t(97) = 3.83$, $p < .001$, $d = 0.38$, the frustrated rating, $t(97) = 10.15$, $p < .001$, $d = 1.02$, and the bored rating, $t(98) = 7.34$, $p < .001$, $d = 0.74$.

Overall, this pattern of results is consistent with hypothesis 1b, indicating that participants were able to distinguish the emotional tone of the content instructor from each of the other emotions. This was true for both the virtual instructors and the human instructors.

3.2.3. Recognizing the emotion of the frustrated instructor

The first columns of Table 3 show the mean rating (and standard deviation) for each emotion based on the frustrated videos. An ANOVA on the ratings for the frustrated videos produced a significant main effect, $F(3, 600) = 198.06$, $p < .001$, partial $\eta^2 = 0.21$. Follow up t-tests (with Bonferroni correction) showed the rating for frustrated was significantly higher than for each of the other emotions, including the happy rating, $t(200) = 15.01$, $p < .001$, $d = 1.12$, the content rating, $t(200) = 14.50$, $p < .001$, $d = 1.02$, and the bored rating, $t(201) = 4.10$, $p < .001$, $d = 0.29$.

The second and third columns of Table 3 show the mean rating (and standard deviation) for each emotion based on the frustrated videos split by type of instructor. An ANOVA based on only the data for the virtual instructor produced a significant main effect, $F(3, 306) = 84.23$, $p < .001$, partial $\eta^2 = 0.56$. Follow-up t-tests (with Bonferroni correction) showed that for the virtual instructor, ratings for frustrated were significantly higher than the happy ratings, $t(102) = 9.84$, $p < .001$, $d = 1.15$, and the content ratings, $t(102) = 8.78$, $p < .001$, $d = 0.86$, but not significantly different from the bored ratings, $t(102) = -0.24$, $p = .811$. An ANOVA based on only the data for the human instructor produced a significant main effect, $F(3, 291) = 125.42$, $p < .001$, partial $\eta^2 = 0.56$. For the human instructor, the frustrated rating was significantly higher

Table 3
Means and SDs of Ratings for the Frustrated Instructors.

	Combined		Virtual		Human	
	M	SD	M	SD	M	SD
Happy	1.84*	1.12	1.91*	1.05	1.77*	1.18
Content	1.95*	1.16	2.02*	1.08	1.87*	1.24
Frustrated	**3.78**	**1.02**	**3.43**	**1.00**	**4.14**	**0.91**
Bored	3.45*	0.91	3.46	0.89	3.45*	0.94

Note. Asterisk (*) indicates significantly lower rating than for frustrated.

than for each of the other emotion ratings, including the happy rating, t (97) = 13.18, $p <$ .001, d = 1.33, the content rating, t(97) = 12.15, $p <$ .001, d = 1.23, and the bored rating, t(98) = 6.12, $p <$ .001, d = 0.53.

Overall, these results support hypothesis 1c, indicating that participants could distinguish the emotion being displayed by the frustrated instructor from each of the other emotions. Specifically, participants were able to differentiate frustrated emotional tone from the rest of the emotions displayed. This pattern was found for the combined data and for human instructors, but for virtual instructors, level of arousal was more difficult to distinguish such that an active negative emotion (frustrated) was difficult to distinguish from a passive negative emotion (bored).

3.2.4. Recognizing the emotion of the bored instructor

The first columns of Table 4 show the mean rating (and standard deviation) for each emotion based on the bored videos. An ANOVA on the ratings for the bored videos produced a significant main effect, F(3, 600) = 351.62, $p <$ .001, partial η^2 = 0.64. Follow up t-tests (with alpha set at 0.017 to allow for a Bonferroni correction) showed that the bored rating was significantly higher than the happy rating, t(201) = 23.82, $p <$ .001, d = 1.67, the content rating, t(201) = 23.18, $p <$ .001, d = 1.63, and the frustrated rating, t(200) = 11.35, $p <$ .001, d = 0.84.

The second and third columns of Table 4 show the mean rating (and standard deviation) for each emotion based on the bored videos split by type of instructor. An ANOVA based on only the data for the virtual instructor produced a significant main effect, F(3, 306) = 179.16, $p <$ .001, partial η^2 = 0.64. Follow-up t-tests (with Bonferroni correction) showed that for the virtual instructor, the bored rating was significantly higher than happy rating, t(102) = 17.12, $p <$ .001, d = 1.69, content rating, t(102) = 16.44, $p <$ .001, d = 1.62, and frustrated rating, t(102) = 10.45, $p <$ .001, d = 1.02. An ANOVA based on only the data for the human instructor produced a significant main effect, F(3, 291) = 195.35, $p <$ .001, partial η^2 = 0.67. Similar to the foregoing results, for the human instructor, the bored rating was significantly higher than happy rating, t(98) = 16.51, $p <$ .001, d = 1.65, the content rating, t(98) = 16.35, $p <$ .001, d = 1.64, and the frustrated rating, t(97) = 5.84, $p <$ .001, d = 0.59.

This pattern is consistent with hypothesis 1d, indicating that participants were able to detect the emotional tone of the bored instructor. This was true for both the virtual instructor and the human instructor.

Overall, the results are largely consistent with hypothesis 1, especially when we focus on the combined data, which provides the most power, and the human instructor data. We conclude that participants were able to recognize the emotional tone displayed by the instructor reasonably well, both for a human instructor and a virtual instructor. However, in some cases involving virtual instructors, people could not distinguish between the active and passive versions of a positive emotion (i.e., seeing the happy instructor as both happy and content) or a negative emotion (i.e., seeing the frustrated instructor as both frustrated and bored). This may be attributed to the somewhat muted display of gesture by the virtual instructors.

3.3. Can people recognize the target emotion equally well with human and virtual instructors?

The second goal of this study was to determine whether participants can recognize the target emotion of the instructor equally well with human and virtual instructors. In short, we wanted to understand if participants were able to perceive the target emotion of virtual instructors as well as for human instructors (consistent with hypothesis 2). To investigate this, multiple ANCOVAs were run determining the effect of the instructor type on the ratings of the target emotion, with age as a covariate.

For the happy videos, there was a main effect of instructor type, F(1, 199) = 15.26, $p <$ .001, d = 0.53, with participants rating the human instructor (M = 4.15, SD = 0.72) as happier than the virtual instructor (M = 3.75, SD = 0.78), in contrast to hypothesis 2a. For the content videos, there was no main effect of instructor type, F(1, 199) = 1.34, p = .248, consistent with hypothesis 2b. For the frustrated videos, there was a main effect of instructor type, F(1, 199) = 24.78, $p <$ .001, d = 0.75, with participants rating the human instructor (M = 4.14, SD = 0.90) as more frustrated than the virtual instructor (M = 3.43, SD = 1.00), in contrast to hypothesis 2c. Lastly, in the bored videos, there was no main effect of instructor type, F(1, 199) = 0.28, p = .596, consistent with hypothesis 2d.

We conclude that participants perceived emotions with high activity levels (i.e., happy and frustrated) as stronger for human instructors than for animated pedagogical agents, perhaps because the on-screen agent did not make as heavy use of gesture and facial expression as did the human instructor. Overall, the results are consistent with hypothesis 2 for low-activity emotions (i.e., content and bored) but not high-activity emotions (i.e., happy and frustrated).

4. Discussion

4.1. Empirical contributions

This study found that learners generally were able to recognize the emotional tone of an instructor and were able to differentiate four individual emotions, especially more positive emotions from more negative emotions. Effect sizes were high (often above 1 standard deviation), which is important because it shows that emotions can be recognized easily. Furthermore, we also found that learners were able to recognize emotional tone similarly with human instructors and virtual instructors, although emotions that involved high activity (such as happy and frustrated) were more easily identified when presented by the human instructor.

This study shows that emotion is something that learners are aware of when it comes to how instructors present information. The emotions displayed by instructors are strongly recognized by learners in an online lesson. Additionally, the emotional tone of an instructor can be discerned by learners when the instructor is either a human or an on-screen agent, indicating that emotion portrayal is an important aspect to consider in creating instructional videos involving human or virtual instructors.

4.2. Theoretical contributions

This research helps support Russell's (2003) model of core affect using online lessons. Participants in this study were able to discriminate between emotions that came from each of the four quadrants (positive/active, positive/passive, negative/active, and negative/passive) created in Russell's (2003) model. This study supports not only that these emotions are distinct from one another, but also that people can decipher each of the individual emotions from either a human instructor or a virtual instructor, especially along the valence dimension (i.e., positive versus negative).

This research also supports the media equation hypothesis (Reeves &

Table 4
Means and SDs of Ratings for the Bored Instructors.

	Combined		Virtual		Human	
	M	SD	M	SD	M	SD
Happy	1.67*	1.04	1.70*	1.00	1.63*	1.08
Content	1.76*	1.03	1.85*	0.98	1.66*	1.08
Frustrated	3.20*	1.14	2.85*	1.16	3.58*	1.00
Bored	**4.36**	**0.84**	**4.35**	**0.86**	**4.37**	**0.83**

Note. Asterisk (*) indicates significantly lower rating than for bored.

Nass, 1996), which posits that people are able to accept media (such as computers and online lessons) as real people and places, as long as the media seems to act in a human way (i.e., have a human voice, use human-like gestures, etc.). This study supports this theory by demonstrating that in some situations people are able to recognize human emotions from a virtual instructor essentially as well as they are able to recognize emotions from a human instructor. Overall, research on the emotional tone of on-screen instructors adds a new approach to our understanding of the nature of emotion (Fox et al., 2018).

4.3. Practical implications

This study provides information that is relevant for designing online video lessons. First of all, this study demonstrates that the emotions of an instructor can be recognized by learners in an online lesson. With this knowledge in hand, there can be more research done aimed at understanding the effect of instructor emotion on learning processes and outcomes. Furthermore, this study shows that people are able to recognize the emotion of the instructor, regardless of if that instructor is human or virtual. These results suggest that an instructor of an online lesson can either be human or virtual, and learners will still be able to recognize the emotional tone displayed by the instructor. It also provides a foundation for future research investigating how the emotions displayed by different types of instructors (human vs. virtual) may affect learning differently.

4.4. Limitations and future directions

A main limitation of the study is that virtual instructors are not always made and implemented in the same way. This may affect the generalizability of the results this study. In this study, the virtual instructors were created based on videos of a human actor presenting the lesson in four different emotional tones. Yet, this is not the only way to create an online virtual instructor. Some virtual instructors may be created in different ways that change the voice, the gestures, the facial expressions, the body movements, and other characteristics. Due to this, the results of this study may not always hold true for all virtual instructors. Our virtual instructor may not have displayed facial expression and gesture quite as strongly as our human instructor, so different results might have been found if the expressions and gestures of the virtual instructor are strengthened. Future research should investigate how using different methods of designing virtual instructors may affect the way in which people are able to perceive the emotion of the instructor.

Another limitation is that although the human and virtual instructors had the same voice, they were not identical in all features, such as perspective, clothing, hair, and background color. This could have contributed to the significant differences in ratings that were found between the human and virtual instructors. Future research is needed to determine the influence of these kinds of factors.

Furthermore, a limitation of this study is the fact that we did not find many differences when investigating the active versus passive dimension of emotion. Participants were much better at distinguishing positive and negative emotions (e.g., happy versus frustrated) from one another, but not as good at distinguishing active from passive emotions (e.g., happy versus content). This could be due to people being less sensitive to this dimension; however, it could also be due to the way in which our instructor and agent portrayed the emotions. It is possible that participants had a hard time distinguishing between the active and passive emotions due to how the emotions were portrayed. This is especially true for the virtual instructor; it is possible that the translation from human to virtual may have lost the distinguishing components of the active versus passive dimension. Future research should investigate how using different instructors, human and animated, may impact how well participants are able to differentiate between the active and passive dimension in order to better understand if the difficulty to distinguish

between the two is due to perceptual processes or poor design.

Another limitation of this study is that the results may not hold true for all cultures. Different cultures use and display emotions differently than the way done in this study. This may lead to varied results for different cultural groups. Future research should investigate how culture may play a role in recognizing various emotional tones in online instructional material.

Additionally, future research should investigate if there is an effect of the instructor's emotion on student learning. This is important to understand because emotions are a key part of the human experience and thus the emotion of an instructor could play a key role in education. As suggested by Pekrun and Stephens (2010), emotions can be connected to both activities and outcomes involved with achievement, called achievement emotions. Students experience emotions while in school, so it is expected that emotions may play a role in learning. Understanding how the emotions portrayed by an instructor play a role in learning is essential to understand what may be promoting or hindering student success in an online classroom.

Lastly, research should be done to understand if the effect of instructors' emotions on learning is dependent on whether the instructor is human or virtual. This study shows that people are able to recognize the emotions similarly, whether the emotion is displayed by a human or a virtual instructor. Yet, the question remains if the instructor type moderates the effect of the instructor's emotion on learning.

CRediT authorship contribution statement

Declaration of competing interest

The authors report no conflicts of interest.

Acknowledgements

Appendix A

Video clips:

Human Happy: https://www.youtube.com/watch?v=1Py5qrT47Mo.

Human Content: https://www.youtube.com/watch?v=grGdY-7ZmGg.

Human Frustrated: https://www.youtube.com/watch?v=ztrC6g-wUPo.

Human Bored: https://www.youtube.com/watch?v=Eas28JcRSBg.

Agent Happy: https://www.youtube.com/watch?v=JXiPpsm7lPA.

Agent Content: https://www.youtube.com/watch?v=1gTXeu6UuP8.

Agent Frustrated: https://www.youtube.com/watch?v=xfQ1n5lVcBg.

Agent Bored: https://www.youtube.com/watch?v=NYQc9zsyVEk.

References

Brom, C., Starkova, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Psychology Review, 25*, 100–119.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Los Altos, CA: Malor Books.

Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon Press.

Fox, A. S., Lapate, R. C., Shackman, A. J., & Davidson, R. J. (Eds.). (2018). *The nature of emotion* (2nd ed.). New York: Oxford University Press.

Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (in press). Fostering positive emotion and history knowledge with location-based augmented reality and tour-guide prompts. Learning and Instruction.

Knörzer, L., Brünken, R., & Park, B. (2016). Emotions and multimedia learning: The moderating role of learner characterstics. *Journal of Computer Assisted Learning, 32* (6). https://doi.org/10.1111/jcal.12158.

Kramer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you–The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies, 71*, 335–349.

Loderer, K., Pekrun, R., & Lester, J. (in press). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. Learning and Instruction.

Loderer, K., Pekrun, R., & Plass, J. L. (2019). Emotional foundations of game-based learning. In J. Plass, R. E. Mayer, & B. Homer (Eds.), *Handbook of game-based learning* (pp. 111–152). Cambridge, MA: MIT Press.

(in press-a) Mayer, R. E. (2020a). *Multimedia learning* (3rd ed.). New York: Cambridge University Press.

(in press-b) Mayer, R. E. (2020b). *Searching for the role of emotions in e-learning*. Learning and Instruction.

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–370). New York: Cambridge University Press.

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12–18.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309–326.

Nass, C., & Brave, S. (2005). *Wired for speech*. Cambridge, MA: MIT Press.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 27–42). New York: Cambridge University Press.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419.

Park, B., Knorzer, L., Plass, J. L., & Brunken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study ion the use of anthropomorphisms. *Computers & Education, 86*, 30–42.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R. (2016). Academic emotions. In K. R. Wentzel, & D. B. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 120–144). New York, N.Y.; London: Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). New York, N.Y.; London: Taylor and Francis.

Pekrun, R., & Stephens, S. J. (2010). Achievement emotions in higher education. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 257–306). New York: Springer. https://doi.org/10.1007/978-90-481-8598-6_7.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29*, 128–140.

Plass, J. L., Homer, B. D., MacNamara, A., Ober, T., Rose, M., Pawar, S., Hovey, C. M., & Olsen, A. (in press). Emotional design for digital games for learning: The affective quality of expression, color, shape, and dimensionality. Learning and Instruction.

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*, 339–359.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Y. Tettegah, & M. P. McCreery (Eds.), *Emotions, technology, and learning* (pp. 131–161). London: Academic Press.

Reeves, B., & Nass, C. (1996). *The media equation*. New York: Cambridge University Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.

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.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Spinger.

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.

Uzum, A. M., & Yildirim, Z. (2018). Exploring the effect of using different levels of emotional design features in multimedia science learning. *Computers & Education, 119*, 112–128.

Wong, R.M., Adesope, O.O. (in press). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. Educational Psychology Review.