

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/348646724>

Do Learners Recognize and Relate to the Emotions Displayed By Virtual Instructors?

Article in *International Journal of Artificial Intelligence in Education* · January 2021

DOI: 10.11007/s40593-021-00238-2

CITATIONS

0

READS

131

6 authors, including:



Alyssa Lawson

University of California, Santa Barbara

5 PUBLICATIONS 4 CITATIONS

SEE PROFILE



Richard Mayer

University of California, Santa Barbara

500 PUBLICATIONS 67,959 CITATIONS

SEE PROFILE



Nicoletta Adamo-Villani

Purdue University

95 PUBLICATIONS 645 CITATIONS

SEE PROFILE



Bedrich Benes

Purdue University

208 PUBLICATIONS 3,495 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Isotropic Surface Remeshing [View project](#)



Learning Glass: A New Platform for Promoting STEM Engagement and Learning [View project](#)



Do Learners Recognize and Relate to the Emotions Displayed By Virtual Instructors?

Alyssa P. Lawson¹  · Richard E. Mayer¹ · Nicoletta Adamo-Villani² · Bedrich Benes² · Xingyu Lei² · Justin Cheng²

Received: 13 June 2020 / Revised: 1 January 2021 / Accepted: 6 January 2021

Published online: 20 January 2021

© International Artificial Intelligence in Education Society 2021

Abstract

There has been much research on the effectiveness of animated pedagogical agents in an educational context, however there is little research about how the emotions they display contribute to a learner's understanding of the lesson. The positivity principle suggests that learners should learn better from instructors with positive emotions compared to those with negative emotions. Additionally, the media equation theory (Reeves and Nass 1996) would suggest this principle should be true for animated instructors as well. In an experiment, students viewed a lesson on binomial probability taught by an animated instructor who was happy (positive/active), content (positive/passive), frustrated (negative/active), or bored (negative/passive). Learners were able to recognize positive from negative emotions, rated the positive instructors as better at facilitating learning, more credible, more human-like, and more engaging. Additionally, learners who saw positive instructors indicated they tried to pay attention to the lesson and enjoyed the lesson more than those who saw negative instructors. However, learners who saw positive instructors did not perform better on a delayed test than those who saw negative instructors. This suggests that learners recognize and react to the emotions of the virtual instructors, but research is needed to determine how the emotions displayed by virtual instructors can promote better learning outcomes.

Keywords Animated pedagogical agents · Affective processes · Emotion · Emotional tone · Learning

✉ Alyssa P. Lawson
Alyssa.lawson@psych.ucsb.edu

¹ Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

² Computer Science Department, Purdue University, Lafayette, IN, USA

Introduction

Emotion in Animated Pedagogical Agents

Animated pedagogical agents (or virtual instructors) are lifelike onscreen characters intended to provide guidance or instruction in learning episodes. Over the past 20 years, researchers have developed numerous onscreen agents (Cassell et al. 2000; Johnson and Lester 2016; Johnson et al. 2000) and examined features that improve learning outcomes (e.g., Wang et al. 2008). Our focus in the present study is to examine how affective and social cues from a virtual instructor play a role in the learning process.

In particular, we examine whether the positivity principle applies to virtual instructors. The positivity principle posits that people recognize when instructors display positive emotions during instruction, report better rapport with positive instructors, report better learning activity with positive instructors, and attain better learning outcomes with positive instructors. The positivity principle is in line with research on emotional design, showing that increasing the positive emotional tone of onscreen characters can improve learning in a computer-based lesson or game (Mayer and Estrella 2014; Plass et al. 2020; Plass and Kaplan 2015; Plass et al. 2014; Um et al. 2012).

For example, consider an online lesson in which an animated character presents a video lecture, such as exemplified in Fig. 1. The lesson involves spoken words from the instructor and printed words and graphics in the slides that she is standing next to as she lectures. Based on Russell's (1980, 2003) model of core affect, the instructor displays one of four emotional stances through her voice and gestures: happy, content, frustrated, or bored. According to the model, emotions can vary along two orthogonal dimensions, which we refer to as valence (running from negative to positive) and activity (running from passive to active). Pekrun and colleagues (Pekrun and Linnenbrink-Garcia 2012; Pekrun and Perry 2014; Loderer et al. 2021) have provided some support for the psychological mechanisms underlying these dimensions within a theory of achievement motivation, particularly the valence dimension. Happy and content are positive emotions whereas frustrated and bored are negative emotions. More specifically, happy is positive and active, content is positive and passive,

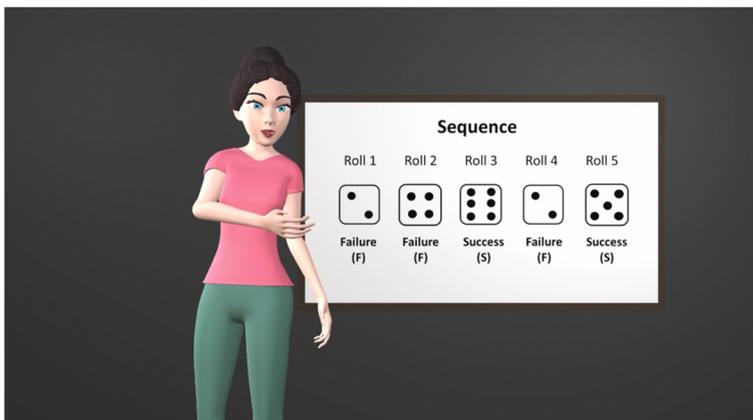


Fig. 1 Animated instructor teaching lesson on binomial probability

frustrated is negative and active, and bored is negative and passive. According to the media equation theory (Reeves and Nass 1996) people will accept an onscreen computer-generated character as a social partner in an equivalent way as a human as a social partner, and thereby we expect learners to be influenced by the emotion displayed by the virtual instructor. In short, we expect learners to be sensitive to affective cues displayed by virtual instructors.

Throughout 20 years of the development of the cognitive theory of multimedia learning (Mayer 2014a, 2020a), an overarching goal has been to discover evidence-based principles for the design of multimedia instructional messages. A multimedia instructional message is a communication involving words and pictures that is intended to promote learning. Although research on multimedia instructional messages commonly involves media such as printed text and static graphics or narrated animations (Mayer 2014b), in the present study we focus on the increasingly important medium of instructional video (Derry et al. 2014; Mayer et al. 2020). At the college level, instructional videos play a central role in online courses including in MOOCs, as resources in course management systems such as for flipped classrooms, and as alternatives for face-to-face instruction required by situations such as the recent pandemic.

According to Mayer's (2014a, 2020a) cognitive theory of multimedia learning, a multimedia instructional message is a communication consisting of words and pictures that is intended to foster learning. In the original formulation, the focus is on two cognitive aspects of the multimedia instructional message: the instructional content (i.e., what material is presented) and the instructional method (i.e., how it is presented). However, more recently, research has added a focus on affective and social cues in multimedia instructional messages, as reflected in the cognitive affective model of e-learning (Mayer 2014b, 2020a, 2021).

Cognitive Affective Model of e-Learning

Although cognitive factors have been the primary focus of research on technology-based instruction, there is growing interest in incorporating affective and social factors including the Cognitive Affective Theory of Learning with Media (Moreno and Mayer 2007), Social Agency Theory of Multimedia Learning (Mayer 2014b, 2020a), Control Value Theory of Achievement Motivation (Pekrun and Linnenbrink-Garcia 2012; Pekrun and Perry 2014), and Integrated Cognitive Affective Model of Learning with Multimedia (Plass and Kaplan 2016).

For purposes of the present study, we focus on the newly proposed Cognitive Affective Model of e-Learning that is designed specifically for learning from video lectures with onscreen instructors (Lawson et al. 2021; Mayer 2020b). As shown in Fig. 2, the cognitive affective model of e-learning involves a sequence of five events. In the first event, the instructor displays a positive emotional stance during learning, such as displaying a happy or content emotion. This leads to the learner recognizing the emotional stance of the instructor (event 2). From here, when the instructor does display a positive emotion, the learner develops a social connection with the instructor (event 3). When the learner begins to feel this social connection, the learner experiences more enjoyment and exerts more effort to learn from the instructor (event 4). Lastly, this causes the learner to perform well on tests of learning (event 5).

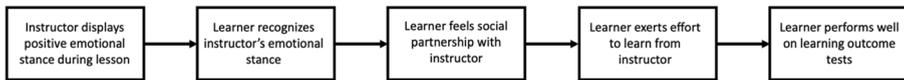


Fig. 2 Cognitive affective model of e-Learning

Predictions

Our predictions are in line with the steps of the cognitive affective model of e-learning. As this framework sheds light only on the valence dimension (positive and negative), our predictions focus on this dimension. Analyses of the activity dimension are exploratory. In the first step, we predict that positive instructors will be rated higher on positive emotions and negative instructors will be rated higher on negative emotions (hypothesis 1). This hypothesis breaks down into four separate hypotheses; one for each emotion. Happy instructors will be seen as more positive than negative (hypothesis 1a), content instructors will be seen as more positive than negative (hypothesis 1b), bored instructors will be seen as more negative than positive (hypothesis 1c), and frustrated instructors will be seen as more negative than positive (hypothesis 1d).

Matching the second step of the model, we predict that participants will rate the positive instructors higher in the four categories of the Agent Persona Index (API; Baylor and Ryu 2003; hypothesis 2)—the instructor facilitates learning, the instructor is credible, the instructor is humanlike, and the instructor is engaging. Along with step three of the model, we predict that participants will have higher ratings of effort, motivation, and enjoyment in the postquestionnaire when they see a positive instructor compared to when they see a negative instructor (hypothesis 3). Lastly, we predict that positive instructors should lead to higher posttest scores than negative instructors (hypothesis 4).

Method

Participants and Design

The participants were 119 participants recruited from a university in southern California from a psychology subject pool. Their mean age was 19.01 ($SD = 1.24$); 87 were women and 32 were men. The experiment used a 2 (valence of emotion: positive vs. negative) \times 2 (activity of emotion: active vs. passive) between-subjects design. The four groups of participants are as follows: 30 in the active/positive condition (also called the happy instructor condition), 30 in the passive/positive condition (also called the content instructor condition), 30 in the passive/negative condition (also called the bored instructor condition), and 29 in the active/negative condition (also called the frustrated instructor condition). Based on a power analysis, this sample size was determined to be sufficient to find a medium effect size (of $d = 0.50$) when power is 0.80.

Materials

The paper-based materials consisted of a prequestionnaire and a postquestionnaire. The computer-based materials consisted of 4 versions of a video on binomial probability taught by an animated agent and a posttest consisting of 21 questions in a self-paced PowerPoint presentation.

Prequestionnaire

The prequestionnaire collected demographic information from the participant, including major, grade point average (GPA), age, gender, and year in school. It also had participants rate their prior knowledge of statistics on a five-point scale from “Very Low” to “Very High.” Additionally, 11 statements about knowledge about binomial probability and statistics were listed and participants were asked to mark each statement that applied to them, in order to obtain an objective measure of prior knowledge (e.g., “I have taken a statistics class” and “I know how to compute joint probability.”) The total number of marks (ranging from 0 to 11) instituted each participant’s prior knowledge score. The Cronbach’s alpha for prior knowledge was 0.56. The low Cronbach’s alpha was due to the fact that the checklist provided to students was meant to assess for participants’ background knowledge of the topic broadly, rather than assessing their knowledge of binomial probability more specifically. The prequestionnaire was used instead of a pretest because of the potential for a testing effect and a priming effect (Mayer 2020a). According to the testing effect, a pretest is a form of instruction that can cause learning before the lesson is presented. According to the priming effect, a pretest can prime students to pay attention to certain information during the lesson that they would not necessarily pay attention to in the first place. Thus, instead of introducing this bias, the prequestionnaire was used to assess the level at which students had related knowledge of the content of the lesson, consistent with prior work on multimedia learning (Mayer 2020a).

Video Lessons

The video lessons consisted of four versions of a binomial probability lesson. The instructor was an animated young woman, whose behavior was based on the video of a young woman actor from a theatre program giving the same lesson in the four different emotional stances in four separate videos. The animated woman was standing in front of a screen with instructional material displayed as she talked. The voice of the woman was taken from the live action version of the lessons and matched to the appropriate emotion video for the animated instructor. Her gestures, facial expressions, and body positioning were created to mirror as closely as possible each of the four human videos, respectively. For example, for positive emotions, the agent used an open body position and for the negative emotions, the agent used a closed body position. For active emotions, the agent was positioned to look like she was leaning forward while for the passive emotions, the agent was leaning back. Facial expressions and gestures were adjusted to be appropriate for each emotion, corresponding to the respective live action video displaying each emotion. The lesson contained 18 slides and 1510 spoken words. The videos ranged in length from 8 minutes to 35 seconds to 12 minutes and 57 seconds, depending on the emotion being portrayed. A screenshot is provided in Fig. 1. The script is provided in Appendix and is a modified version of a paper-based lesson created by Mayer and Greeno (1972).

In each of the video lessons, the animated instructor’s voice, gestures, body positioning, and facial expression mirrored how the original actress portrayed each of the 4 emotions: happy, content, frustrated, and bored. In previous work (Lawson et al. 2021), the four videos of the animated instructor were pretested using participants from

Amazon Mechanical Turk. In this validation study, participants were shown clips of each of the videos in a random order and were asked to rate how happy, content, frustrated, and bored the instructor seemed. Overall, results showed that the four emotions were generally interpreted correctly, and participants were especially successful in distinguishing positive and negative emotions.

Posttest

The posttest was 21 questions, each presented in a fixed order on separate slides of a PowerPoint Lesson. The questions had participants recall the definition of the different symbols used in the equations, solve problems using formula, answer questions about the binomial probability, and identify unanswerable questions. Participants were given up to 55 minutes to answer these questions and participants could move forward their own pace. Participants earned 1 point for each correct answer they reported. For two of the questions, there were two parts to the answer, so they received 0.5 points for each of the parts answered correctly. From their total score on the posttest, a percent correct was calculated by dividing their total number correct divided by 21. This score was used for the analysis. Cronbach's alpha for the posttest is 0.76. The low Cronbach's alpha can be explained by the posttest assessing learning in a variety of ways and at a variety of levels of transfer, including rote memorization of definitions, filling in equations properly, answering question that required essay answers, solving novel problems, and recognizing impossible problems. This diversity of items—which provides a broad assessment—is more likely to lead to a lower alpha than using questions that were all similar in their level of transfer and mode of responding.

Postquestionnaire

The postquestionnaire included different sections of questions. The first section asked participants to rate the degree to which the instructor in the lesson displayed each emotion (happy, content, bored, and frustrated) on a 5-point scale from “strongly disagree” to “strongly agree.” This section also had participants rate the degree to which the instructor was active and pleasant, both on the same 5-point scale as the previous questions. The next section asked participants to answer 5 questions about their experience with the lesson, including their motivation, the difficulty of the lesson, the effort to understand, the enjoyment of the lesson, and the desire to learn from other similar lessons. All of these questions were rated on a 5-point scale. The next section had questions from the Agent Persona Index (API; Baylor and Ryu 2003). Four subscales were used to assess how the participants rated the instructor in facilitating learning (Cronbach's alpha = 0.84), credibility (Cronbach's alpha = 0.42), being human-like (Cronbach's alpha = 0.79) and engaging (Cronbach's alpha = 0.88).

Apparatus

The apparatus consisted of 4 dell computers with over-the-ear headphones. Each participant was in a separate cubicle with an individual computer that blocked visual contact among participants.

Procedure

Participants were randomly assigned to one of the four conditions and up to four participants were tested independently from one another in each session. First, the researcher explained the study to the participants and had each participant read and sign an informed consent form. Then, the participants were given time to complete the prequestionnaire at their own rate. Once done with that, participants were instructed on how to watch the video and then they watched the entire video. Participants were then thanked and asked to return exactly a week later to complete the second part of the experiment. After a week, participants came back to the lab and first were given instructions on how to complete the posttest. They were then allowed to work through the posttest, one question at a time, at their own pace. They were given a simple calculator to complete calculations with. They could work through each problem on a pre-numbered sheet of paper. Participants took on average 28 minutes and 22 seconds ($SD = 7$ minutes and 36 seconds) to finish the posttest, with the fastest time being 15 minutes and 15 seconds and the slowest time being 50 minutes and 45 seconds. We used a delayed posttest because the goal of education is to promote learning that lasts beyond a few minutes and because deep learning sometimes shows up better on delayed tests (Mayer 2011). Once they completed the posttest, they were given the postquestionnaire packet to complete. Once done with that, participants were thanked and excused from the study. The entire experiment took no more than an hour and a half to complete in total. We obtained IRB approval and adhered to guidelines for ethical treatment of human subjects.

Results

Do the Groups Differ on Basic Characteristics?

A preliminary issue for analysis is whether the random assignment created groups equivalent in basic characteristics. Concerning age, there were no statistically significant differences between the groups based on valence, $F(1, 115) = 1.96, p = .164$, nor based on activity, $F(1, 115) = 0.04, p = .837$, and no significant interaction, $F(1, 115) = 0.05, p = .817$. Concerning prior knowledge, there were no statistically significant differences between the groups based on valence, $F(1, 115) = 0.01, p = .918$, nor based on activity, $F(1, 115) = 0.001, p = .975$, and no interaction, $F(1, 115) = 1.07, p = .304$. Additionally, concerning gender, there were no significant differences between the 4 groups, $\chi^2(3, N = 119) = 3.80, p = .284$. We conclude that the groups were similar in basic characteristics.

Do the Learners Recognize the Emotion of the Instructor?

The first step of the cognitive affective model of e-learning explains that learners recognize the emotion of the instructor (first step in Fig. 2). Table 1 shows the means and standard deviations of the emotion ratings for each of the groups. To analyze this, 2 (valence: positive versus negative) \times 2 (activity: active versus passive) ANOVAs were conducted to determine whether participants were able to recognize the emotion of the

Table 1 Mean rating and standard deviation of the instructors' emotions

Emotion of instructor	Mean rating (and SD)							
	Happy rating (Positive/Active)		Content rating (Positive/Passive)		Bored rating (Negative/Passive)		Frustrated rating (Negative/Active)	
	M	SD	M	SD	M	SD	M	SD
Happy instructor	3.83	.87	3.87	0.82	2.00*	1.02	1.53*	1.27
Content instructor	3.90	0.80	4.17	.75	2.27*	1.29	1.63*	1.18
Bored instructor	1.60*	0.89	1.79*	1.05	4.60	.86	3.67*	1.22
Frustrated instructor	3.03*	1.27	3.21*	1.18	3.38*	1.29	2.28	1.29

Note. Bolded font indicates the rating of the target group. Asterisk (*) indicates score is significantly different (at $p < .05$) than the target group (indicated in bold font) in the same column

instructor. The first column of Table 1 displays the means and standard deviations for the happy ratings by the four groups. For the happy ratings, there was a statistically significant effect of valence, $F(1, 115) = 75.31, p < .001, d = 1.67$, such that participants who received positive instructors gave higher happy ratings ($M = 3.87, SD = 0.83$) than participants who received negative instructors ($M = 2.31, SD = 1.03$), consistent with hypothesis 1a. Additionally, there was a statistically significant effect of activity, $F(1, 115) = 14.67, p < .001, d = 0.53$, such that participants who received active instructors gave higher happy ratings ($M = 3.44, SD = 1.15$) than participants who received passive instructors ($M = 2.75, SD = 1.43$). There was also a significant interaction, $F(1, 115) = 17.67, p < .001$. To follow-up, a one-way ANOVA was conducted. The one-way ANOVA was significant, $F(3, 115) = 36.20, p < .001$. Dunnett's test (with $p < .05$) was conducted to analyze the differences among the four groups. The mean happy rating for the happy instructor group was not significantly different from the happy rating of the other positive group (i.e., the content instructor group, $p = .987$), but was significantly higher than the happy rating of the two negative groups (i.e., the bored instructor group, $p < .001, d = 2.53$, and the frustrated instructor group, $p = .006, d = 0.74$). Consistent with the positivity principle and supporting hypothesis 1a, participants who learned with instructors displaying positive emotions (i.e., happy or content) gave a higher happy rating than participants who learned with instructors displaying negative emotions (i.e., frustrated or bored).

Column 2 in Table 1 displays the means and standard deviations for the content ratings by the four groups. For the content ratings, there was a statistically significant effect of valence, $F(1, 114) = 73.55, p < .001, d = 1.40$, such that participants who received positive instructors gave higher content ratings ($M = 4.02, SD = 0.79$) than those who received negative instructors ($M = 2.50, SD = 1.32$), consistent with hypothesis 1b. Additionally, there was a statistically significant effect of activity, $F(1, 114) = 9.92, p = .002, d = 0.42$, in which participants who received active instructors gave higher content ratings ($M = 3.54, SD = 1.06$) compared to those who received passive instructors ($M = 3.00, SD = 1.50$). Lastly, there was a significant interaction, $F(1, 114) = 23.48, p < .001$. To follow-up, a one-way ANOVA was conducted. The one-

way ANOVA was significant, $F(3, 114) = 35.48, p < .001$. Dunnett's test (with $p < .05$) was conducted to analyze the differences among the four groups. The mean content rating for the content instructor group was not significantly different from the content rating of the other positive group (i.e., the happy instructor group, $p = .248$), but was significantly higher than the content rating of the two negative groups (i.e., the bored instructor group, $p < .001, d = 2.62$ and the frustrated instructor group, $p = .001, d = 0.98$). Consistent with the positivity principle and supporting hypothesis 1b, participants who learned with instructors displaying positive emotions gave a higher content rating than participants who learned with instructors displaying negative emotions. However, there was confusion among the active/passive dimension in that participants who learned with active instructors (i.e., happy or frustrated) gave higher content rating than participants who learned with passive instructors (i.e., content or bored).

Column 3 in Table 1 shows the means and standard deviations for the bored ratings by the four groups. For the bored ratings, there was a statistically significant effect of valence, $F(1, 115) = 90.74, p < .001, d = 0.49$, such that participants who received negative instructors gave higher bored ratings ($M = 4.00, SD = 1.25$) than those who received positive instructors ($M = 3.38, SD = 1.29$). Additionally, there was a statistically significant effect of activity, $F(1, 115) = 14.56, p < .001, d = 0.52$, such that participants who received passive instructors gave higher bored ratings ($M = 3.43, SD = 1.51$) than those who received active instructors ($M = 2.68, SD = 1.35$). Lastly, there was a significant interaction, $F(1, 115) = 5.99, p = .016$. To follow-up, a one-way ANOVA was conducted. The one-way ANOVA was significant, $F(3, 115) = 37.39, p < .001$. Dunnett's test (with $p < .05$) was conducted to analyze the differences among the four groups. The mean bored rating for the bored instructor group was significantly higher than the bored rating of the other negative group (i.e., the frustrated instructor group, $p < .001, d = 1.12$), and significantly higher than the bored rating of the two positive groups (i.e., the happy instructor group, $p < .001, d = 2.76$, and the content instructor group, $p < .001, d = 2.43$). Consistent with the positivity principle and supporting hypothesis 1c, participants who learned with instructors displaying negative emotions gave a higher bored rating than participants who learned with instructors displaying positive emotions.

Column 4 of Table 1 shows the means and standard deviations for the frustrated ratings by the four groups. For the frustrated ratings, there was a statistically significant effect of valence, $F(1, 115) = 58.37, p < .001, d = 1.26$, such that participants who received negative instructors gave higher frustrated ratings ($M = 2.98, SD = 1.35$) than those who received positive instructors ($M = 1.58, SD = 0.79$). Additionally, there was a statistically significant effect of activity, $F(1, 115) = 17.84, p < .001, d = 0.60$, such that the participants who received passive instructors gave higher frustrated ratings ($M = 2.65, SD = 1.39$) than those who received active instructors ($M = 1.90, SD = 1.09$). Lastly, there was a significant interaction, $F(1, 115) = 12.62, p = .001$. To follow-up, a one-way ANOVA was conducted. The one-way ANOVA was significant, $F(3, 115) = 29.53, p < .001$. Dunnett's test (with $p < .05$) was conducted to analyze the differences among the four groups. The mean frustrated rating for the frustrated instructor group was significantly higher than the frustrated rating for the positive groups (i.e., the happy instructor group, $p = .013, d = 0.72$, and the content instructor group, $p = .038, d = 0.64$). However, the mean frustrated rating for the frustrated instructor group was significantly lower than the

frustrated rating for the other negative group (i.e., the bored instructor group, $p < .001$, $d = 1.20$). Consistent with the positivity principle and hypothesis 1d, participants who learned with instructors displaying negative emotions gave a higher frustrated rating than participants who learned with instructors displaying positive emotions. However, there was confusion among the active/passive dimension once again in that participants who learned with passive instructors gave higher frustrated rating than participants who learned with active instructors.

Overall, there is evidence supporting the positivity principle in that participants were able to distinguish positive emotions (happy and content) from negative emotions (bored and frustrated). Even so, learners did struggle when it came to identifying the activity level of the emotion, specifically for the ratings of content and frustrated. In this section we conducted a 2×2 ANOVA on each emotion rating, followed up by a one-way ANOVA on each emotion rating with Dunnett's test in order to directly test our predictions. In the 2×2 ANOVAs, main effects and interactions inform the cognitive affective theory of e-learning, although we acknowledge that interactions serve to qualify any main effects. This is why we included subsequent one-way ANOVAs with a Dunnett's test, which allow us to test specific a priori predictions. In the Dunnett's test we compared each group to the target group for each emotion rating (i.e., the target group had an agent who displayed the specific emotion that was being rated by the participants).

Do Learners Develop a Stronger Social Partnership with Positive Instructors?

The next step in the cognitive affective model of e-learning proposes that learners feel a social partnership with the instructor, which we predict will be stronger when the instructor is positive (i.e., step 2 in Fig. 2). To test this, we conducted ANOVAs on the four subcomponents of the API (Baylor and Ryu 2003). Means and standard deviations are reported in Table 2. The first subcomponent assessed how well the instructor facilitated learning. Column 1 of Table 2 displays the means and standard deviations for this subcomponent. There was a statistically significant effect of valence, $F(1, 113) = 41.36$, $p < .001$, $d = 1.14$, with participants who learned with positive instructors ($M = 3.07$, $SD = 1.00$) rating their instructor higher at facilitating learning than participants who learned with negative instructors ($M = 2.02$, $SD = 0.87$). There was also a statistically significant effect of activity, $F(1, 113) = 7.88$, $p = .006$, $d = 0.43$, with participants who learned with active instructors ($M = 2.77$, $SD = 0.84$) rating their instructor as better at facilitating learning than participants who learned with passive instructors ($M = 2.32$, $SD = 1.22$). Lastly, there was a significant interaction, $F(1, 113) = 9.31$, $p = .003$. To understand the interaction, t -tests were run using Bonferroni corrections ($\alpha = 0.025$), separating the analyses based on valence. For the positive emotions, there was no significant difference in ratings between the happy instructor group ($M = 3.05$, $SD = 0.72$) and the content instructor group ($M = 3.09$, $SD = 1.30$), $t(56) = -0.15$, $p = .881$. For the negative emotions, the frustrated instructor group ($M = 2.50$, $SD = 0.86$) rated their instructor as better at facilitating learning than the bored instructor group ($M = 1.55$, $p = .58$), $t(48.69) = 4.97$, $p < .001$, $d = 1.30$. Consistent with the positivity principle, instructors displaying positive emotions were seen as better at facilitating learning than instructors displaying negative emotions.

The second subcomponent assessed how credible the instructor was. Column 2 of Table 2 displays the means and standard deviations for this subcomponent. There was a statistically significant effect of valence, $F(1, 114) = 11.71, p = .001, d = 0.63$, with participants who learned from positive instructors ($M = 4.00, SD = 1.92$) rating their instructor as more credible than participants who learned from negative instructors ($M = 3.06, SD = 0.89$). However, there was no statistically significant effect of activity, $F(1, 114) = 0.50, p = .480$. There was a significant interaction, $F(1, 114) = 5.77, p = .018$. To understand the interaction, t -tests were run using Bonferroni corrections ($\alpha = 0.025$), separating the analyses based on valence. For the positive emotions, there was no significant difference in ratings between the happy instructor group ($M = 3.77, SD = 0.75$) and the content instructor group ($M = 4.23, SD = 2.59$), $t(57) = -0.92, p = .364$. For negative emotions, the frustrated instructor group ($M = 3.49, SD = 0.93$) rated their instructor as more credible than the bored instructor group ($M = 2.65, SD = 0.64$), $t(57) = 4.07, p < .001, d = 1.06$. Consistent with the positivity principle, instructors displaying positive emotions were seen as more credible than instructors displaying negative emotions.

The third subcomponent assessed how human-like the animated instructor was. Column 3 of Table 2 displays the means and standard deviations for this subcomponent. There was a statistically significant effect of valence, $F(1, 114) = 10.17, p = .002, d = 0.56$, with participants who received positive instructors ($M = 2.82, SD = 0.88$) rating their instructor as more human-like than participants who received negative instructors ($M = 2.34, SD = 0.84$). There was no statistically significant effect of activity, $F(1, 114) = 0.78, p = .379$. There was, however, a significant interaction, $F(1, 114) = 16.56, p < .001, d = 0.15$. To understand the interaction, t -tests were run using Bonferroni corrections ($\alpha = 0.025$), separating the analyses based on valence. For positive emotions, the content instructor group ($M = 3.05, SD = 0.92$) rated their instructor as similarly human-like to the happy instructor group ($M = 2.58, SD = 0.78$), $t(57) = -2.13, p = .037, d = 0.55$. However, for negative emotions, the frustrated instructor group ($M = 2.71, SD = 0.76$) rated their instructor as more human-like than the bored instructor group ($M = 1.97, SD = 0.76$), $t(57) = 3.73, p < .001, d = 0.97$. Consistent with the positivity principle, the instructors displaying positive emotion were seen as more human-like than the instructors displaying negative emotion.

The fourth and final subcomponent assessed how engaging the instructor was. Column 4 of Table 2 displays the means and standard deviations for this

Table 2 Mean rating and standard deviation of the subsections of the API

Emotion of instructor	Mean rating (and SD)							
	Facilitate Learning		Credible		Human-like		Engaging	
	M	SD	M	SD	M	SD	M	SD
Happy instructor	3.05	0.72	3.77	0.75	2.58	0.78	3.07	0.67
Content instructor	3.09	1.22	4.23	2.59	3.05	0.92	3.09	0.86
Bored instructor	1.55	0.58	2.65	0.64	1.97	0.76	1.41	0.49
Frustrated instructor	2.50	0.86	3.49	0.93	2.71	0.76	2.37	0.91

subcomponent. There was a statistically significant effect of valence, $F(1, 114) = 73.95$, $p < .001$, $d = 1.46$, with participants who learned with positive instructors ($M = 3.08$, $SD = 0.77$) rating their instructor as more engaging than participants who learned with negative instructors ($M = 1.88$, $SD = 0.87$). There was a statistically significant effect of activity, $F(1, 114) = 11.57$, $p = .001$, $d = 0.47$, with participants who learned with active instructors ($M = 2.71$, $SD = 0.87$) rating their instructor as more engaging than participants who learned with passive instructors ($M = 2.25$, $SD = 1.09$). Lastly, there was a significant interaction, $F(1, 114) = 12.33$, $p = .001$. To understand the interaction, t -tests were run using Bonferroni corrections ($\alpha = 0.025$), separating the analyses based on valence. For positive emotions, there was no significant difference between the happy instructor group ($M = 3.07$, $SD = 0.67$) and content instructor group ($M = 3.09$, $SD = 0.86$), $t(56) = -0.08$, $p = .941$. For the negative emotions, the frustrated instructor group ($M = 2.37$, $SD = 0.91$) rated their instructor as more engaging than the bored instructor group ($M = 1.41$, $SD = 0.49$), $t(42.66) = 5.03$, $p < .001$, $d = 1.32$. Consistent with the positivity principle, instructors displaying positive emotions were seen as more engaging than instructors displaying negative emotions.

Overall, hypothesis 2 and the positivity principle were supported. Positive instructors were rated as better at facilitating learning, more credible, more human-like, and more engaging.

Do Learners Report More Effort, Motivation, and Enjoyment for Positive Instructors?

The next step in the cognitive affective model of e-learning is that learners exert more effort to learn from the instructor, which we predict will be more likely for learners with positive instructors. Means and standard deviations are displayed in Table 3. To assess learners' effort into the lesson, multiple questions from the posttest were analyzed using ANOVAs. First, participants were asked to rate their agreement to the statement, "I was motivated to pay attention to the lesson I just watched." Column 1 of Table 3 displays the means and standard deviations for this question. There was a statistically significant effect of valence, $F(1, 115) = 26.15$, $p < .001$, $d = 0.89$, with participants reporting paying more attention when learning with positive instructors ($M = 3.03$, $SD = 0.97$) than with negative instructors ($M = 2.08$, $SD = 1.15$). There was a statistically significant effect of activity, $F(1, 115) = 12.95$, $p < .001$, $d = 0.60$, with participants reporting

Table 3 Mean rating and standard deviation of the postquestionnaire questions

Emotion of instructor	Mean rating (and SD)									
	Motivated		Difficult		Effort		Enjoyed		Like More	
	M	SD	M	SD	M	SD	M	SD	M	SD
Happy instructor	3.13	0.97	2.97	1.27	2.86	0.92	2.59	1.18	2.28	1.22
Content instructor	2.93	1.02	2.60	1.07	2.50	0.90	2.47	0.90	2.30	0.99
Bored instructor	1.53	0.86	2.67	1.12	2.40	0.90	1.63	0.72	1.40	0.68
Frustrated instructor	2.66	1.14	3.38	0.94	2.38	1.19	1.97	0.98	2.10	1.21

paying more attention when learning with active instructors ($M = 2.90$, $SD = 1.08$) than with passive instructors ($M = 2.23$, $SD = 1.17$). Additionally there was a significant interaction, $F(1, 115) = 6.30$, $p = .013$. To understand the interaction, t -tests were run using Bonferroni corrections ($\alpha = 0.025$), separating the analyses based on valence. For the positive emotions, there was no difference between the ratings of the happy instructor ($M = 3.13$, $SD = 0.97$) and the ratings of the content instructor ($M = 2.93$, $SD = 1.02$), $t(58) = 0.78$, $p = .439$. For the negative emotions, participants reported paying more attention to the frustrated instructor ($M = 2.66$, $SD = 1.14$) than the bored instructor ($M = 1.53$, $SD = 0.86$), $t(52.02) = 4.25$, $p < .001$, $d = 1.12$. Consistent with the positivity principle, participants reported that they paid more attention to the material when the instructor was positive than when the instructor was negative.

Participants were then asked to rate their agreement to the statement, “The information in the lesson was difficult for me.” Column 2 of Table 3 displays the means and standard deviations for this question. There was no statistically significant effect of valence, $F(1, 114) = 1.39$, $p = .241$. There was a statistically significant effect of activity, $F(1, 114) = 7.00$, $p = .009$, $d = 0.49$, with participants rating the active instructors ($M = 3.17$, $SD = 1.13$) as more difficult than the passive instructors ($M = 2.63$, $SD = 1.09$). Additionally, there was no significant interaction, $F(1, 114) = 0.73$, $p = .396$. Not consistent with the positivity principle, participants reported a similar level of difficulty between the positive and negative instructors.

Participants were then asked to rate their agreement to the statement, “I put in a lot of effort to understand the information in the lesson.” Column 3 of Table 3 displays the means and standard deviations for this question. There was no statistically significant effect of valence, $F(1, 114) = 2.57$, $p = .112$. There was no statistically significant effect of activity, $F(1, 114) = 0.88$, $p = .349$. Lastly, there was no significant interaction, $F(1, 114) = 1.11$, $p = .294$. Not consistent with the positivity principle, participants reported a similar level of effort expended between the positive and negative instructors.

Participants were then asked to rate their agreement to the statement, “I enjoyed learning about this information.” Column 4 of Table 3 displays the means and standard deviations for this question. There was a statistically significant effect of valence, $F(1, 114) = 17.02$, $p < .001$, $d = 0.76$, with participants reporting more enjoyment when learning with positive instructors ($M = 2.53$, $SD = 1.04$) compared to learning with negative instructors ($M = 1.80$, $SD = 0.87$). There was no statistically significant effect of activity, $F(1, 114) = 1.64$, $p = .203$. There was no significant interaction, $F(1, 114) = 0.36$, $p = .548$. Consistent with the positivity principle, participants reported enjoying the lesson more with a positive instructor compared to a negative instructor.

Lastly, participants were asked to rate their agreement to the statement, “I would like more lessons like this one.” Column 5 of Table 3 displays the means and standard deviations for this question. There was a statistically significant effect of valence, $F(1, 114) = 7.80$, $p = .006$, $d = 0.51$, with participants reporting higher levels of agreement when learning with positive instructors ($M = 2.29$, $SD = 1.10$) compared to when learning with negative instructors ($M = 1.75$, $SD = 1.03$). There was no statistically significant effect of activity, $F(1, 114) = 3.13$, $p = .080$. There was no significant interaction, $F(1, 114) = 3.59$, $p = .061$. Consistent with the positivity principle, participants reported that they would like more similar lessons if the instructor was positive than negative.

Although somewhat mixed, the postquestionnaire results support the positivity principle and hypothesis 3 when the focus is on affective perceptions of the lesson (motivated, enjoyed, and like more) but not for cognitive perceptions (difficulty and effort). In sum, in partial support of hypothesis 3, participants reported that, with a positive instructor, they were more motivated to pay attention, enjoyed the lesson more, and would like more similar lessons but did not report experiencing more effort or experiencing less difficulty.

Do Learners Learn More From Positive Instructors?

In the last step in the cognitive affective model of e-learning, learners should have a better understanding of the material presented in the lesson. We predict this will be more true for positive instructors compared to negative instructors. The mean and standard deviation of the posttest are reported in Table 4. The posttest was examined using a 2×2 ANOVA to determine whether there were any differences based on groups. There was no statistically significant effect of valence, $F(1, 115) = 1.65$, $p = .201$, no statistically significant effect of activity, $F(1, 115) = 1.15$, $p = .286$, and no significant interaction, $F(1, 115) = 0.04$, $p = .852$. Not consistent with the positive principle and hypothesis 4, there were no differences between the performance on the posttest between the different groups.

Discussion

Empirical Contributions

The present study shows that learners recognize and relate to whether a virtual instructor displays positive or negative emotional tone. Learners were able to differentiate the positive instructors from the negative instructors consistently across the four emotions. However, learners struggled more with identifying the active/passive dimension. Additionally, positive instructors were rated as better at facilitating learning, more credible, more human-like, and more engaging. Furthermore, positive instructors encouraged students to pay more attention to the lesson, promoted more enjoyment of the lesson, and increased students' desire to learn more from lessons similar to this one. However, emotional tone did not have an effect on performance on a delayed test.

Table 4 Mean rating and standard deviation of the postquestionnaire questions

Emotion of instructor	Mean posttest score (and SD)	
	M	SD
Happy instructor	0.44	0.21
Content instructor	0.48	0.15
Bored instructor	0.43	0.17
Frustrated instructor	0.40	0.19

Theoretical Implications

The results are partially consistent with the cognitive affective model of e-learning. Each of the first three predictions was upheld to some degree but the fourth prediction was not. This may indicate that learners may need something more from animated instructors in order to lead to the last step, improved learning.

Practical Implications

This study has practical implications for how to design online learning experiences that involve onscreen agents. In particular, this study confirms the call to focus on the social and emotional features of onscreen agents in addition to the cognitive information-presenting features (e.g., Mayer et al. 2006; Wang et al. 2008). Consistent with the positivity principle, there is some evidence that virtual instructors should exhibit a positive emotional tone during instruction. In light of the finding that learners rated the positive instructors as more able teachers and more trustworthy, it may be beneficial to create positive virtual instructors for virtual classrooms. This study shows that this goal can be accomplished through voice and gesture. However, more research is necessary to determine what specifically in a voice and in a gesture is considered positive by learners.

Limitations and Future Directions

This was a short lesson, which took about 10 minutes for participants to view. It may be the case that in a course the impact of an instructor's emotional tone could change over a period of time. Not only could the emotional tone of an instructor overtime impact learners, but also how the emotional tone may affect the rapport built between the instructor and the learner. For example, maybe an instructor who is happy every day when lecturing has a more impactful benefit than an instructor who is happy presenting only one lecture. Having an instructor who is often happy while lecturing could build better rapport with students compared to one who is only happy once or inconsistently happy. Future research should investigate how the emotion of a virtual instructor may influence students' perceptions and learning over longer periods, like students would expect for a classroom setting.

Additionally, it is useful to determine whether these results generalize to other content areas, including those outside the field of statistics. Future research should investigate how the emotional tone of a virtual instructor impacts learning in lessons from a variety of fields.

There is also a limitation in generalizing these findings across all types of pedagogical agents. The agents in the present study were made from modeling a real-life actress giving a statistics lesson. However, this is not necessarily the only way to create pedagogical agents. Additionally, the pedagogical agent in our instructional video was an animated human, but onscreen agents may not have to be human to display the same emotions. Due to this, the results of this experiment may not generalize to other ways of designing pedagogical agents or other types of pedagogical agents. Future research should investigate the robustness of the positivity principle and the findings of this experiment across many different types of pedagogical agents.

Furthermore, participants seemed to struggle identifying the activity dimension for the content instructor and the bored instructor. This could have been due to several reasons. First, students may be less sensitive to recognizing the difference between an active instructor and passive instructor, particularly when the animated instructor is more passive. However, there was a week delay between a participant seeing the emotion of the instructor and rating the instructor's emotion. Students may have forgotten much of lesson and how the instructor presented the material during the retention interval, so it would be useful to replicate this study with an immediate test. More research should be done investigating how the ratings of instructors' emotions are influenced by the passage of time.

The benefit of having pedagogical agents that are responsive to the emotional experiences of learners has been a focus of prior research (e.g., Calvo et al. 2015; D'Mello et al. 2010, 2011; Woolf et al. 2010). By tracking the learners' emotions and having pedagogical agent respond to the emotional experiences, learners, especially those with low prior knowledge, are able to feel more confident and less frustration (Woolf et al. 2010), as well as perform better on posttests (D'Mello et al. 2010). Future research should aim to connect how affect-sensitive tutors can be improved using the information discovered in this paper. Particularly, it would be interesting to understand the instructional impact of pedagogical agents that respond to students' emotions only with positive emotions as compared to pedagogical agents that respond to students' emotions with both positive and negative emotions.

The emotional tone of the instructor affected learners' perceptions of the affective features of the lesson but not the cognitive features, which suggests that learners have more accurate access to their affective processing than their cognitive processing. Future research is warranted to address the larger issue of a possible dissociation between metacognitive awareness of affective and cognitive features of online learning.

To understand more about the relationship between the media equation theory and emotions in instructors, future research should focus on comparing how learners react to human instructors compared to virtual instructors. The media equation theory would suggest that learners react similarly to virtual instructors as they do to human instructors. However, this study did not address this question. To fully understand it, future research should directly compare the impact of emotional tone for human instructors and virtual instructors.

Acknowledgements This project was supported by Grant 1,821,833 from the National Science Foundation.

Compliance with Ethical Standards

Conflict of Interest The authors report no conflicts of interest.

Appendix A Video Script

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 5 or 6 and you lose if you roll 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.

First, you need to understand trials and outcomes. A trial is something you do. For example, you roll a die. An outcome is what happens on the trial. For instance, if you roll a die (the trial), the outcome could be that you rolled a 4.

Second, you also need to think about success and failure. A success is defined, by you, as one or more of the possible outcomes. For example, a success of rolling the die could be that you roll a number greater than 4. That means, if you roll a die, and get a 5 or 6, a success has occurred. On the other hand, a failure occurs on any trial that is not a success. So, if you defined success as rolling a number greater than 4, failure would occur if you rolled a 1, 2, 3, or 4.

Next, we should figure out the probability of success. The probability of success is the number of success outcomes divided by the total number of outcomes (including the success outcomes) if all the outcomes have an equal chance. In this case, there are 6 equally likely outcomes and 2 of them are successes, so the probability of success is 2 out of 6 or one-third. We can expect a 5 or a 6 to come up on about one-third of the times the die is rolled. The probability of success can be symbolized by the letter P .

Similarly, there is a probability of failure. This is the probability of success subtracted from 1. So, in our example, the probability of failure is 1 minus one-third which is two-thirds. The probability of failure can be symbolized as 1 minus P .

Now you know how to determine the probability of success (symbolized as P) and the probability of failure (symbolized as 1 minus P).

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, and then a 5. The sequence would be failure, failure, success, failure, success.

A sequence, like the previous example, has a probability of occurring, which is called the joint probability of a sequence. This can be found by multiplying the probabilities of each individual event. Let's take the previous example. We had failure, failure, success, failure, success. Now, we multiply the probability of each happening, so we get two-thirds (for failure), times two-thirds (for failure), times one-third (for success), times two-thirds (for failure), times one-third (for success). We can also write this as one-third squared times two-thirds cubed. So, the joint probability of this particular sequence occurring is 8 out of 243.

We can compute the probability for any specific sequence. So, let's say the number of trials in a sequence can be symbolized by the letter N and the number of successes in those trials is called R and the number of failures is N minus R . To figure out the probability of any sequence, you can use the formula displayed on the screen. We multiply the probability of success (P) by itself R times, then we multiply the probability of failure (1 minus P) by itself N minus R times, and we finally multiply those two numbers together. This is called the joint probability of a sequence.

Now you know how to compute the joint probability of a sequence of successes and failures. The next step is to figure out how many different sequences (that is, patterns of successes and failures) have that same number of successes out of N trials. For example, there are three different ways that we can have 2 successes from 3 trials:

success, success, failure.
 success, failure, success.
 failure, success, success.

As you can see, in each sequence, there are 2 successes and 1 failure. The number of different sequences having R successes in N trials is called the number of combinations. In this example, there are 3 combinations for a sequence having 2 successes out of 3 trials.

The number of combinations may be simple to work out by hand when there are just a few trials, like our previous example, but what if I asked you how many different combinations can occur for 2 successes in 5 trials? In cases like this, having a formula to find the number of combinations is quite helpful. This formula is N factorial divided by R factorial times N minus R factorial. This equation includes a factorial symbol (indicated by an exclamation point). This factorial symbol means multiply the number before the exclamation mark times the number minus one, then times the number minus two, and so on down to 1. For example, 5 factorial equals 5 times 4 times 3 times 2 times 1, which equals 120.

Now, let's finish finding the number of combinations that can occur for 2 successes in 5 trials. So, 5 factorial is equal to 120, which we just found out. Then, we divided that by 2 factorial (which is 2 times 1) times 5 - 2 factorial, or 3 factorial (which is 3 times 2 times 1). That gives us 120 divided by 12, which equals 10. This means there are 10 ways to get 2 successes in 5 trials.

Now you see how to compute the joint probability of a particular sequence that has R successes in N trials (such as failure, failure, success, failure, success) how to compute the number of combinations in which a sequence has R successes in N trials (such as 10 ways to get 2 successes out of 5 trials).

As the final step in computing binomial probability you just put those two parts together. You can figure out the probability of getting R successes out of N trials by multiplying the number of combinations for a sequence that has R successes out of N trials by the joint probability of any one of those sequences. When you do this, you are finding the probability of R successes in N trials. So, if you put that all together you get the formula on the screen. This is what we call a binomial probability.

Let's do an example to see how well our formula for binomial probability works. Suppose I want to find the probability of rolling a die 5 times and having 5 or 6 come up exactly 2 times. In this case, we want to know the binomial probability of 2 successes in 5 trials, when the probability of success is $1/3$. The binomial probability equals the number of combinations that have 2 successes in 5 trials times the joint probability of this sequence.

The number of combinations is 5 factorial divided by 2 factorial times 3 factorial, which equals 5 times 4 times 3 times 2 times 1 divided by 2 times 1 times 3 times 2 times 1 which equals 120 divided by 12 which is 10.

The joint probability of any one sequence is one-third times one-third times two-thirds times two-thirds times two-thirds, which equals 8 divided by 243.

Multiply the number of combinations times the joint probability of a sequence. We get 10 times 8 divided by 243 which equals 80 divided by 243 (or about 0.33).

This means you have about a one-third chance of rolling a die 5 times and getting 5 or 6 to come up exactly 2 times.

Now you know how to determine the probability of R successes out of N trials when the probability of success is P.

References

- Baylor, A. L., & Ryu, J. (2003). The API (Agent Persona Instrument) for assessing pedagogical agent persona. *Technology, Instruction, Cognition & Learning*, 2, 291–314.
- Calvo, R. A., D’Mello, S., Gratch, J. M., & Kappas, A. (Eds.). (2015). *The Oxford handbook of affective computing*. New York: Oxford University Press.
- Cassell, J., Sullivan, J., Churchill, E., & Prevost, S. (Eds.). (2000). *Embodied conversational agents*. Cambridge: MIT Press.
- Derry, S. J., Sherin, M. G., & Sherin, B. L. (2014). Multimedia learning with video. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 785–812). Cambridge: Cambridge University Press.
- D’Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L., & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn’t effective at promoting deep learning. *International Conference on Intelligent Tutoring Systems* (pp. 245–254).
- D’Mello, S. D., Lehman, B., & Graesser, A. (2011). A motivationally supportive affect-sensitive autotutor. *New Perspectives on Affect and Learning Technologies*, 3, 113–126.
- Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26(1), 25–36.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education*, 11(1), 47–78.
- 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. <https://doi.org/10.1016/j.chb.2020.106554>.
- Loderer, K., Pekrun, R., & Lester, J. (2021). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, 70. <https://doi.org/10.1016/j.learninstruc.2018.08.002>.
- Mayer, R. E. (2011). *Applying the science of learning*. Boston: Pearson.
- Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge: Cambridge University Press.
- Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, embodiment, and image principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). Cambridge: Cambridge University Press.
- Mayer, R. E. (2020a). *Multimedia learning* (3rd ed.). Cambridge: Cambridge University Press.
- Mayer, R. E. (2020b). Searching for the role of emotions in e-learning. *Learning and Instruction*, 70. <https://doi.org/10.1016/j.learninstruc.2019.05.010>.
- Mayer, R. E., & Greeno, J. G. (1972). Structural differences between outcomes produced by different instructional methods. *Journal of Educational Psychology*, 63(2), 165–173.
- Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, 33, 12–18.
- Mayer, R. E., Johnson, L., Shaw, E., & Sahiba, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human Computer Studies*, 64, 36–42.
- Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, 68, 837–852. <https://doi.org/10.1007/s11423-020-09749-6>.
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19, 309–326.
- 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: Taylor and Francis.
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259–282). Boston: Springer.

- Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In S. Y. Tettegah & M. P. McCreery (Eds.), *Emotions, Technology, and Learning* (pp. 131–161). Cambridge: Academic.
- Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Y. Tettegah & M. Garteimer (Eds.), *Emotions, technology, design, and learning* (p. 131–161). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-801856-9.00007-4>.
- 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. (2020). Emotional design for digital games for learning: The affective quality of expression, color, shape, and dimensionality. *Learning and Instruction, 70*. <https://doi.org/10.1016/j.learninstruc.2019.01.005>.
- 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.
- 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.
- Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human Computer Studies, 66*, 96–112.
- Woolf, B. P., Arroyo, I., Cooper, D., Burleson, W., & Muldner, K. (2010). Affective tutors: Automatic detection of the response to student emotion. In R. Nkambou, R. Mizoguchi, & J. Bourdeau (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 207–227). Berlin: Springer.

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