2015

Designing effective video-based modeling examples using gaze and gesture cues

Kim Ouwehand
Erasmus University

Tamara van Gog
Erasmus University, vangog@fsw.eur.nl

Fred Paas
University of Wollongong, fredp@uow.edu.au

Designing effective video-based modeling examples using gaze and gesture cues

Abstract

Research suggests that learners will likely spend a substantial amount of time looking at the model's face when it is visible in a video-based modeling example. Consequently, in this study we hypothesized that learners might not attend timely to the task areas the model is referring to, unless their attention is guided to such areas by the model's gaze or gestures. Results showed that the students in all conditions looked more at the female model than at the task area she referred to. However, the data did show a gradual decline in the difference between attention toward the model and the task as a function of cueing: students who observed the model gazing and gesturing at the task, looked the least at the model and the most at the task area she referred to, while those who observed the model looking straight into the camera, looked most at the model and least at the task area she referred to. Students who observed a human model only gazing at the task fell in between. In conclusion, gesture cues in combination with gaze cues effectively help to distribute attention between the model and the task display in our video-based modeling example.

Keywords

effective, gesture, gaze, examples, designing, modeling, cues, video

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Ouwehand, K., van Gog, T. & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. Educational Technology and Society, 18 (4), 78-88.

Designing Effective Video-Based Modeling Examples Using Gaze and Gesture Cues

Kim Ouwehand[1*], Tamara van Gog[1] and Fred Paas[1,2]

[1]Institute of Psychology, Erasmus University Rotterdam, The Netherlands // [2]Early Start Research Institute, University of Wollongong, Australia // ouwehand@fsw.eur.nl // vangog@fsw.eur.nl // paas@fsw.eur.nl

[*]Corresponding author

ABSTRACT

Research suggests that learners will likely spend a substantial amount of time looking at the model's face when it is visible in a video-based modeling example. Consequently, in this study we hypothesized that learners might not attend timely to the task areas the model is referring to, unless their attention is guided to such areas by the model's gaze or gestures. Results showed that the students in all conditions looked more at the female model than at the task area she referred to. However, the data did show a gradual decline in the difference between attention toward the model and the task as a function of cueing: students who observed the model gazing and gesturing at the task, looked the least at the model and the most at the task area she referred to, while those who observed the model looking straight into the camera, looked most at the model and least at the task area she referred to. Students who observed a human model only gazing at the task fell in between. In conclusion, gesture cues in combination with gaze cues effectively help to distribute attention between the model and the task display in our video-based modeling example.

Keywords

Gestures, Video-based human modeling, Eye tracking, Split attention, Cognitive load

Introduction

Over the past decade, learning from videos in which a human model demonstrates and (often) explains how to complete a certain task, has rapidly gained popularity, both in formal and informal educational settings (e.g., YouTube). Such so-called video-based modeling examples provide an opportunity for example-based learning, which is a very effective type of instruction, especially for novice learners (for a review, see van Gog & Rummel, 2010). However, video-modeling examples come in many forms, and little is known about design characteristics that make such examples effective in terms of attention guidance and learning (van Gog & Rummel, 2010). For instance, in video examples in which the model is standing next to a whiteboard or smartboard on which the learning task that the model is explaining is visualized (a typical modern classroom situation), it is possible that the presence of the model creates a type of split-attention effect. The split-attention effect is the adverse effect on learning that is found when students have to mentally integrate information from multiple sources (Ayres & Sweller, 2014). On the other hand, gaze direction and pointing gestures made by the model can automatically trigger attention shifts (Sato, Kochiyama, Uono, & Yoshikawa, 2009). In this way, gaze and gesture cues might be able to timely guide the learners' attention toward relevant aspects of the learning material and thereby alleviate such split attention. The question addressed in the present study is: What do learners attend to in a modeling example in which the model is visible, and can the model effectively guide learners' attention by gazing or gesturing at parts of the task?

The model as a potential source of split attention

The reason why seeing the model in the video example might evoke a division of attention between the model and the task that the model is referring to, is that people's attention is automatically drawn to other people's faces. There is probably no other object that is looked at as often as the human face, and face perception might well be the most highly developed visual skill in humans, who possess an extensive neural brain circuit involved in face perception and processing (Haxby, Hoffman, & Gobbini, 2000). Moreover, it has been shown that humans prefer to look at faces from a very young age (Tzourio-Mazoyer et al., 2002).

In a study by Gullberg and Holmqvist (2006), in which observers had to listen to and recall an event described by a visible speaker, it was shown that observers focused primarily on the speaker's face. Eye tracking was used to investigate the amount of viewing time spent looking at a speaker's face in three conditions: (1) the speaker was telling about the event directly to the addressee, (2) a video (recorded in condition 1) of the speaker was presented at

life-size or, (3) that same video was presented on a 28 inch TV screen. Results showed that over 90% of viewing time was spent looking at the speaker's face (95.6%, 94.2% and 90.8% in condition 1, 2, and 3 respectively). Although observers had to recall the event the speaker talked about, the speaker did not demonstrate a task, so this study did not investigate how we attend to human modeling examples in which a task is demonstrated and explained to learners.

Even though the findings reviewed above suggest that the model's face is likely to receive a substantial amount of attention, it is unlikely that learners would look at the model 90% of the time, since they know they have to observe the demonstration and will be tested on their ability to perform that task themselves later on. Indeed, in a recent study using video-based modeling examples in which it was demonstrated how to solve a puzzle problem by manipulating objects (the model was seated behind a table; the puzzle's objects were placed on the table), half of the participants saw a version of the example in which the face of the model was visible and the other halve saw a version of the same example in which the face of the model was not visible. Learners who saw the example video in which the model's face was visible, were found to look at the model's face only about 20% of the time, but they outperformed those who did not see the model's face, after observing the example twice (van Gog, Verveer, & Verveer, 2014). These findings suggest that the attention allocated to the model does not have to result in a negative effect on learning, and that learners are quite able to efficiently divide their attention between the model and the task.

It should be noted though, that in demonstrating this puzzle problem-solving task, the model was gazing at, gesturing at, and manipulating physical objects. This is very different from lecture-style modeling examples in which a model is standing next to a whiteboard on which slides illustrating the steps in the problem-solving procedure are projected and advanced by the model clicking a remote. In such examples, if the model continues to look into the camera, there might be a higher risk of split attention, because learners have to visually search on the screen what the model is talking about, which imposes unnecessary cognitive load during learning (Wouters, Paas, & van Merriënboer, 2008). Furthermore, when learners are looking at the model's face, they might not attend timely to the task areas the model is referring to, which might result in a) problems integrating the model's explanation into a coherent mental model of the task, and b) not noticing certain changes in the problem-solving states shown in the slides, especially if the information shown in the slides is transient (i.e., prior steps are no longer visible after each new step/slide is presented; see Sweller, Ayres, & Kalyuga, 2011, on the transient information effect). The question is then, whether we would indeed find evidence that learners may have trouble attending timely to the relevant aspects of the task, and whether gaze cues and gesture cues could help to efficiently guide learners' attention through such lecture-style video-based modeling examples.

The model's gaze and gestures as attention guiding cues

In an instructional setting, making deictic gestures (pointing and tracing gestures) has been found to enhance learning (Macken & Ginns, 2014). We suggest that deictic gestures of a video-based model can function as cues to direct learners' attention toward relevant aspects of the task on crucial moments during the instruction. Research has shown that our attention to faces mainly focuses on the eyes (Vecera & Johnson, 1995) and that eye gaze is a powerful attentional cue; we tend to automatically follow other people's gaze in order to look at what they are looking at (for reviews see Birmingham & Kingstone, 2009; Langton, Watt, & Bruce, 2000). Indeed, even though the aforementioned study by Gullberg and Holmqvist (2006) showed that in general, speakers' gestures were hardly fixated at all (less than 1%); observers did relatively often fixate on those gestures that the speakers looked at themselves.

The fact that gestures were hardly fixated in the Gullberg and Holmqvist (2006) study (although it is possible that the gestures were processed through peripheral vision) is quite surprising, because gestures fulfil an important communicative function. For instance, gestures have been found to improve learning (because they capture and guide attention; Valenzeno, Alibali, & Klatzky, 2003) and can communicate information not conveyed in speech (Singer & Goldin-Meadow, 2005). In animations in which a humanoid pedagogical agent gave explanations of the learning content, Mayer and DaPra, (2012) found an embodiment effect, indicating that animated agents producing humanlike behaviour, such as emotional expression, biological movement, gestures and eye gaze, led to better learning outcomes. This effect has also been found with animated pedagogical agents (Moreno, Reislein, & Ozogul, 2010). Moreno et al. (2010) compared learning from a narrated animation with (1) an animated pedagogical agent that produced pointing gestures toward key aspects of the learning material, (2) the same animation in which the gestures

were replaced with arrow cues, and (3) static visualisations. They found that instruction with a gesturing pedagogical agent led to superior learning compared with instruction using a non-gesturing agent or static visualisations.

Furthermore, research has shown that gestures accompanying speech are perceived as an integrated whole with speech (Kelly, Creigh, & Bartolotti, 2010), and processed in parallel with the head and eye movements (Langton et al., 2000) they accompany. In sum, these results suggest that both gaze and gesture cues are automatically processed and integrated with speech (i.e., quite effortlessly, without imposing much working memory load). These cues might therefore be very useful in video-modeling examples to ameliorate the potential effects of the model's presence as a source of split attention, by guiding the learners' attention efficiently through the examples.

The present study

The present study investigated this assumption by measuring learners' visual attention allocation toward the model and the task aspects in the slides that the model was referring to in her verbal explanation. Participants watched a video-based modeling example showing a human model verbally explaining a novel problem-solving task and either looking straight into the camera (no cue condition), or making occasional gaze shifts toward specific task areas on the screen (gaze cue condition), or making occasional gaze shifts accompanied by pointing gestures toward the screen (gesture + gaze cue condition; see Figure 1 for an impression).

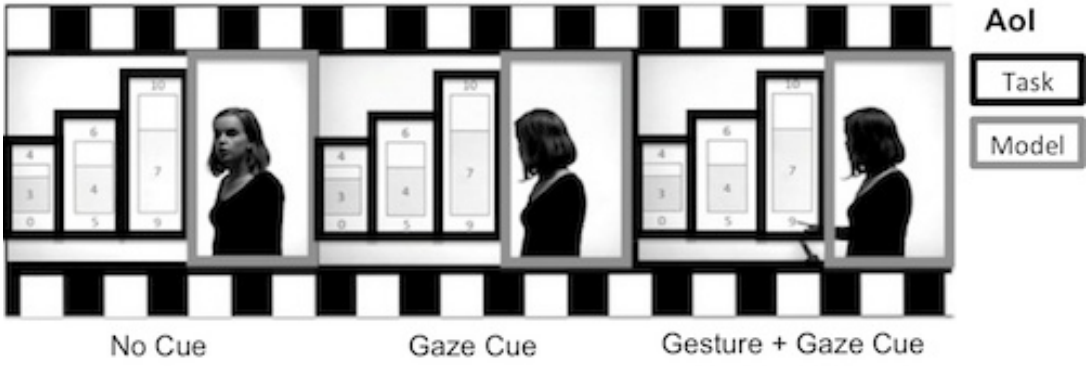


Figure 1. Snapshot of the instruction conditions with AoI's

For those scenes of the video-modeling example in which the model was referring to a specific part of the task on the screen (i.e., the small, medium, or large jug), we investigated how learners' attention allocation (fixation time) was distributed between the model and the task area referred to in that scene. In the scenes in which the model was referring to one of the jugs, students should ideally spend a substantial proportion of time looking at that task Area of Interest (AoI) instead of looking at the model, and cueing might assist them in shifting their focus to the task AoI, with gesture cues being more specific than gaze cues.

It was therefore hypothesized that participants in the no cue condition would spend more time looking at the model and less at the task AoI than those in the gaze cue condition, who would in turn spend more time looking at the model and less at the task AoI than those in the gesture + gaze cue condition. In addition, it was expected that the distribution of attention between the model and the relevant task (screen) areas would be least optimal in the no cue (split-attention) condition, more optimal in the gaze cue condition, and most optimal in the gesture + gaze cue condition. That is, learners in the no cue condition would first have to process what the model was talking about, then shift their attention toward the screen and then search the information on the current slide to determine the right task area, by which time the model might already be at a next step. In the gaze cue condition, distributing attention should be more optimal, because the model's gaze shift toward the screen would automatically induce an attention shift of the learners, meaning they would look less at the model. However, they might not look more at the task AoI, because they would still have to search for the relevant task area on the current slide as this might not be obvious. This visual search is prevented in the gesture + gaze cue condition, in which attention is not only automatically drawn to the screen, but also to the right aspect of the task on the current slide, which should therefore lead to the most optimal distribution of attention. Besides visual attention, participants' performance and perceived mental effort

on subsequent isomorphic and transfer problem solving was measured to explore whether optimal attention distribution would also lead to optimal performance and effort.

Method

Participants and design

Participants were 35 Dutch undergraduate Psychology students who participated for course credits. All participants had normal or corrected-to-normal vision. Despite successful calibration, one participant had to be excluded due to too much missing eye tracking data, leaving a sample of 34 participants for analysis (20 women, 14 men, M_{age} = 22.7 *SD* = 1.97, age range: 20–28).

Participants were randomly assigned to one of three video-based modeling example conditions. In all conditions participants studied videos of a human model standing next to a screen displaying the problem-solving task, while verbally explaining and demonstrating the solution procedure that was illustrated by a series of slides projected onto the screen. Depending on the assigned condition, the model either (1) made no gestures or gaze shifts and looked into the camera while talking (i.e., no-cue condition), (2), made no gestures, but occasionally looked at relevant task areas on the screen when these were being mentioned (i.e., gaze cue condition), or (3) looked at and made pointing and tracing gestures toward the relevant task areas on the screen when these were being mentioned (i.e., gesture + gaze cue condition). Figure 1 provides an illustration of each instruction condition.

Materials

Problem-solving tasks and video-based modeling examples

The problem-solving task consisted of an adapted version of the water-redistribution paradigm of Schmid, Wirth, and Polkehn (2003), which is based on Luchins' (1942) water jug task. Participants were presented with three jugs with a certain maximum content (displayed above each jug) containing certain amounts of water (displayed inside each jug), which they were instructed to redistribute until a goal state (displayed below each jug) would be reached (see Appendix 1 for an example). Problem solving was constrained by one task rule: The entire content of the donating jug would always be emptied into the receiving jug (i.e., no partial contents could be redistributed), unless the receiving jug would not have enough capacity for the content of the donating jug, in which case the receiving jug would be filled to the brim, leaving the donating jug with the residual. The problems used for the present experiment consisted of three-step water-redistribution problems that could only be solved with a counterintuitive strategy. Carder, Handley, and Perfect (2008) explain the counterintuitive strategy with the evaluation factor (EVF), which is the sum of differences between the current and goal states of all jugs. For example, in Figure 1, the EVF is 6 (3 + 1 + 2). A step that decreases the EVF is called perceptually consistent, because it directly brings the problem solver perceptually closer to the goal state. A counterintuitive step increases the EVF, but is sometimes a necessary step in the solution pathway. Hence, problems that should be solved with a counterintuitive strategy require problem solvers to look more than one move ahead (Bull, Espy, & Senn, 2004) and are therefore more demanding for working memory than problems that can be solved with a perceptually consistent strategy (Carder et al., 2008).

A computerized version of this water-redistribution task (Schmid et al., 2003) was created in E-prime 2.0. Participants could redistribute water through mouse clicks on the jugs. In Figure 1, for example, in order to pour water from jug A into jug B, participants first had to click on the jug they wanted to pour water from (i.e., the donating jug, in this case A; the water in this jug changed to a darker color as a visual confirmation that it was selected) and secondly, on the jug they wanted to pour water into (i.e., the receiving jug, in this case B). With the second click, the water levels of the jugs changed according to the task rule.

For each instruction condition, a video-based modeling example was created, in which the same female model explained a problem-solving task while standing next to a screen depicting the task (a typical lecture situation). In all three conditions, the model gave the same verbal explanation (see Appendix 1). The problem state depicted on the slide that was projected on the screen changed automatically to the next problem state (i.e., slide) when the model mentioned a problem-solving step being performed, so no interaction of the model with the screen was required. The

video examples in all conditions were divided in 33 scenes, consisting of six scenes in which participants were expected to look at the model (because no task-relevant areas on screen were referred to), and 27 scenes in which the model referred to task-relevant areas. Task-relevant areas were referred to verbally in the no cue condition, verbally combined with simultaneous gaze shifts in the gaze cue condition, or verbally combined with simultaneous gaze shifts and gestures in the gesture + gaze cue condition (see Figure 1). The video-based modeling examples were recorded with a digital video camera and edited in Final Cut Pro 7.0.3. All videos had the same duration of two min and were presented in E-prime 2.0.

Mental effort

After each problem participants rated how much mental effort they invested in solving it, which is an indicator of experienced cognitive load. The mental effort rating scale consisted of labeled values ranging from 0 (no effort) to 9 (extremely high effort) and was adapted from Paas (1992; see also Paas, Tuovinen, Tabbers, & van Gerven, 2003). The mental effort rating scale was also presented in E-prime 2.0 and participants responded by pressing a number on the keyboard that corresponded to the amount of mental effort they perceived to have invested in the task.

Eye-tracking equipment

The video-based modeling examples and problem-solving tasks were presented in E-prime on the 21-inch display of a Tobii 2150 (50 Hz) eye tracker, which registered participants' eye movements while they studied the modeling examples. Participants sat approximately at a 70 cm distance from the screen. To show the videos full screen they were presented with a 600 x 800 resolution. The system was recalibrated in IView prior to each example, with a 5-point calibration.

Procedure

The experiment was conducted in individual sessions of approximately 15 minutes. Participants first read a short written instruction about the basic task rules, for which they received three min. Subsequently, the system was calibrated and participants were instructed to sit as still as possible while they studied the modeling example for the first time. They were then presented with an isomorphic problem to solve (during which they could move freely). Then they rated how much mental effort they perceived to have invested in solving that problem. After this, participants studied the modeling example for the second time. They were then presented with a new isomorphic problem to solve (during which they could move freely) after which they rated how much mental effort they perceived to have invested in solving that problem. Finally, participants were presented with two transfer problems, in which the same procedure could be used to solve the problem, but the jugs had different positions, so participants could not just copy the procedure exactly as they observed it. Each transfer problem was followed by the mental effort rating scale. Participants received a maximum of one min per problem for all problems presented during the experiment.

Data analysis

Eye-movement data

The video examples were divided into 33 scenes. There were two types of scenes. In six scenes the main Area of Interest (AoI) was the model (if she was providing explanations not directly referring to the task) and in 27 scenes the main AoI was a part of the task, that is one of the three jugs (with the accompanying numbers above, in and under the jug) that the model was referring to either verbally only, verbally with gaze shifts or verbally with gaze shifts and gestures (see Figure 1). Fixations were defined as gaze points that fell within a radius of 30 pixels and together had a duration of more than 60 ms, and for each AoI in each scene, fixation duration was calculated. Fixation duration on the model and fixation duration on the relevant task area in each scene (i.e., the area being referred to by the model in that scene, which could vary across scenes) were summed only for those 27 scenes in which the task area was the main AoI, and subsequently transformed into a percentage of total fixation duration on those scenes.

Learning outcomes

For each isomorphic problem solved, a performance score was computed by dividing the number of steps in the shortest possible solution (i.e., three), by the actual number of steps a participant took to solve the problem. For example, if participant A solved a three-step problem in three steps and participant B solved the same problem in 15 steps, this would result in a score of 1 for A and 0.2 for B. The same formula was applied for transfer performance, but here, one score was obtained by determining the average performance score over the two transfer problems.

Results

Eye movement data

Table 1 shows the means and standard deviations for the fixation duration (percentage) on the model, on the task areas she referred to (averaged across scenes; hereafter called relevant task area), and on the remaining task areas (averaged across scenes). Because the overall data include those scenes in which the model did not refer to aspects of the task, the data from the task scenes only are more relevant for our hypothesis, and these were analysed with a 3 x 2 x 2 ANOVA with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor and object of attention (model vs. task) and time (first example vs. second example) as within-subjects factors.

Table 1. Percentage fixation duration in task scenes

		No cue (n = 11)		Gaze cue (n = 12)		Gesture + Gaze cue (n = 11)	
Time	Object of Attention	*M*	*SD*	*M*	*SD*	*M*	*SD*
First	Model	45.86	17.59	33.72	13.67	29.40	10.30
	Relevant Task Area	12.81	6.18	17.94	6.96	17.90	13.55
	Remaining Task Areas	24.76	8.47	25.23	9.70	23.63	5.47
	Other	16.57	9.05	23.10	11.69	29.07	16.14
Second	Model	40.75	23.15	35.17	12.34	26.96	9.95
	Relevant Task Area	15.01	10.56	16.77	7.39	20.07	9.86
	Remaining Task Areas	24.77	11.04	24.99	9.70	21.61	5.47
	Other	19.47	10.01	23.07	14.26	31.37	14.23

Note. Fixations on "other" areas are fixations to white space above, below, or next to the task and the model.

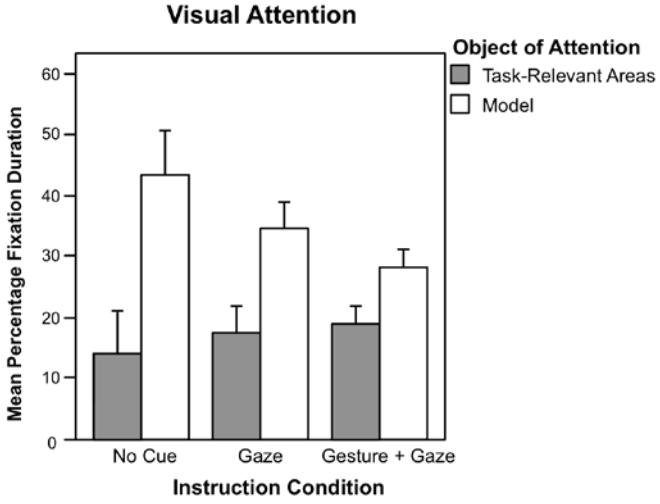


Figure 2. Interaction between instruction condition and object of attention (error bars represent standard errors + 2 SE)

The analysis showed no main effect of instruction condition, $F(2, 31) = 1.51$, $MSE = 184.07$, $p = .236$, $\eta_p^2 = .09$, or time, $F(1, 31) = 0.24$, $MSE = 33.55$, $p = .629$, $\eta_p^2 < .01$, a main effect of object of attention, $F(1, 31) = 39.08$, *MSE*

= 299.28, $p < .001$, $\eta_p^2 = .56$, and an interaction between object of attention and instruction condition, $F(2, 31) = 3.81$, $MSE = 299.28$, $p = .033$, $\eta_p^2 = .20$ (see Figure 2). There was no interaction of instruction condition and time, $F(2, 31) = 0.24$, $MSE = 33.55$, $p = .786$, $\eta_p^2 = .02$, object of attention and time, $F(2, 31) = 0.63$, $MSE = 129.85$, $p = .343$, $\eta_p^2 = .02$, or instruction condition, time and object of attention, $F(2, 31) = 0.59$, $MSE = 129.85$, $p = .560$, $\eta_p^2 = .04$.

We followed up on the significant Instruction Condition x Object of Attention interaction with multiple comparisons between instruction conditions on the attention distribution between model and task-relevant areas. We calculated a measure of attention distribution by subtracting the total fixation duration toward task-relevant areas from the total fixation duration toward the model. Results show a significant difference between the no cue and the gesture + gaze cue group, $t(20) = 2.57$, $p = .023$, $d = 1.10$, but no difference between the no cue and gaze cue group, $t(21) = 1.48$, $p = .153$, $d = 0.61$, or the gaze cue and gesture + gaze cue group, $t(21) = 1.47$, $p = .157$, $d = 0.62$. These results indicate that participants in the gesture + gaze cue group had a smaller attentional bias toward the model compared with the task-relevant areas than participants in the no cue group. Figure 2 depicts the interaction between instruction condition and object of attention.

Learning outcomes

Table 2 shows the means and standard deviations of the performance and mental effort data on the isomorphic and transfer problems. Performance and mental effort measures of isomorphic problem solving were analyzed by 3 x 2 mixed ANOVAs with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor and time (problem solving after the first and second time participants watched the video) as within-subjects factor. Performance and mental effort measures of transfer problem solving were analyzed by ANOVAs with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor.

Isomorphic problem solving performance and mental effort

For performance, results showed no main effect of instruction condition, $F(2, 31) = 2.86$, $MSE = 0.23$, $p = .072$, $\eta_p^2 = .16$, a main effect of time, $F(2, 31) = 26.21$, $MSE = 0.13$, $p < .001$, $\eta_p^2 = .46$, but no interaction, $F(2, 31) = 0.53$, $MSE = 0.13$, $p = .595$, $\eta_p^2 = .03$. The analysis of perceived mental effort invested in solving the isomorphic problems showed no main effect of instruction condition, $F(2, 31) = 1.66$, $MSE = 12.25$, $p = .208$, $\eta_p^2 = .10$, a main effect of time, $F(1, 31) = 13.80$, $MSE = 6.08$, $p = .001$, $\eta_p^2 = .31$, but no interaction, $F(2, 31) = 0.45$, $MSE = 6.08$, $p = .643$, $\eta_p^2 = .03$. As Table 2 shows, these results reflect improved performance and decreased perceived mental effort on problem solving after the second compared with the first example.

Transfer problem solving performance and mental effort

Results showed no main effect of instruction condition on transfer performance, $F(2, 31) = 1.35$, $MSE = 0.20$, $p = .275$, $\eta_p^2 = .08$, or perceived mental effort invested in solving the transfer problems, $F(2, 31) = 0.54$, $MSE = 9.33$, $p = .588$, $\eta_p^2 = .03$.

Table 2. Learning and transfer performance, and effort

	Condition	No cue ($n = 11$)		Gaze cue ($n = 12$)		Gesture + Gaze cue ($n = 11$)	
	Time	*M*	*SD*	*M*	*SD*	*M*	*SD*
Performance	First	0.42	0.51	0.50	0.52	0.10	0.32
	Second	0.81	0.39	0.90	0.29	0.65	0.46
	Transfer	0.69	0.46	0.65	0.44	0.49	0.48
Mental Effort	First	5.33	3.39	4.50	3.75	7.00	2.49
	Second	3.17	2.69	2.83	2.59	4.10	2.96
	Transfer	3.75	3.00	3.54	3.05	4.35	3.23

Conclusion

The present study focused on the question of whether gaze and gesture cues would improve the distribution of visual attention when studying a video-based modeling example in which a human model explained how to solve a novel problem. The data showed a clear trend in line with our hypothesis that students looked more at the model than at the task-relevant AoI, and that gaze and gesture cues can help shift attention from the model to what she is talking about; students in the no cue condition, looked most at the model and least at the task, while students in the gesture + gaze cue condition looked most at the task and least at the model compared with the other two conditions, and the gaze cue condition falling in between. Thus the attention toward the model gradually decreased and the attention toward the task gradually increased from the no cue, to the gaze cue to the gesture + gaze cue condition. Or in other words, participants who learned from a human model that occasionally gestured and gazed toward the task screen had a smaller attentional bias toward the model compared with the task-relevant areas than participants that learned from a model that did not gesture or gaze at the task.

Our main focus in this study was on the effect of gaze and gesture cues on learners' visual attention distribution, but we also explored whether type of instruction affected learning outcomes, although, in contrast to the eye movement data, for the performance and mental effort data this sample size was probably too low to have sufficient power to detect possible differences. Indeed, we found no significant effects on learning outcomes as measured by performance and mental effort on the isomorphic and transfer problems. It is a likely assumption that students who spent more time looking at the model than at what the model is talking about would not be able to smoothly integrate the visual and verbal information provided in the example and that this would hamper their learning (see also Mayer & DaPra, 2012; Moreno et al., 2010). Therefore, future research should replicate this experiment with larger sample sizes in order to address the question of whether better distribution of visual attention between the model and the task-related areas she is referring to, would indeed improve learning.

In sum, this study confirmed that when learning from videos, the model's face attracts a substantial amount of learners' attention, and showed that providing cues, gestures in particular seem effective in redirecting learners' attention from the model to the task areas the model is referring to. Given that the use of lecture-style online instructional videos is rapidly increasing, these findings contribute toward the development of design guidelines for such videos.

References

Ayres, P., & Sweller, J. (2014). The Split-attention principle in multimedia learning. In R. E. Mayer (Ed.) *The Cambridge handbook of multimedia learning* (pp. 206–226). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139547369.011

Birmingham, E., & Kingstone, A. (2009). Human social attention. *Annals of the New York Academy of Sciences, 1156,* 118–140. doi:10.1111/j.1749-6632.2009.04468.x

Bull, R., Espy, K. A., & Senn, T. E. (2004). A Comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry, 45,* 743–754. doi:10.1111/j.1469-7610.2004.00268.x

Carder, H. P., Handley, S. J., & Perfect, T. J. (2008). Counterintuitive and alternative moves choice in the Water Jug task. *Brain and Cognition, 66,* 11–20. doi:10.1016/j.bandc.2007.04.006

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition, 14,* 53–82. doi:10.1075/pc.14.1.05gul

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The Distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233. doi:10.1016/S1364-6613(00)01482-0

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience, 22,* 683–694. doi:10.1162/jocn.2009.21254

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4,* 50–59. doi:10.1016/S1364-6613(99)01436-9

Luchins, A. S. (1942). Mechanization in problem solving: The Effect of Einstellung. *Psychological Monographs, 54*(6), i-95. doi:10.1037/h0093502

Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical Teacher*, *36*, 596–601. doi:10.3109/0142159X.2014.899684

Mayer, R. E., & DaPra, C. S. (2012). An Embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied, 18*, 239–252. doi:10.1037/a0028616

Moreno, R., Reislein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention during learning: A Test of the persona hypothesis. *Journal of Media Psychology: Theories, Methods, and Applications*, *22*, 52–60. doi:10.1027/1864-1105/a000008

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A Cognitive-load approach. *Journal of Educational Psychology, 84,* 429–434. doi:10.1037/0022-0663.84.4.429

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. doi:10.1207/S15326985EP3801_8

Sato, W., Kochiyama, T., Uono, S., & Yoshikawa, S. (2009). Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. *NeuroImage*, *45*, 984–992. doi:10.1016/j.neuroimage.2008.12.052

Schmid, U., Wirth, J., & Polkehn, K. (2003). A Closer look at structural similarity in analogical transfer. *Cognitive Science Quarterly, 3*, 57–89. Retrieved from http://www.informatik.uos.de/schmid/pub-ps/csq-rev.pdf

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*, 85–89. doi:10.1111/j.0956-7976.2005.00786.x

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. doi:10.1007/978-1-4419-8126-4

Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *NeuroImage, 15*, 454–461. doi:10.1006/nimg.2001.0979

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A Lesson in symmetry. *Contemporary Educational Psychology, 28*, 187–204. doi:10.1016/S0361-476X(02)00007-3

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*, 155–174. doi:10.1007/s10648-010-9134-7

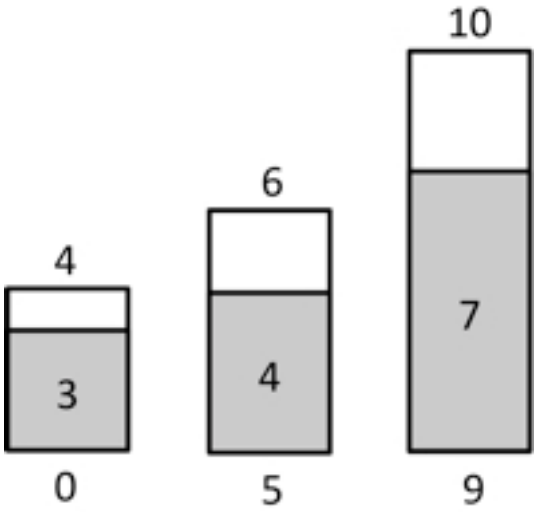
Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, *72*, 323–327. doi:10.1016/j.compedu.2013.12.004

Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, *2*, 59–87. doi:10.1080/13506289508401722

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: A Review of guidelines based on cognitive load. *Review of Educational Research, 78*, 645–675. doi:10.3102/0034654308320320

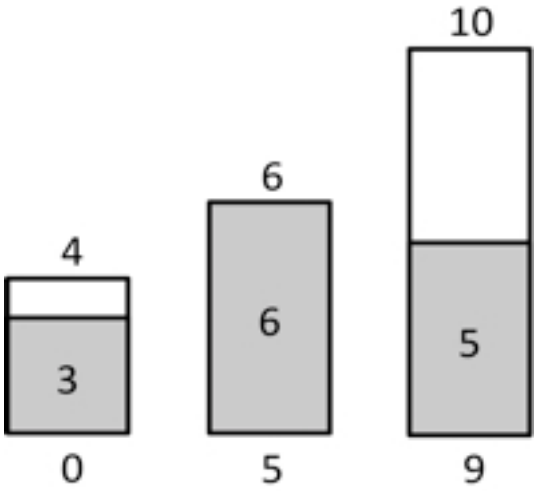
Appendix 1

Verbal script of the video



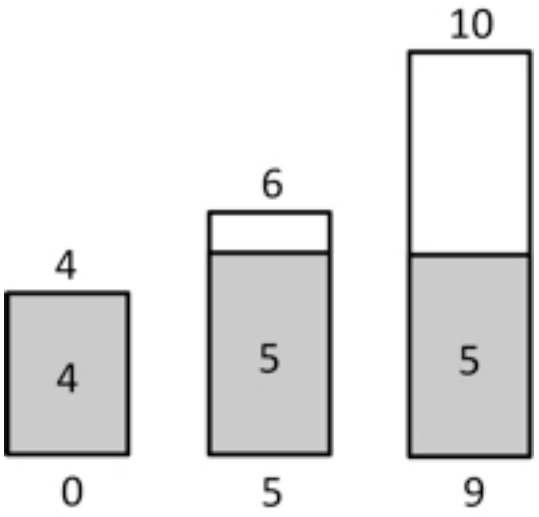
"The next problem can be solved in three steps. The correct solution can be found if you focus on the goal amount of the large jug. You can see the solution in the next formula; the current quantity of the large jug – the quantity that can be added to the medium jug + the maximum quantity of the small jug, or 7 – 2 + 4 = 9.
The first step is to pour water from the large jug into the medium jug. The medium jug will reach a quantity of 4 + 2 = 6. The large jug will reach a quantity of 7 – 2 = 5.
After the first step, the jugs look like this."

Next slide appears.



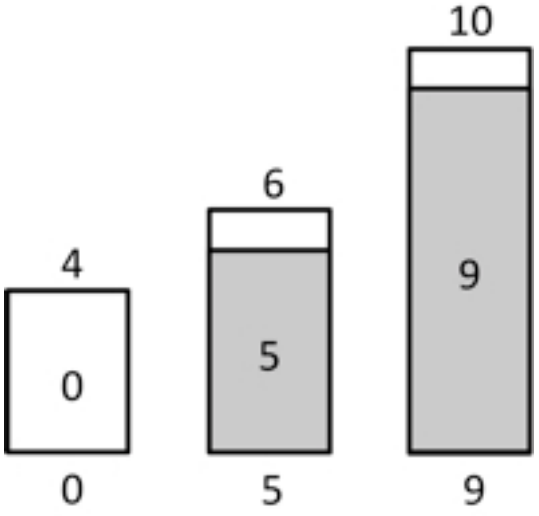
"The second step is to pour water from the medium jug into the small jug. The medium jug will reach a quantity of 6 - 1 = 5, which is equal to its goal amount. The small jug will reach a quantity of 3 + 1 = 4. After the second step, the jugs look like this."

Next slide appears.



"The final step is to pour water from the small jug to the large jug. The small jug will reach a quantity of 4 - 4 = 0, which is equal to its goal amount. The large jug will reach a quantity of 5 + 4 = 9, which is equal to its goal amount. After the final step, the jugs look like this."

Next slide appears



"The problem is now solved."

End of video.