Computers in Human Behavior xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Social robots and virtual agents as lecturers for video instruction

Jamy Li [a,b,*], René Kizilcec [a], Jeremy Bailenson [a], Wendy Ju [b]

[a] Department of Communication, 450 Serra Mall, Building 120, Stanford, CA 94305, USA
[b] Center for Design Research, 424 Panama Mall, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Embodied pedagogical agent
Pedagogical social robot
Human–robot interaction
Virtual agent
Talking head
Video instruction

ABSTRACT

One emerging convention in video lectures is to show presentation slides with an inset video of the instructor's head. Substituting a robot or a digital agent for the video of the instructor could radically decrease production time and cost; thus, the influence of a digital agent or robot on the learner should be evaluated. Agent-based alternatives for a talking head were assessed with an experiment comparing human and agent lecturers in a video from a popular online course. Participants who saw the inset video of the actual lecturer replaced by an animated human lecturer recalled less information than those who saw the recording of the human lecturer. However, when the actual lecturer was replaced with a social robot, knowledge recall was higher with an animated robot than a recording of a real robot. This effect on knowledge recall was moderated by gender. Attitudes were more positive toward human lecturers than toward robots. An initial proof-of-concept demonstrates that although a human lecturer is preferable, robotic and virtual agents may be viable alternatives if designed properly.

1. Introduction

Video is a rich medium for communicating educational content. Instructional video has been widely used to augment traditional face-to-face education both in the classroom and online (Allen & Seaman, 2010; Berk, 2009). An emerging trend in video instruction is the use of a prerecorded video of the instructors' head presented over lecture slides (Fig. 1a). This format has been especially popular in Massive Open Online Courses (MOOCs), which provide free educational content over the internet using lecture videos. Although the "talking head" displayed in these videos is typically a human lecturer, this approach entails high video production costs. Employing an artificial agent as a talking head in instructional video may supplement face-to-face learning by providing a low-cost, accessible alternative to a video of a human lecturer. While many researchers have explored the use of virtual agents in pedagogical contexts (cf. Clark & Mayer, 2011), social robots are a form factor that may be perceived as particularly playful or nonjudgmental compared to a virtual human. Can a virtual or robotic agent be used instead of a human lecturer to deliver video instruction?

The MOOC is an influential application domain for agent-based video instruction. Several hundred MOOCs have been created, with a median enrollment rate of 40,000 students per course (Jordan, 2014). Online courses such as MOOCs allow educators to deliver lectures, tutorials, and assessments to a large audience of students. A core component of these courses is a set of instructor-narrated videos (Guo, Kim, & Rubin, 2014). Students spend the majority of their time watching video content (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2013). Some of the largest costs and time commitments have gone into video strategy for "extended" MOOCs (xMOOCs), such as those featured on platforms like Coursera or edX. Development costs for xMOOCs vary from $38,000 to $325,000 (Hollands & Tirthali, 2014, p. 12) with the largest expenses being videography and the hiring of teaching assistants (Lewin, 2013). Modern MOOC studios such as those at Harvard and elsewhere have estimated costs of $4300 per hour of high-quality produced video (Hollands & Tirthali, p. 11). Despite these high costs, video instruction is a timely medium for instruction in the age of YouTube and other online media. Past empirical studies have found video instruction to be an effective form of learning (McNeil & Nelson, 1991). Students feel more connected to an instructor when communicating via video messages than when no video is used (Borup, West, & Graham, 2012), have higher retention and motivation with video-based than with text-based instruction (Choi & Johnson, 2005), attempt more follow-up questions when lecture slides feature video of the instructor's face than when the face is missing (Guo et al., 2014), and prefer instruction with the instructor's face compared to without (Kizilcec, Papadopoulos, & Sritanyaratana, 2014). Thus, MOOCs could be improved by "post-production editing to display the instructor's head at opportune times in the video." (Guo et al., 2014, p. 42; cf. Kizilcec, Bailenson, & Gomez, 2015)

* Corresponding author at: Department of Communication, 450 Serra Mall, Building 120, Stanford, CA 94305, USA.
E-mail address: jamy@stanford.edu (J. Li).

http://dx.doi.org/10.1016/j.chb.2015.04.005
0747-5632/

Fig. 1. Participants viewed a lecture featuring (a) a human, (b) an animated human, (c) a robot or (d) an animated robot. Each video included head movement and hand gestures modeled on the human lecturer. Conditions (a) and (b) included mouth movement and natural human voice while conditions (c) and (d) employed a robotic voice without mouth movement.

Robotic and virtual agents may also improve the accessibility of pedagogical content. One of the main goals of online video instruction is to promote accessibility in education by offering high quality instruction to disadvantaged populations (Mackness, Mak, & Williams, 2010). Some evidence, however, suggests that MOOCs are preferentially used by the educated few: a survey of participants from 32 MOOCs offered by the University of Pennsylvania found that over 80% of students who took classes on Coursera had a two- or four-year post-secondary degree (Emanuel, 2013), although other courses have had better success at targeting less educated populations (e.g., King, Robinson, & Vickers, 2014). One method to improve diversity among online education participants is to localize content (White, Davis, Dickens, Leon Urrutia, & Sanchez Vera, 2014), which may be more easily achieved using social agents. Developing instructor tools that facilitate simple and inexpensive production of supplemental videos can help broaden the pool of instructors. On the student side, recent field (Kizilcec et al., 2015) and lab (Lyons, Reysen, & Pierce, 2012) studies have found that many participants in an online course prefer videos without the instructor's face, perhaps because they experience higher cognitive load when the face is included (Homer, Plass, & Blake, 2008). Artificial agents that have lower agency than a real person (particularly social robots that may appear "neutral" without a particular gender or ethnicity) may provide more standardized and less distracting cues than a human instructor. What impacts might such a substitution engender?

The current work is an initial investigation into the use of virtual and robotic agents to deliver video instruction. Four video-based conditions were compared: a human instructor (H), an animated human-like character (AH), a physically-embodied robot (R) and an animated robot-like character (AR). Fig. 1 shows still images of each condition and Table 1 compares advantages and disadvantages of each format (Section 3.2 contains links to video stimuli). To our knowledge, no prior work has compared the effectiveness of instructor videos with embodied pedagogical agents and pedagogical social robots in a video instruction scenario. If students respond in the same way to human lecturers as they do to agents playing the same role, then video production for MOOCs and other learning environments can be simplified. Conversely, if agents elicit worse learning outcomes than human

Table 1
Comparison of advantages and disadvantages of talking head alternatives.

Criterion	Image of talking head			
	Human appearance, video (H)	Human appearance, animation (AH)	Robotic appearance, video (R)	Robotic appearance, animation (AR)
Instructor visible?	Yes	No (virtual agent stand-in)	No (physical robot stand-in)	No (virtual robot stand-in)
Cost	High	Low	Low (high if robot purchased)	Low
Robot programming	No	No	Yes	No
Modeling/animation	No	Yes	No	Yes
Video equipment	Yes	No	Yes	No
Filming location	Yes	No	Yes	No
Video-editing	Yes	Minimal	Yes	Minimal
Instructor effort/time	High	Low	Low	Low
Lecturing skill needed	High	Low	Low	Low
Ease of manipulating appearance	Low (make-up, attire of human)	High (gender, race, ethnicity of model)	Low (different face masks or projected faces of robot)	High (morphology of robot)
Social cues	Instructor's cues	Programmed cues	Programmed cues	Programmed cues
Learning outcomes	H = no H (Craig et al., 2002; Mayer, 2005) H > no H (Atkinson, 2002; Mayer & DaPra, 2012) H > AH (Berry et al., 2005) H = AH (Moreno et al., 2001)	AH < H (Berry et al., 2005) AH = no AH (Craig et al., 2002)	R > no R (Brown et al., 2013; Kanda et al., 2004)	
Attitudinal measures	H > R (Park et al., 2011a, 2011b)		R < H (Park et al., 2011a, 2011b)	

lecturers, there is greater justification for the costs and effort associated with filming the instructor. This topic is of importance to the online education community as a proof-of-concept assessing alternative methods of content delivery and to the media psychology community as a way to explore interactions with agents that differ in their morphology and visual realism.

2. Embodied agents for video instruction

Embodied agents are taken to be physical robots or virtual characters that have a visually-identifiable body and communicate using voice, gesture or facial expression for instructional purposes; they are distinguished from software agents that demonstrate autonomy but have no visible "body." Embodied agents may be particularly suitable for videos in which a "talking head" accompanies lecture slides given the minimal variability in movement and facial expression of the teacher, the one-way communication from teacher to student and the high financial and time costs of video production with a real person. Two types of embodied agents are explored here: virtual agents and social robots.

2.1. Embodied pedagogical agents

Virtual agents – animated characters that are rendered using computer graphics software – have been used to deliver educational content as part of a vision that a virtual agent can advantageously influence learning (for reviews, see Clark & Mayer, 2011; D'Mello et al., 2008; Krämer & Bente, 2010). These "embodied pedagogical agents" (Paiva & Machado, 1998; Johnson et al., 2000) support face-to-face interaction in learning environments and have been extensively applied to computer-based Intelligent Tutoring Systems (ITS) (Baylor & Kim, 2004), Internet applications (Shaw, Johnson, & Ganeshan, 1999) and Virtual Learning Environments (VLE) that utilize immersive head-mounted displays or virtual reality rooms (Dede, 2009). Past work suggests the presence of a lifelike animated character, even one with minimal expressiveness, can have a positive influence on students' learning experience – a phenomenon called the "persona effect" (Lester et al., 1997). For example, people experience more positive emotions when interacting with a virtual agent that gives positive versus negative feedback (Pour, Hussain, AlZoubi, D'Mello, & Calvo, 2010), perform better on problem-solving tasks after playing an educational game with a polite versus an impolite agent (Wang et al., 2008) and report increased comfort and better concentration with an agent compared to media such as text and still images (cf. Dehn & van Mulken, 2000).

The effect of embodied pedagogical agents on learning outcomes, however, is inconclusive. Students who played an educational game about plants did not perform better on knowledge recall or attitudinal measures when they engaged with an agent compared to without one, and although knowledge transfer was higher, this was attributed to differences in game procedure rather than the presence of the agent itself (Moreno et al., 2001). Knowledge transfer was also found to be higher when participants saw the image of an agent and the agent provided instructions about a word task than when no image was shown and the instructions were presented as text (Atkinson, 2002). Other work has found that people who viewed a human-voiced agent that gestured and had facial expressions performed better on a test of knowledge transfer (but not knowledge retention) compared to either a nonmoving character or no character (Mayer & DaPra, 2012); a follow-up study revealed that the effect was only found when the agent spoke with a human voice and not a machine voice, which was attributed to a machine-voice being a negative social cue. Students who learned about climate with a virtual agent also did not perform better on knowledge retention, transfer or recall compared to when no agent was present (Craig, Gholson, & Driscoll, 2002). Past empirical work has suggested that the effect of including versus excluding a moving image of an instructor on learning outcomes is minimal (Mayer, 2005). Further, it has been proposed that for virtual agents, it is not the visual presence of the character itself but its vocalization and the instructional method embedded in its use that leads to improved learning (Krämer & Bente, 2010). It therefore appears that how a virtual agent is included may matter as much as whether one is included.

Past work has identified perceptual differences between animated agents and real people in educational settings. Berry, Butler, and De Rosis (2005) found knowledge retention to be lower with an animated agent who delivered a persuasive health message compared to a video of a real person delivering the same message, though both versions were preferred over voice-only delivery. Moreno et al. (2001) found no differences in students' learning outcomes when they viewed a bug-like animated character compared to a video of a real human. Such differences may exist because distinct brain regions are activated for videos featuring animated characters compared to those that include real people (Han, Jiang, Humphreys, Zhou, & Cai, 2005). We therefore ask whether students' learning outcomes are better with a talking head video of a human instructor or an animated version of the instructor.

RQ1. Do students retain more information when instructed with a video or an animation of a human instructor?

2.2. Pedagogical social robots

Social robots – devices with mechanical moving parts that interact in socially appropriate ways – have been employed in educational settings, particularly to teach children (for reviews, please see Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013 and Leite, Martinho, & Paiva, 2013). These "pedagogical social robots" take the role of a human instructor, although there is substantial work linking education and robotics in which the robot is the learner as well (e.g., Nicolescu & Mataric, 2001). The state of the art is to employ social robots in three main application domains: robotics/computation, in which the programming or operation of the robot is the subject itself; science education, in which the robot is used as an intermediary to study kinematics or other information; and language education, in which the robot teaches a second language (cf. Mubin et al., 2013).

Development in social robots has included significant work on creating anthropomorphic (i.e., human-like) versions of a head and upper body. A variety of platforms have been developed, with different appearances (e.g., degree of human, animal or machine likeness), morphologies (e.g., presence and shape of specific facial features), technical capabilities (e.g., degrees of freedom of movement) as well as a variety of other considerations (for reviews, please see DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002; Fong, Nourbakhsh, & Dautenhahn, 2003; Al Moubayed, Beskow, Skantze, & Granström, 2012). More recently, robotic heads have combined digital projection with a physical model of a face (e.g., Al Moubayed et al., 2012). Humanoid robot heads are of practical use for social face-to-face interaction in which nonverbal facial signals are important. While lifelike robotic heads may not be necessary for social interaction, a robotic head is particularly helpful for discourse in which emotions are conveyed or personality is expressed (Miwa, Okuchi, Itoh, Takanobu, & Takanishi, 2003). Robot head platforms mimic behaviors and emotional responses common in interpersonal interaction, such as facial expressions that employ Action Units based on the physiology of the human

Please cite this article in press as: Li, J., et al. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* (2015), http://dx.doi.org/10.1016/j.chb.2015.04.005

face (Kobayashi, Ichikawa, Senda, & Shiiba, 2003), natural neck and lip motion (Lutkebohle et al., 2010) and shared gaze for mutual attention (Mutlu, Forlizzi, & Hodgins, 2006). The behavior of a robot head may be designed by analyzing video-recorded input of a person's facial movements to support face-to-face communication with a robot (Jaeckel, Campbell, & Melhuish, 2008). Unlike virtual agents that are necessarily displayed on a digital screen, a robot's head may be co-located in a user's physical space, which may lead to various perceptual benefits; for example, perception of gaze has been found to be superior with a three-dimensional surface compared to a two-dimensional one (Kuratate, Matsusaka, Pierce, & Cheng, 2011).

Past work suggests social robots have the potential to improve learning outcomes. A longitudinal study by Kanda, Hirano, Eaton, and Ishiguro (2004) found that higher interaction time with a robot that spoke English was a positive predictor of post-study proficiency among Japanese school children. Brown, Kerwin, and Howard (2013) found students performed better on math exercises when a robotic tutor was present and exhibited various behaviors than when it was absent. School children also responded positively to a humanoid robot that taught English as a second language (Chang, Lee, Chao, Wang, & Chen, 2010) and performed better on language learning assessments when a robotic tutor provided socially supportive versus unsupportive behavior (Saerbeck, Schut, Bartneck, & Janse, 2010). In work related to higher-level education, robots have been found to be less desirable than human lecturers. In Park et al. (2011a, 2011b), university students who listened to a lecture delivered by a human instructor reported greater social attraction and acceptability of feedback than those who watched the same lecture by Aldebaran Robotics' Nao robot, particularly when the feedback from the robot was neutral or negative instead of positive. This was in spite of the fact that the robot employed the same voice and gestures as the human instructor, modeled on recordings from a previous session. Following this line of research, we assess videos of social robots in addition to virtual characters as a presentation method.

RQ2. Do students retain more information when instructed with a video of a human instructor or a video of a robot?

2.3. Visual realism and morphology

Investigating agent alternatives to a human instructor enables looking at the dimensions of visual realism and morphology. A video version of an instructor is compared to an animated version to test whether a computer-animated lecturer is comparable to its real life counterpart. Previous experiments have not found perceptual differences between filmed and animated video stimuli (e.g., Li & Chignell, 2011) and that more realistic virtual faces have only a small effect on performance of an interface (cf., Yee, Bailenson, & Rickertsen, 2007); however, it is unclear whether this would apply to online educational settings, in which a video of the human instructor may enhance the credibility of the lecture. In addition, two morphologies of an instructor are compared: the instructor had the visual appearance either of a human or of a robot. A social robot is included to evaluate whether it may be beneficial to video instruction. Both social actor theory (Reeves & Nass, 1996) and social influence theory (Blascovich, 2002) are relevant to the experience of robotic and virtual agents. Social actor theory, which predicts that people have social responses to technology, would suggest that people treat an agent instructor as if it were a real person. Social influence theory, which focuses on people's perceptions of embodied agents, would suggest that visual realism and morphology are two among many factors that characterize the degree to which people consider another person to be "present" in a virtual environment.

3. Materials and methods

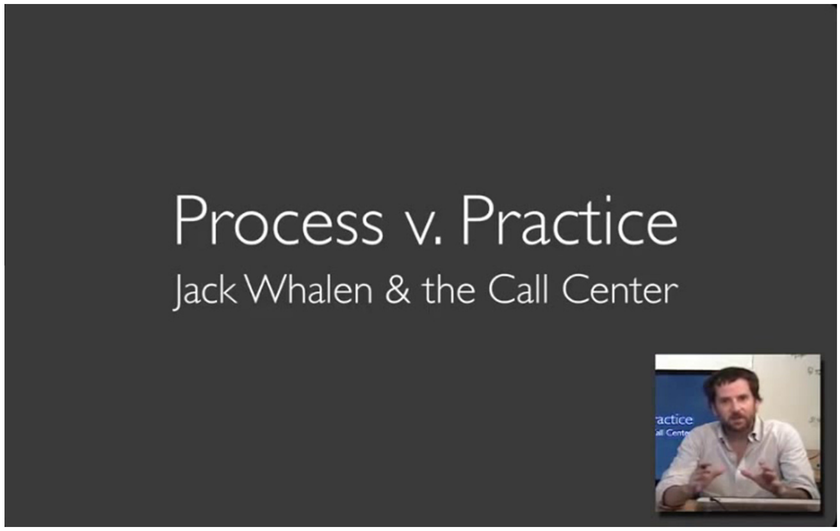
3.1. Design

A 2 (morphology: human or robot) $\times$ 2 (visual realism: non-virtual or virtual) factorial experiment was conducted to compare the effect of agent-based alternatives to a "talking head" video of an instructor on knowledge recall and attitudinal measures (Fig. 1). For morphology, the embedded video of an online lecture consisted of either the image and voice of a human (human condition) or a robot (robot condition). For visual realism, the video consisted of a real-world scene filmed using a camcorder (non-virtual) or an animation of a virtual scene rendered using computer graphics (virtual, also referred to as animated in this work). This led to four conditions for the image of the instructor: human (H), human animation (AH), robot (R) or robot animation (AR).

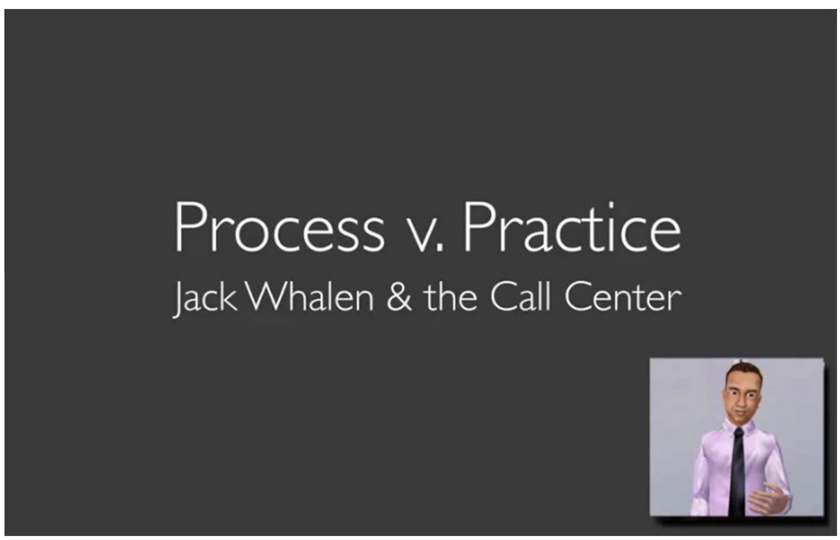
3.2. Video stimuli

A video on "Participant Observation" was obtained from the online Coursera course entitled "Human–Computer Interaction" with permission from the course instructor. A pre-test ($N = 6$) listing all Coursera course topics in August 2014 showed that the course received average ratings of appropriateness for a robot lecturer ($M = 5.20$, SD = 1.64 on a 10-item scale for the single item "Please rate how well you think a robot would be able to teach the topic"; group ratings, $M = 5.37$, SD = 1.87). Video stimuli were edited using Final Cut Pro X to control for verbal dialog, size of image and content in the following manner. All conditions contained the lecture slides as shown in the original lecture. The human condition (Fig. 1a) used the original embedded video of the course instructor, Professor Scott Klemmer. The background displayed behind the instructor in the inset video was not modified. The animated human condition (Fig. 1b) employed an agent created using Codebaby (www.codebaby.com). The audio from the original lecture was imported into the software, which automatically generated gestures and head movements for a stock agent selected based on its resemblance to the human lecturer; gestures were later edited to more closely match movements made in the original video and the resulting clip embedded over the lecture slides. Audio from the human instructor's voice was used in this condition. A background image was selected that resembled the background used in the original lecture.

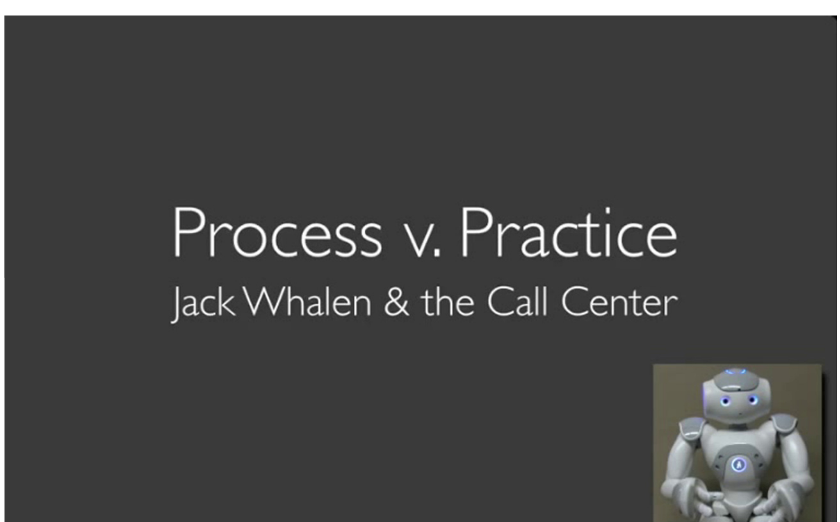
The robot condition (Fig. 1c) used a video recording of Aldebaran Robotics' Nao robot. The robot spoke a transcribed script that matched the human instructor's speech, although synthesized audio using the robot's default speech engine was used to avoid a potential mismatch between the robot's appearance and its voice. The robot's head, arm and hand movements were created by a member of the research team, who reproduced the human instructor's movements as closely as possible. The animated robot condition (Fig. 1d) used a screen-captured recording of the 3D model of Nao displayed on its software-based graphical user interface Choregraphe. The background of both robot conditions matched the body color of the robot. Videos were edited to exclude a section in which the instructor displayed a keyboard with post-it notes in the video frame, as we were not able to reproduce it across all conditions. Videos were approximately 13 min in length and the experiment took approximately 30 min. [Print version: The videos are available for reference at: http://goo.gl/kVFcl8 // Digital version: To access the videos used in this experiment, please click the images visible below.]



Video 1. h

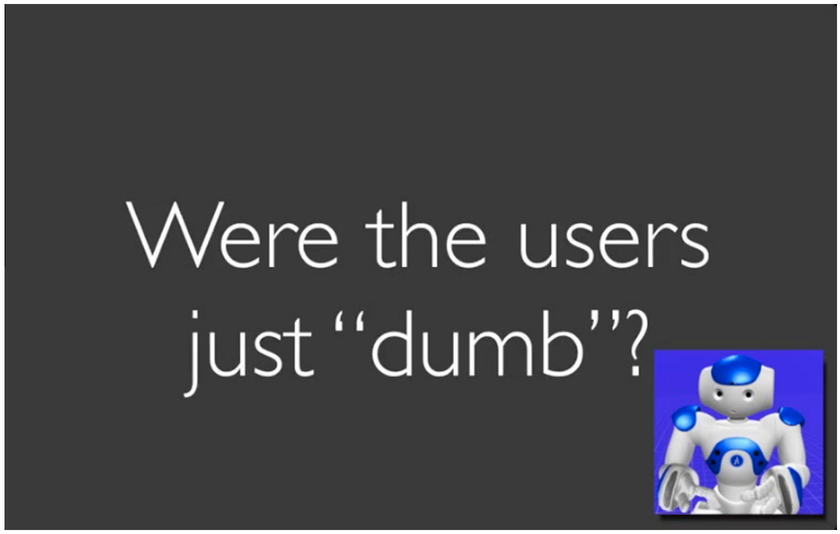


Video 2. ah



Video 3. r

Please cite this article in press as: Li, J., et al. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* (2015), http://dx.doi.org/10.1016/j.chb.2015.04.005



Video 4. ar

3.3. Participants

Forty students enrolled in an undergraduate Communication class at Stanford University between ages 18 and 25 ($M = 20.48$, $SD = 1.43$) participated in the experiment for course credit. Each condition was gender-balanced with five females and five males in each cell. No participants had previously taken the online course from which the lecture was taken.

3.4. Procedure

The study was conducted in a university computer lab instead of a naturalistic home environment to ensure that participants would view the full lecture video. Participants were randomly assigned to conditions (H, AH, R, AR). Each participant sat at a lab computer that was set up for the experiment. Consent, experimental stimuli and questionnaires were implemented using a Qualtrics online survey. Participants were instructed that they would be viewing a lecture and to "imagine that you are a student in the course." After the video, participants were asked knowledge recall questions followed by questions about their attitudes toward the instructor.

3.5. Measures

Knowledge recall and participant attitudes were assessed. Knowledge recall was measured as the number of correct responses (out of fourteen) between two multiple choice and three select-all-that-apply test questions (two of which came from the Coursera course; Appendix A). Similar to Kizilcec et al. (2014), one point was awarded per correctly selected multiple choice answer and per correctly selected or unselected all-that-apply option, then the sum was divided by the maximum possible score (14 points total). Social presence was measured using five items adapted from Lee, Peng, Jin, and Yan (2006). Reliability for the social presence scale was high (standardized Cronbach's alpha = .83). Interpersonal attraction (or "liking") was measured using four items from Byrne (1971) with high reliability (alpha = .95). The instructor's presentation skills, the instructor's enthusiasm and the overall lecture experience were assessed using single-item questions taken from the evaluation materials delivered after completing the actual online course (e.g., "Rate the instructor's presentation skills"). All attitudinal measures were assessed using ten-point Likert scales.

4. Results

4.1. Knowledge recall

RQ1 and RQ2 asked whether participants' knowledge recall scores would be different between a video of a human lecturer and its agent alternatives. Knowledge recall scores differed significantly across experimental conditions in a linear regression with main effects for morphology, visual realism and their interaction.[1] The analysis revealed a significant morphology × realism interaction, $t(36) = 2.21$, $p = .033$, $\eta^2 = .12$. Participants who saw a human instructor recalled more information if it was a recording of a real person (non-virtual) than an animation (virtual), while participants who saw a robot instructor recalled more information if the robot was animated (virtual) rather than recorded (non-virtual). No significant main effects were found. Fig. 2 illustrates this result and Table 2 provides descriptive statistics for recall scores in each condition.

An exploratory analysis of gender differences revealed that gender moderated the interaction effect of instructor presentation on recall. Fig. 3 illustrates recall scores in each condition for male and female participants. Women exhibited no significant variation across conditions, $F(3,16) = 0.34$, $p = 0.80$. In contrast, men showed substantial differences in recall across conditions. Fitting the same linear regression model for male participants revealed stronger results than in the pooled analysis, especially for the morphology × realism interaction effect, $t(16) = -2.44$, $p = .026$, $\eta^2 = .41$. Recall scores were lower with a virtual instead of a non-virtual human instructor ($t(16) = -2.44$, $p = .026$) and lower with a video of a robot instead of a human instructor ($t(16) = -2.44$, $p = .026$); if the instructor was a virtual robot, however, recall scores were higher ($t(16) = 3.33$, $p = .004$). Thus, the significant interaction effect on knowledge recall was driven by male participants.

4.2. Attitudinal measures

Overall, the results for attitudinal measures are in favor of human lecturers independent of visual realism. Ratings of social

[1] The participant with the lowest recall performance scored 2.4 standard deviations below the average recall score in their condition, but provided average ratings on attitudinal measures. The observation was included in the analysis, because it appeared to derive from the same data generating process based on a histogram of recall scores and residuals. It is noted that omission of this observation reduces the significance of the morphology × realism interaction effect to $p = .059$.

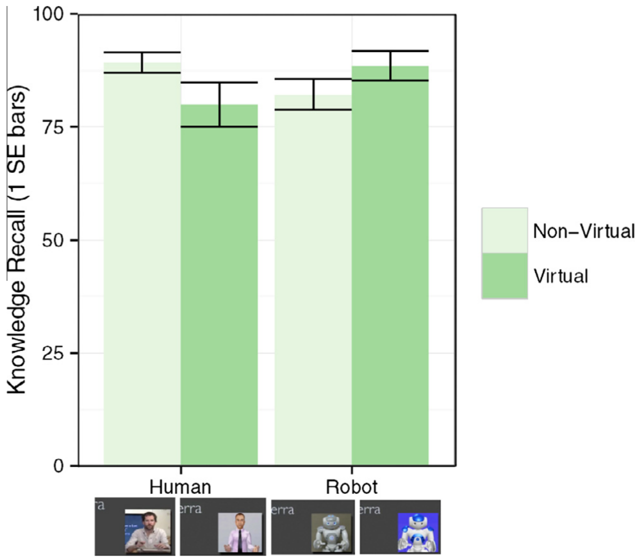


Fig. 2. Participants' knowledge recall (% questions correct) was higher with the non-virtual rather than virtual version of the human instructor, but lower with the non-virtual rather than virtual version of the robot agent. Error bars show SE.

Table 2
Descriptive statistics by experimental condition for each outcome measure.

Morphology	Visual realism			
	Non-virtual		Virtual	
	M	SD	*M*	SD
Knowledge recall				
Human	.893	.069	.800	.154
Robot	.821	.108	.886	.102
Social presence				
Human	5.74	1.41	5.16	1.80
Robot	3.94	1.62	4.12	1.85
Interpersonal attraction				
Human	5.68	2.27	5.63	2.02
Robot	2.85	1.69	3.58	2.46
Presentation				
Human	5.44	1.94	6.00	1.89
Robot	3.50	1.84	3.50	2.27
Enthusiasm				
Human	4.00	1.94	5.90	1.97
Robot	3.30	1.34	3.90	2.64
Overall experience				
Human	5.80	1.48	5.56	2.13
Robot	4.20	1.23	4.40	2.41

presence, liking (interpersonal attraction), presentation skills, and the overall lecture experience were higher for human than robot lecturers, as evidenced by a significant main effect of morphology in linear regressions: social presence, $t(36) = -2.40$, $p = .022$, $\eta^2 = .16$; liking, $t(36) = -2.97$, $p = .005$, $\eta^2 = .26$; presentation skills, $t(35) = -2.12$, $p = .041$, $\eta^2 = .25$; overall experience, $t(36) = -1.92$, $p = .063$, $\eta^2 = .13$. The finding for interpersonal attraction is illustrated in Fig. 4. Participants rated the instructor's enthusiasm somewhat higher with the animated human compared to the video of the human instructor, $t(36) = 2.10$, $p = .043$, $\eta^2 = .10$. No significant interaction effects were found for any attitudinal measures.

5. Discussion and future work

This work evaluated student responses to an online video lecture in which the human lecturer was replaced with either a virtual character or a robot. Our results suggest that agents have the

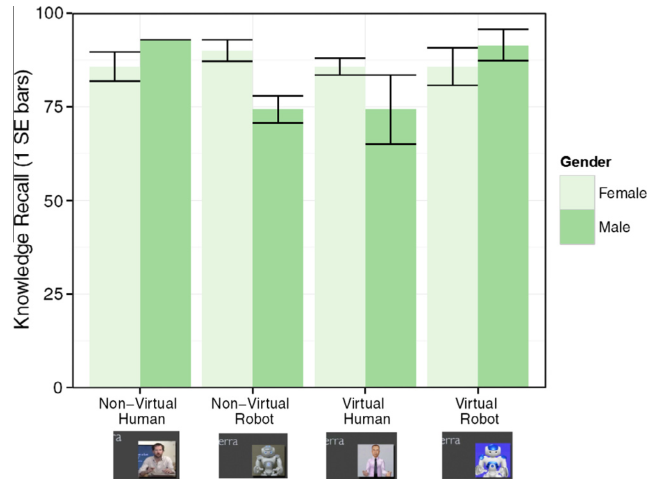


Fig. 3. Participants' knowledge recall (% questions correct) was higher with the non-virtual rather than virtual version of the human instructor, but lower with the non-virtual rather than virtual version of the robot agent. Error bars show SE.

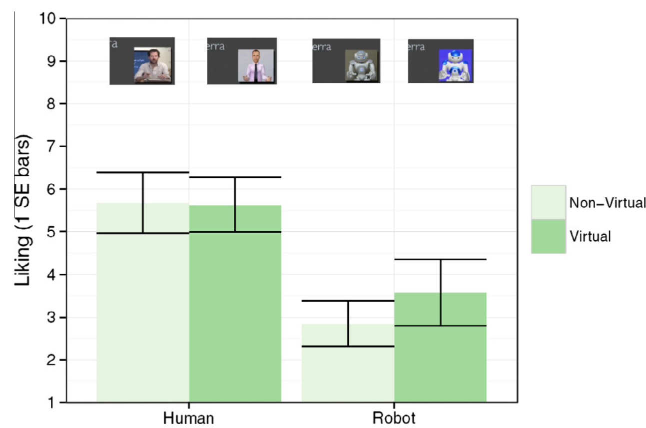


Fig. 4. Participants' ratings of liking (1–10 Likert scale) were higher with a human rather than robot instructor. Error bars show SE.

potential to be effective alternatives to human instructors in instructional videos if designed well. Knowledge recall was similarly high with either a video of the human instructor or a virtual robot stand-in, but lower with either an animated human or a video of a robot. This finding implies that certain presentations of the instructor hindered information retention more than others, given that narration and slide content was held constant between conditions. A potential explanation for the decrease in recall ability is an increase in extraneous cognitive processing (Sweller, 1994) that results from the apparent inconsistency of an animated human instructor (with a human voice) and a video recording of a robot instructor (with a synthesized voice). The novelty of these presentation formats could distract students and thereby hinder information retention. Prior work found that pedagogical agents affect learning outcomes based on the instructional method embedded in their use (Moreno et al., 2001) and the vocalization process (Bente et al., 2008) rather than their visual representation. The current work provides preliminary evidence that the consistency of the instructor's representation with student expectations could be an additional factor that affects learning. Further analysis suggested knowledge recall was only affected for male participants. Building on prior work in the agent literature (cf., Yee et al., 2007), higher visual realism (here, a video recording instead of an animation of the instructor) increased recall performance, but only for males. It is unclear why gender moderates the effects on

Please cite this article in press as: Li, J., et al. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* (2015), http://dx.doi.org/10.1016/j.chb.2015.04.005

knowledge recall in this study. Notably, the gender of affected students matches the instructor's gender, which warrants further research. While agent substitutions for a human talking head may be acceptable in some circumstances, additional work is needed to derive explicit use cases.

This work has several limitations. First, results may not be generalizable beyond the stimuli used in the experiment. Only one instance of a talking head is used for each category (e.g., virtual agents with differing ethnicities or robots with differing morphologies are not compared). A robot that better resembles a real person, for example, may elicit a more favorable response than found in this work. Similarly, while we aimed to employ an animated human that resembled the human lecturer as closely as possible, there were several differences in appearance, including in the formality of attire and presence of facial hair, which may have confounded results. An additional concern is that the specific robot employed in this study appeared cute and friendly yet did not elicit greater liking among participants. The robot's resemblance to a child or toy may have detracted from its status as a course lecturer, particularly as it played the role of a knowledgeable expert; results may differ depending on which role the agent plays, such as a mentor or motivational partner (Baylor & Kim, 2005). While characteristics such as the morphology, race or gender of an agent influence perception (e.g., Baylor & Kim, 2004; Gulz & Haake, 2006), we tested an agent with a human appearance to investigate the effect of visual realism. We therefore selected a commercially available animated agent and social robot as an initial step to evaluate the potential of agent-delivered online lectures.

This study was a laboratory experiment rather than a field study in a naturalistic learning environment (e.g., students in an online course). This provided more control of participants' attention and contextual factors while they viewed the videos. In a real-world situation with more distractors, however, students could be more strongly influenced by verbal rather than visual cues, particularly over longer periods of time (the present stimuli lasted only 13 min). The scope of this work was also limited to evaluating the influence of "talking head" videos and did not consider variations of the primary video format (e.g., videotape of a live classroom lecture or digital tablet-style animated video; cf., Guo et al., 2014). In this evaluation, a more comprehensive assessment of learning outcomes, besides the specific measure of knowledge recall employed, could help distinguish between different types of learning and is meant as future work.

An open question that warrants further exploration is why participants expressed more positive attitudes toward the human than robot lecturer. The reason could include the use of synthetic speech, the degree of facial expressiveness and the degree of visual similarity with a person. The framing of the robot's agency – whether it is an autonomous agent or an avatar controlled by a person – could also affect students' performance and learning experience. Moreover, the provision of a rationale for the robot – for instance, as a way to standardize experience among learners and teachers, particularly for MOOCs that have been translated from another language – might change perception of the robot and the lesson.

6. Conclusion

Artificial agents can enable educators to deliver video instruction without the need for video recording and editing – if these robots and virtual agents are acceptable to learners. This work demonstrated that students' recall of instructional video content may be affected when a video of a human lecturer is replaced with either an embodied pedagogical agent or a pedagogical social robot. The findings highlight that the effectiveness of agent-based alternatives to human lecturers in video instruction relies on appropriately-designed agents.

Acknowledgements

This research was funded by Stanford University's Institute for Research in the Social Sciences (IRiSS) and the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors thank Professor Scott Klemmer for use of the video from his online course entitled "Human-Computer Interaction."

Appendix A

Scale and item
Knowledge recall (standardized Cronbach's alpha = .31)
Check all statements that are generally true. Self report generally yields a more accurate picture of behavior than participant observation...
The example of Walmart asking its customers for feedback on aisle/organization was used to illustrate...
Which of the following do we obtain from participant observation? Goals of users | Current practice of users | Guidance on what people ask for | Quantitative data on user attitudes
Which of the following inspired participant observation techniques used in human computer interaction?...
Check all statements that are generally true. The apprenticeship model is suitable for participant observation...
Social presence (standardized Cronbach's alpha = .83)
I felt as if I were interacting with an intelligent being.
I felt as if I were accompanied by an intelligent being.
I felt as if I were alone. (*reversed*)
I paid attention to the instructor.
I felt involved with the instructor.
Interpersonal attraction (standardized Cronbach's alpha = .95)
I liked the instructor.
I think I could work with the instructor.
I would like to spend more time with the instructor.
I think the instructor could be a friend of mine.
Presentation skills
Rate the instructor's presentation skills.
Enthusiasm
Rate the instructor's enthusiasm.
Overall experience
Rate your overall lecture experience.

References

Allen, I. E., & Seaman, J. (2010). *Learning on demand: Online education in the United States, 2009*. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416.

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(2), 95–115.

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In: *Intelligent tutoring systems* (pp. 592–603). Berlin Heidelberg: Springer.

Bente, G., Rüggenberg, S., Krämer, N. C., & Eschenburg, F. (2008). Avatar-assisted net-working. Increasing social presence and interpersonal trust in net-based collaborations. *Human Communication Research, 34*(2), 287–318.

Berk, R. A. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning, 5*(1), 1–21.

Berry, D. C., Butler, L. T., & De Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies, 63*(3), 304–327.

Please cite this article in press as: Li, J., et al. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* (2015), http://dx.doi.org/10.1016/j.chb.2015.04.005

Blascovich, J. (2002). A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th international conference on collaborative virtual environments* (pp. 25–30). ACM.

Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. *The Internet and Higher Education, 15*(3), 195–203.

Brown, L., Kerwin, R., & Howard, A. M. (2013). Applying behavioral strategies for student engagement using a robotic educational agent. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE international conference on* (pp. 4360–4365). IEEE.

Byrne, D. E. (1971). *The attraction paradigm*. Academic Press (vol. 11).

Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society, 13*(2), 13–24.

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *The American Journal of Distance Education, 19*(4), 215–227.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*(2), 428.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66–69.

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies, 52*(1), 1–22.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the 4th conference on designing interactive systems: Processes, practices, methods, and techniques* (pp. 321–326). ACM.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., & Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states. In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems*.

Emanuel, E. J. (2013). Online education: MOOCs taken by educated few. *Nature, 503*(7476). 342–342.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems, 42*(3), 143–166.

Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents—A look at their look. *International Journal of Human-Computer Studies, 64*(4), 322–339.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on learning @scale conference* (pp. 41–50). ACM.

Han, S., Jiang, Y., Humphreys, G. W., Zhou, T., & Cai, P. (2005). Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage, 24*(3), 928–935.

Hollands, F. M., & Tirthali, D. (2014). MOOCs: Expectations and reality. Full Report. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University, NY.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*(3), 786–797.

Jaeckel, P., Campbell, N., & Melhuish, C. (2008). Facial behaviour mapping—From video footage to a robot head. *Robotics and Autonomous Systems, 56*(12), 1042–1049.

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.

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distance Learning, 15*(1).

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction, 19*(1), 61–84.

King, C., Robinson, A., & Vickers, J. (2014). Online education: Targeted MOOC captivates students. *Nature, 505*(7481). 26–26.

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

Kizilcec, R. F., Papadopoulos, K., Sritanyaratana, L., 2014. Showing face in video instruction: Effects on information retention visual attention, and affect. In *Proceedings of the SIGCHI conference on human factors in computing systems, CHI'14*. Toronto, Canada: ACM.

Kobayashi, H., Ichikawa, Y., Senda, M., & Shiiba, T. (2003). Realization of realistic and rich facial expressions by face robot. In *Intelligent robots and systems, proceedings of the 2003 IEEE/RSJ international conference on* (vol. 2, pp. 1123–1128). IEEE.

Krämer, N. C., & Bente, G. (2010). Personalizing e-learning. The social effects of pedagogical agents. *Educational Psychology Review, 22*(1), 71–87.

Kuratate, T., Matsusaka, Y., Pierce, B., & Cheng, G. (2011). "Mask-bot": a life-size robot head using talking head animation for human-robot communication. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on* (pp. 99–104). IEEE.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics, 5*(2), 291–308. Springer.

Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication, 56*(4), 754–772.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI conference on human factors in computing systems* (pp. 359–366). ACM.

Lewin, T. (2013). Students rush to web classes, but profits may be much later. The New York Times.

Li, J., & Chignell, M. (2011). Communication of emotion in social robots through simple head and arm movements. *International Journal of Social Robotics, 3*(2), 125–142.

Lutkebohle, I., Hegel, F., Schulz, S., Hackel, M., Wrede, B., Wachsmuth, S., & Sagerer, G. (2010). The bielefeld anthropomorphic robot head "Flobi". In *Robotics and automation (ICRA), 2010 IEEE international conference on* (pp. 3384–3391). IEEE.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior, 28*(1), 181–186.

Mackness, J., Mak, S. & Williams, Roy (2010). The ideals and reality of participating in a MOOC. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. De Laat, D. McConnell, T. Ryberg (Eds.). *Proceedings of the 7th international conference on networked learning 2010*. Lancaster: University of Lancaster, pp. 266–275. ISBN 9781862202252.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied, 18*(3), 239.

McNeil, B. J., & Nelson, K. R. (1991). Meta-analysis of interactive video instruction: A 10 year review of achievement effects. *Journal of Computer-Based Instruction*.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325*(5938), 284–288.

Miwa, H., Okuchi, T., Itoh, K., Takanobu, H., & Takanishi, A. (2003). A new mental model for humanoid robots for human friendly communication introduction of learning system, mood vector and second order equations of emotion. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on* (Vol. 3, pp. 3588–3593). IEEE.

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

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning, 1*.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoid robots, 2006 6th IEEE-RAS international conference on* (pp. 518–523). IEEE.

Nicolescu, M. N., & Mataric, M. J. (2001). Learning and interacting in human-robot domains. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 31*(5), 419–430.

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems* (pp. 114–130). Berlin Heidelberg: Springer.

Paiva, A., & Machado, I. (1998). Vincent, an autonomous pedagogical agent for on-the-job training. In *Intelligent tutoring systems* (pp. 584–593). Berlin Heidelberg: Springer.

Park, E., Kim, K. J., & Del Pobil, A. P. (2011a). The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom. In *Social robotics* (pp. 135–141). Berlin Heidelberg: Springer.

Park, E., Kim, K. J., & Del Pobil, A. P. (2011b). Do children see robots differently? A study comparing eye-movements of adults vs. children when looking at robotic faces. In *IT convergence and services* (pp. 421–427). Netherlands: Springer.

Pour, P. A., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In *Intelligent tutoring systems* (pp. 264–273). Berlin Heidelberg: Springer.

Reeves, B., & Nass, C., 1996. *The Media Equation: How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge University Press.

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1613–1622). ACM.

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2013). Who does what in a massive open online course? *Communications of the ACM*.

Shaw, E., Johnson, W. L., Ganeshan, R. (1999). Pedagogical agents on the web. In *Proceedings of the third annual conference on autonomous agents* (pp. 283–290). ACM.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

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*(2), 98–112.

White, S., Davis, H., Dickens, K. P., Leon Urrutia, M., & Sanchez Vera, M. (2014). MOOCs: What motivates the producers and participants? *Communications in Computer and Information Science* (3).

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1–10). ACM.