Computers & Education 113 (2017) 98–107

Contents lists available at ScienceDirect

Computers & Education

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

# Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning

Margot van Wermeskerken [a, b, *], Tamara van Gog [a, b]

[a] Department of Education, Utrecht University, The Netherlands
[b] Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, The Netherlands

A R T I C L E  I N F O

*Article history:*
Received 24 March 2016
Received in revised form 31 March 2017
Accepted 20 May 2017
Available online 24 May 2017

*Keywords:*
Example-based learning
Eye tracking
Gaze guidance
Modeling
Demonstration

A B S T R A C T

Although the use of video examples in which an instructor demonstrates how to perform a task has become widespread in online and blended education, specific guidelines for designing such examples to optimize learning are scarce. One design question concerns the presence of the instructor or the instructor's face in the video; because faces attract attention, this might hinder learning by drawing students' attention away from the demonstration. Yet, a recent study suggested that seeing the instructor's face in demonstration video examples may help learning, presumably because the instructor's gaze offers guidance as to what s/he is attending to, which may allow anticipating what s/he is going to do. Using a different task, the main aim of the present study was to see if we could replicate this finding by comparing learning outcomes after observing video examples in which the instructor's face was not visible, or was visible and offered gaze guidance. In addition, we aimed to explore whether the effect —assuming we replicated it— would indeed be due to gaze guidance; we therefore added a third, exploratory condition in which the instructor's face was visible but offered no gaze guidance (i.e., staring straight into the camera). Students' eye movements were recorded in all conditions. We did not replicate prior findings with regard to learning outcomes: learning was neither facilitated nor compromised when seeing the instructor's face. The eye movement data suggested that learners are able to efficiently distribute their attention between the instructor's face and the task he is demonstrating.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

The use of video modeling examples in which an instructor (the model) demonstrates how to perform a task has become widespread in online (e.g., massive open online courses —MOOCs) and blended learning environments. Video modeling examples differ from other types of instructional video (e.g., podcasts: Kay & Kletskin, 2012; web-lectures: Chen & Wu, 2015; Korving, Hernández, & De Groot, 2016) in that they are recorded demonstrations of how to perform a certain task ("how to" videos), thereby falling in the realm of observational learning and example-based learning (Renkl, 2014; Van Gog & Rummel, 2010). Video modeling examples take many different forms. For instance, they might not show the instructor at all, but only show what s/he is writing (e.g., Kostons, Van Gog, & Paas, 2012; see also the examples on www.khanacademy.org), or what s/

* Corresponding author. Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands.
*E-mail address:* m.m.vanwermeskerken@uu.nl (M. van Wermeskerken).

http://dx.doi.org/10.1016/j.compedu.2017.05.013
0360-1315/© 2017 Elsevier Ltd. All rights reserved.

he is doing (clicking, typing) on the computer (e.g., McLaren, Van Gog, Ganoe, Karabinos, & Yaron, 2016), often (though not necessarily) accompanied by a verbal explanation. When the instructor is present, video examples can either take a "lecture-style" form, with the instructor standing next to a screen on which slides are projected that visualize each step in the task completion process (e.g., Hoogerheide, Van Wermeskerken, Loyens, & Van Gog, 2016), or the form of a demonstration involving object manipulation, in which case the video can show the instructor entirely or only partly (e.g., only showing the instructor's hands: Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013).

Despite the widespread use of video modeling examples in online education, specific guidelines for designing such examples to optimize learning are scarce. One prominent design question that has also been raised regarding other kinds of instructional video (e.g., weblectures: Kizilcec, Bailenson, & Gomez, 2015; Korving et al., 2016) concerns whether the presence of the instructor or the instructor's face in the video matters for learning. Using demonstration (i.e., object manipulation) examples, the present study investigated the effects of seeing the instructor's face and gaze on students' attention allocation and learning outcomes.

### 1.1. Effects of seeing the instructor's face and gaze on attention and learning

This question of whether seeing the instructor's face affects learning, is not a trivial one. After all, our attention is automatically drawn to other people's faces (e.g., Beattie, Webster, & Ross, 2010; Gullberg & Holmqvist, 2006) and even to human-like faces, such as faces of animated pedagogical agents (Louwerse, Graesser, McNamara, & Lu, 2009; see also; Levy, Foulsham, & Kingstone, 2012). Consequently, seeing the instructor's face in a video example is likely to attract students' attention and if this goes at the expense of attending to the task the instructor is demonstrating, learning might be hampered. Given that information in video examples is often transient, it is important that students attend *timely* to the part of the demonstration that the model is referring to, as they need to integrate what is being visually demonstrated with what is being verbally explained, for learning to occur (Mayer, 2014). As such, the presence of an instructor may result in a kind of 'split attention effect' (Ayres & Sweller, 2014),[1] as learners have to divide their attention between the instructor and the demonstration, which may hamper learning and may result in higher levels of cognitive load (see also Homer, Plass, & Blake, 2008).

However, one could also argue that presence of the instructor in video modeling examples might prime a stronger social response in learners (i.e., a feeling of social presence; Gunawardena, 1995) that leads to deeper cognitive processing and better learning outcomes. Moreover, if the instructor is not merely looking into the camera, but switches between looking at the camera (i.e., the learner) and the task s/he is demonstrating, then learners' attention might be guided towards the task, because humans tend to automatically follow another person's gaze (e.g., Kuhn & Martinez, 2012; Langton, Watt, & Bruce, 2000).

Indeed, research on animated pedagogical agents (i.e., humanoid or cartoon characters that provide instructions in animations) suggests that whether or not seeing an agent on screen is effective for learning, may depend on what the agent is doing. The mere presence of an agent had mixed effects on learning, with a small median effect size ($d = 0.20$) across 14 studies (i.e., image principle; Mayer, 2014). When the animated agents employed social cues (e.g., gestures, gaze guidance, eye contact), findings were generally more positive, although the median effect size across 11 studies ($d = 0.36$) was only small-to-medium (i.e., embodiment principle; Mayer, 2014). Such social cues may foster learning by helping learners to direct their attention away from the agent towards the information that the agent is referring to at that moment, which fosters integration of the visual demonstration and verbal explanation.

In line with these findings, recent research using human instructors has shown that, when no social cues are provided, the instructor's face attracts a substantial amount of attention, but does not affect learning (Kizilcec, Papadopoulos, & Sritanyaratana, 2014). However, when social (i.e., gaze) cues are employed by the model, seeing the model's face and gaze might help learners to adaptively switch their attention between the instructor and the task (cf. findings by Ouwehand, Van Gog, & Paas, 2015; regarding "lecture-style" video examples), which might foster their learning (Van Gog, Verveer, & Verveer, 2014).

In the study by Van Gog et al. (2014), learners were presented with a video example on how to solve a puzzle problem twice, with learners attempting to solve the demonstrated puzzle problem on their own after each viewing of the video example. Half of the participants saw a video example in which the instructor's face was visible, the other half did not see the face (only the hands solving the puzzle problem); participants in both conditions heard the verbal explanation that the instructor provided. The instructor's face attracted substantial attention (23% of all fixations on first viewing and 17% on second viewing), yet participants in the condition that did see the instructor's face performed better after having seen the example a second time than participants who did not see the instructor's face. Because seeing the instructor's face also meant seeing her gaze being directed towards an object prior to manipulating it, Van Gog et al. hypothesized that this contributed to the beneficial effects on learning, as learners might have automatically followed the instructor's gaze (cf. Kuhn & Martinez, 2012; Langton et al., 2000), which might have enabled them to

[1] It is 'a kind of' because in contrast to the split attention effect, in which learners have to integrate two mutually referring sources of information, the instructor's physical presence is not a relevant information source for the learning task (only the instructor's voice is).

anticipate on the instructor's subsequent actions and to timely integrate the verbal explanation with the observed actions.

The main aim of the present study was to see if we could conceptually replicate the findings of Van Gog et al. (2014) with a different demonstration task. The task consisted of learning to build a molecule, which required assembling 41 parts (atoms and bonds) in a particular order to get the target molecule (i.e., glutamine). Thus, the number of steps in this task was considerably higher than the number of steps necessary to solve the puzzle problem used by Van Gog et al. (i.e., 15 steps). The two conditions that were used by Van Gog et al. were replicated here, that is, comparing learning outcomes after observing video examples in which the instructor's face was not visible (i.e., *No Face Visible* condition), or was visible and offered gaze guidance (i.e., *Face Visible with Gaze Guidance* condition). Additionally, we exploratively included a *Face Visible without Gaze Guidance* condition (i.e., the model looking straight into the camera and not at the molecule). As in Van Gog et al.'s study, the video example was presented twice and learners had to perform the task themselves immediately after each viewing of the video example. We added a knowledge test after each example, to test whether participants remembered the information the model provided in the verbal explanation. In addition, we aimed to explore whether the effect −assuming we replicated it− would indeed be due to gaze guidance; we therefore added a third, exploratory condition in which the instructor's face was visible but offered no gaze guidance (i.e., he was staring straight into the camera; *Face Visible without Gaze Guidance* condition). We recorded learners' eye movements in all conditions.

### 1.2. Research questions and hypotheses

Regarding the main aim of this study, it was hypothesized that the findings by Van Gog et al. (2014) would be replicated. That is, we expected participants in the *Face Visible with Gaze Guidance* condition to allocate a substantial amount of attention (as measured by dwell times) to the instructor's face, meaning that they would pay significantly less attention to the task than students in the *No Face Visible* condition (H1; cf. Kizilcec et al., 2014; Ouwehand et al., 2015; Van Gog et al., 2014). We also expected participants in the *Face Visible with Gaze Guidance* condition to outperform participants in the *No Face Visible* condition on the building task (consisting of building the molecule oneself; H2a) and the knowledge test (answering questions about the molecule; H2b) − at least after having seen the example twice (cf. Van Gog et al., 2014; Wang & Antonenko, 2017). In addition, it was hypothesized that performance on the building task and knowledge test would increase after each viewing (H3a and H3b, respectively). It remains an open question whether and, if so, how attention allocation would change from first to second example study (Q1; cf. Van Gog et al., 2014).

As for the second aim of this study, we explored to what extent attention allocation was indeed guided by the instructor's gaze. Given that humans are inclined to look at other people's faces and are inclined to engage in eye-contact (Langton et al., 2000; Levy et al., 2012), we expected that learners would look more at the instructor's face (i.e., longer dwell times) when he was constantly looking into the camera as compared to when he was shifting gaze between the camera and the task (H4). In addition, it was hypothesized that learners would switch attention between the instructor's face and the task less frequently when no gaze guidance was offered as compared to when gaze guidance was provided, as learners would be inclined to automatically follow the instructor's gaze when he relocates attention from the camera towards the task and vice versa (H5; cf. Kuhn & Martinez, 2012). To gain more insight into whether and when gaze guidance might help students in timely relocating their attention from the instructor to the task (cf. Ouwehand et al., 2015), we divided the examples into three kinds of episodes for this analysis: episodes in which the instructor was 1) speaking only (i.e., not necessary for learners to look at the task), 2) speaking plus acting on the objects (i.e., necessary for learners to switch between model and task), or 3) acting on the objects only (i.e., not necessary for learners to look at the model).

## 2. Method

### 2.1. Participants and design

Participants were 69 psychology students from a Dutch university ($M_{age} = 20.2$, $SD = 2.8$; 17 male), who participated for a monetary reward (5 Euro) or course credit. Six participants had to be excluded due to bad calibration ($n = 2$; deviation of calibration exceeding 1°), or due to too high prior knowledge ($n = 4$; i.e., score of ≥3 out of 6 on the pretest, see below). The distribution of the remaining 63 participants among conditions was as follows: No Face Visible ($n = 24$), Face Visible with Gaze Guidance ($n = 24$), explorative Face Visible without Gaze Guidance condition ($n = 15$[2]). All participants had normal or corrected-to-normal vision.

[2] Because this explorative condition was included for eye movement analyses only, it had fewer participants than the conditions in which performance data were analyzed, which required more participants to have acceptable power.

### 2.2. Apparatus and materials

#### 2.2.1. Pretest

A pretest consisting of six short questions was used to assess participants' prior chemistry knowledge relevant for the task shown in the example (e.g., "What is the abbreviation of carbon?", "How many bonds can [carbon/oxygen/hydrogen] form?"). Each correctly answered question was assigned one point ($M = 0.52$, $SD = 0.92$). Participants who deviated more than two standard deviations from the mean (i.e., score $\geq 3$) were considered not to be novices and therefore excluded from the analyses.

#### 2.2.2. Eye tracking equipment

The video examples were presented to participants in SMI Experiment Center Version 3.3 on a monitor with a resolution of $1680 \times 1050$ pixels ($480 \times 300$ mm). Participants' eye movements during example study were tracked with a SMI RED250 eye tracker, which recorded binocularly at 250 Hz with SMI iView software (Version 2.8; SMI = SensoMotoric Instruments GmbH, Teltow, Germany).

#### 2.2.3. Demonstration video examples

The demonstration video examples showed a male instructor seated behind a table on which the objects (i.e., atoms and bonds from the Molymod® Organic Teacher Set, Spiring Enterprises Ltd, Billingshurst, UK) were placed in transparent containers grouped by type (i.e., carbon, hydrogen, oxygen, nitrogen, sulfur, phosphorus, single (short) bonds, and multiple (long, flexible) bonds). The videos started with the instructor explaining some basic characteristics of the atoms in front of him, by taking out one piece at a time, showing it to the camera and saying what it was (i.e., carbon atom) and its characteristic (i.e., how many bonds it could form; e.g., "This is carbon and carbon can form four bonds."). Subsequently, the instructor explained and demonstrated how to build the molecule glutamine (i.e., $C_5H_{10}N_2O_3$), which is an amino acid that consists of 20 atoms and 21 bonds. Each step of the building procedure was preceded by a brief explanation (e.g., first make a chain of five carbon atoms; on either end of the chain, one oxygen atom should be attached with the use of two flexible bonds; etc.).

The video example in the *No Face Visible* condition (video area on screen = $772 \times 279$ pixels) was created by cutting off the upper part of the video in the *Face Visible with Gaze Guidance* condition (video area on screen = $772 \times 525$ pixels). Thus, these example videos were identical in the size of the task area, in the timing of actions and verbalizations, and in total duration (247s). The video example in the explorative *Face Visible without Gaze Guidance* condition (video area on screen = $772 \times 525$ pixels) had the same verbal instruction but was slightly longer (274s), because connecting components while not looking at what one is doing, takes slightly more time. Screenshots from each condition are depicted in Fig. 1.

#### 2.2.4. Building test

The building test required participants to build the molecule themselves in a maximum of 3 min (based on a pilot study), using the same materials and starting from the same set-up as shown in the video example (i.e., the same spatial arrangements

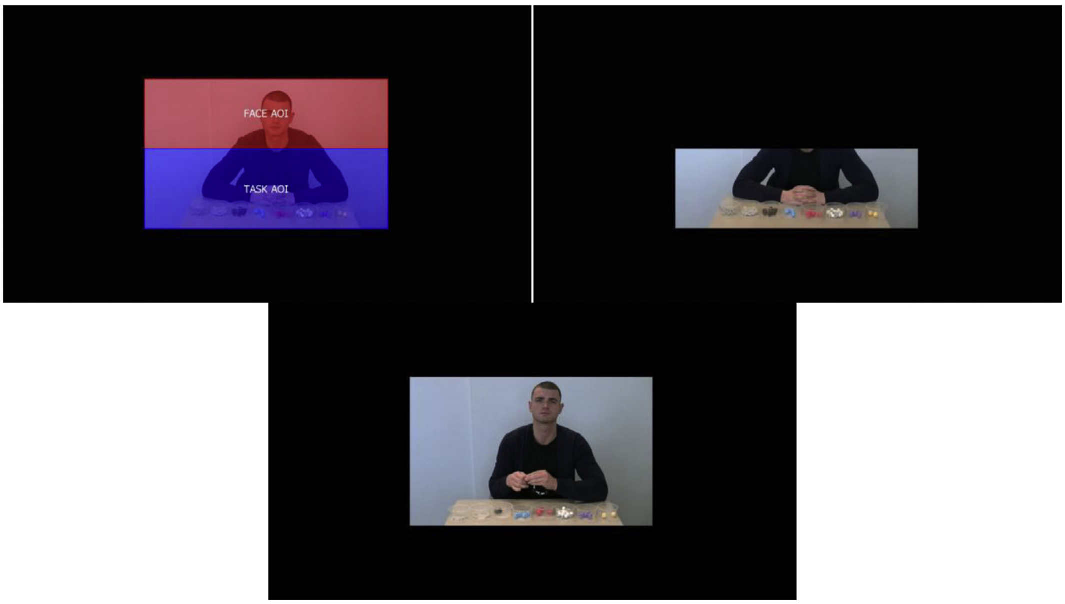


**Fig. 1.** Screen shots of the three conditions. Top left: *Face Visible with Gaze Guidance*; top right: *No Face Visible*; bottom: *Face Visible without Gaze Guidance*. The top left screen shot also shows the Areas of Interest (AoI) Face (red, upper part) and Task (blue, lower part). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of containers that contained the atoms and bonds). Note that –in order not to make the task too easy– both the set-up in the example and the set-up given to the participants during the building test contained additional atoms (and bonds) they did not need for building glutamine. Not only were there more atoms and bonds than needed for building glutamine (i.e., there were more carbon, oxygen, nitrogen, and hydrogen atoms as well as more bonds than needed in the containers), but there were also other kinds of atoms provided that are not used in building a model of glutamine at all (i.e., phosphorus and sulfur atoms).

#### 2.2.5. Knowledge test

The knowledge test consisted of 20 true/false statements, presented on the computer screen using E-Prime (Version 2.0; Psychology Software Tools, Inc., Sharpsburg USA), which assessed how much participants remembered from the examples. It consisted of 6 chemistry statements that required having attended to the instructor's explanation (e.g., glutamine is an amino acid –true), 10 statements concerning the molecule glutamine for which having paid attention to the building procedure could have sufficed (e.g., glutamine contains 4 oxygen atoms –false), and 4 transfer statements that were accompanied by images of chemical molecules that could or could not exist based on rules of atoms and the number of bonds they could engage in which were explained in the video (e.g., O=C=O can exist –true).

### 2.3. Procedure

The experiment was run in individual sessions of 30–40 min. Participants first took a short pre-test to confirm that they were indeed novices with respect to the content of the video examples. Then, participants were seated in front of the eye tracker with their head positioned in a chinrest to minimize head movements. Distance to the monitor's center was approximately 60 cm. Prior to observing each video example, the eye-tracking system was calibrated using a 5-point calibration plus 4-point validation procedure. Subsequently, participants studied the demonstration video associated with their assigned condition, built the molecule and took the knowledge test. Three minutes were allocated for the building test, when these were up (or earlier when the participant had already finished), the experimenter took photographs of the final molecule for later scoring and disassembled the parts. This procedure was then repeated (i.e., calibration, studying the video, and taking both tests).

### 2.4. Data analysis

#### 2.4.1. Eye movement data

Eye movement data were analyzed using SMI BeGaze (Version 3.2; SensoMotoric Instruments GmbH, Teltow, Germany). After removal of the participants with inaccurate calibrations (>1°, see participants section), the average calibration accuracy was 0.42° ($SD = 0.15°$) on the first viewing, and 0.40° ($SD = 0.16°$) on the second viewing; the average tracking ratio was 91.4% ($SD = 3.6\%$) on the first viewing, and 94.5% ($SD = 3.7\%$) on the second viewing.

Two areas of interest (AoI) were created on the videos: Face-AoI (776 × 243 pixels) that captured the instructor's face (upper part of the video) and Task-AoI (776 × 285 pixels)[3] that captured the task area (lower part of the video; i.e., the only visible part in the *No Face Visible* example). To analyze whether participants in the *No Face Visible* condition indeed looked significantly more at the task area than participants in the *Face Visible with Gaze Guidance* condition, the Relative Net Dwell Time was computed by dividing the total net dwell time on a particular AoI by the total duration of the video.

To analyze effects of gaze guidance, the video examples were divided into segments based on whether the instructor was explaining (i.e., Speech versus No Speech) and/or building the molecule (i.e., Action versus No Action). Since there were no occurrences in which the instructor did not do anything, three events could be distinguished: Event 1: Speech, No Action; Event 2: Speech Plus Action; Event 3: Action, No Speech. The segments were aggregated per event type (for a detailed overview of the summed duration of each event type, see Table 1). Note that the videos in the *Face Visible with Gaze Guidance* and *No Face Visible* condition had the exact same duration, but that the *Face Visible without Gaze Guidance* was slightly longer. We therefore corrected for the differences of event duration (i.e., the building procedure [Event 2 and 3] took somewhat more time and the verbal explanation [Event 1 and 2] was somewhat shorter in the *Face Visible without Gaze Guidance* condition compared to the other two conditions) by computing relative measures. First, Relative Net Dwell Time for each event was computed per participant by dividing the total dwell time (in ms) on the Face-AoI or Task-AoI for a particular event by the total duration (in ms) of that event and multiplying this by 100. Note that this relative measure controls for differences in event (and video) duration. Second, the number of transitions between areas of interest was computed for each participant and each event, by dividing the number of transitions from the Face-AoI to the Task-AoI and from the Task-AoI to the Face-AoI by two (so that relocating attention from task to face and back to the task was counted as 1 transition). To account for differences in durations of the events and differences in durations of the video examples, the number of transitions made within an event was divided by the duration (in s) of the corresponding event (resulting in a standardized number of transitions [transitions/s]).

[3] The AoIs were a few pixels larger than the size of the video on the screen, so that they included all of the video.

**Table 1**
Duration of events (ms) for Both video examples.

| | Face Visible with Gaze Guidance & No Face Visible | Face Visible without Gaze Guidance |
|---|---|---|
| Event 1 – Speech, No Action | 54798 (22.1%) | 55997 (20.4%) |
| Event 2 – Speech Plus Action | 49241 (19.9%) | 35359 (12.9%) |
| Event 3 – Action, No Speech | 143908 (58.0%) | 183147 (66.7%) |
| Total Duration | 247947 | 274503 |

### 2.4.2. Building performance

Glutamine consists of 20 atoms and 21 bonds (four of which form double bonds). With each element (atoms and bonds) yielding 1 point, a maximum score of 41 could be obtained, in the case that the glutamine molecule was fully rebuilt. Errors in terms of the number of atoms and bonds that deviated from the target molecule (i.e., less than necessary or incorrect additions) were subtracted from this maximum score, and so were errors with respect to placement or usage of incorrect bonds or elements (in order to account for errors made in the assembly while having used the correct amount of bonds and atoms). So, for instance, if a participant connected the oxygen at the place where a hydrogen atom should have been placed and vice versa, this was counted as two errors because two atoms were not in place.

### 2.4.3. Knowledge test performance

For each participant and each test moment, the number of correct answers to the true/false statements was summed, with a maximum score of 20.

## 3. Results

We used a significance level of 0.05 for all analyses and post-hoc comparisons were Bonferroni-corrected. Partial eta-squared is reported as a measure of effect size for parametric tests, with $\eta_p^2 = 0.01$, $\eta_p^2 = 0.06$, and $\eta_p^2 = 0.14$ corresponding to small, medium, and large effects, respectively. For non-parametric tests, $r$ is reported as an effect size with $r = 0.10$, $r = 0.30$, and $r = 0.50$ denoting small, medium, and large effects, respectively (Cohen, 1988).

### 3.1. Effects of (Not) seeing the instructor's face on attention

To test our hypothesis that participants in the *Face Visible with Gaze Guidance* condition would allocate a substantial amount of attention to the instructor's face, as a consequence of which they would pay significantly less attention to the task than students in the *No Face Visible* condition (H1), and whether this would change from the first to the second viewing (Q1), we analyzed the relative dwell times on the Task-AoI (data for both the Task and Face AoI are presented in Table 2). A 2 (condition: *No Face Visible* or *Face Visible with Gaze Guidance*) x 2 (viewing: 1st or 2nd) mixed factor ANOVA showed a significant main effect of condition, $F(1,46) = 66.94$, $p < 0.001$, $\eta_p^2 = 0.59$, indicating that in line with our hypothesis (H1), the relative dwell time on the Task-AoI was higher in the *No Face Visible* condition than in the *Face Visible with Gaze Guidance* condition. In addition, there was a main effect of viewing (Q1), $F(1,46) = 11.83$, $p < 0.001$, $\eta_p^2 = 0.21$, but this was qualified by a significant interaction between viewing and condition, $F(1,46) = 5.49$, $p = 0.023$, $\eta_p^2 = 0.11$. Post-hoc tests on this interaction revealed that for both viewings the *Face Visible with Gaze Guidance* condition paid less attention to the task than the *No Face*

**Table 2**
Relative Dwell Times (SD) at the Face AoI and Task AoI for each Condition, Event and Viewing.

| | | No Face Visible | Face Visible with Gaze Guidance | Face Visible without Gaze Guidance |
|---|---|---|---|---|
| Event 1 – Speech, No Action | Face AoI – 1 | – | 36.9 (11.4) | 29.5 (9.7) |
| | Face AoI – 2 | – | 43.3 (11.6) | 33.3 (9.4) |
| | Task AoI – 1 | 91.9 (6.9) | 58.7 (11.7) | 64.1 (11.2) |
| | Task AoI – 2 | 91.2 (5.9) | 52.0 (10.9) | 60.2 (9.9) |
| Event 2 – Speech Plus Action | Face AoI – 1 | – | 18.0 (10.1) | 12.7 (8.7) |
| | Face AoI – 2 | – | 22.1 (11.5) | 11.6 (7.2) |
| | Task AoI – 1 | 92.0 (7.7) | 76.6 (11.3) | 77.9 (14.4) |
| | Task AoI – 2 | 91.3 (6.9) | 73.2 (12.3) | 78.2 (10.2) |
| Event 3 – Action, No Speech | Face AoI – 1 | – | 2.5 (1.6) | 6.0 (4.3) |
| | Face AoI – 2 | – | 4.4 (4.6) | 5.8 (4.9) |
| | Task AoI – 1 | 95.7 (4.0) | 94.0 (3.4) | 87.4 (7.4) |
| | Task AoI – 2 | 95.1 (4.2) | 91.6 (7.3) | 87.8 (7.4) |
| Total | Face AoI – 1 | – | 13.3 (4.5) | 11.8 (5.3) |
| | Face AoI – 2 | – | 16.6 (5.8) | 12.3 (5.1) |
| | Task AoI – 1 | 94.0 (5.1) | 82.6 (5.7) | 81.2 (8.5) |
| | Task AoI – 2 | 93.4 (4.8) | 79.1 (7.4) | 80.8 (6.8) |

*Visible* condition (both *p*'s < 0.001). In addition, it was revealed that there was a significant decrease in relative dwell time on the Task-AoI from first to second viewing in the *Face Visible with Gaze Guidance* condition ($p < 0.001$), but not in the *No Face Visible* condition ($p = 0.442$). In other words, students who saw the instructor's face seemed to pay somewhat more attention to the face the second time they observed the example.

### 3.2. Effects of (Not) seeing the instructor's face on learning

To test our hypothesis (H2a,b) that the *Face Visible with Gaze Guidance* condition would have learned more than the *No Face Visible* condition −at least after having seen the demonstration video twice− we analyzed the performance scores on the building and knowledge tests (presented in Table 3).

#### 3.2.1. Building test

For analyzing the building test scores non-parametric tests were conducted due to violations of the normality assumption. Firstly, differences between the conditions were analyzed using Mann-Whitney U tests, for first and second built separately. In contrast to our hypothesis (H2a), this did not reveal significant differences between the *Face Visible With Gaze Guidance* and *No Face Visible* conditions neither on the 1st built, $U = 279.00$, $z = -0.186$, $p = 0.853$, $r = 0.03$, nor on the 2nd built, $U = 268.50$, $z = -0.471$, $p = 0.638$, $r = 0.07$.

Secondly, we tested whether students improved from the first to the second test moment with Wilcoxon Signed Rank tests that were performed for each group separately with building score as dependent variable and test moment (1st versus 2nd) as repeated measure. In line with our hypothesis (H3a), this analysis revealed that each group improved from the first to the second test moment with large effect sizes (Bonferroni adjusted $\alpha = 0.025$; *No Face Visible:* $z = -4.109$, $p < 0.001$, $r = 0.84$; *Face Visible With Gaze Guidance:* $z = -4.259$, $p < 0.001$, $r = 0.87$).

#### 3.2.2. Knowledge test

On knowledge test scores, a 2 (test moment: 1st or 2nd) x 2 (condition: *Face Visible with Gaze Guidance* or *No Face Visible*) mixed ANOVA was conducted to investigate whether there was an effect from seeing the instructor's face on performance and whether participants improved from the first to second viewing. In contrast to our hypothesis (H2b), this analysis did not reveal a main effect of condition, $F(1,46) = 2.28$, $p = 0.138$, $\eta_p^2 = 0.05$. Yet, a main effect of test moment was revealed (H3b), confirming that all participants' test performance improved from 1st to 2nd test, $F(1,46) = 27.93$, $p < 0.001$, $\eta_p^2 = 0.38$. There was no interaction between condition and test moment, $F < 1$.

In sum, these findings suggest that there were no differences in learning outcomes (i.e., performance on building and knowledge test; H2a and H2b, respectively) between conditions and that all participants improved on the building and knowledge test from the first to second test moment (H3a and H3b, respectively).

### 3.3. Effects of gaze guidance on attention

To explore to what extent attention allocation was indeed guided by the instructor's gaze, and when (i.e., under what events), we compared relative dwell times on the instructor's face (H4), as well as the number of transitions from the instructor's face to the task (H5), between the *Face Visible without Gaze Guidance* and *Face Visible with Gaze Guidance* condition for each event.

#### 3.3.1. Dwell times

A $2 \times 2 \times 3$ mixed ANOVA with between-subjects factor condition (*Face Visible With Gaze Guidance* or *Face Visible without Gaze Guidance*) and within-subjects factors viewing (1st or 2nd) and event (Speech, No Action versus Speech Plus Action versus Action, No Speech) was conducted on the relative dwell time on the Face-AoI (presented in Table 2). This analysis revealed main effects of condition, $F(1,37) = 5.20$, $p = 0.028$, $\eta_p^2 = 0.12$, and of viewing, $F(1,37) = 9.16$, $p = 0.004$, $\eta_p^2 = 0.20$. These effects indicate that students in the *Face Visible with Gaze Guidance* condition spent more time looking at the face than students in the *Face Visible without Gaze Guidance* condition, which is not in line with our hypothesis (H4), and that, overall, students spent more time looking at the instructor's face during the second viewing compared to the first viewing. The interaction between condition and viewing was not statistically significant: $F(1,37) = 3.91$, $p = 0.056$, $\eta_p^2 = 0.10$.

**Table 3**
Mean (SD) performance on the building test and knowledge test after both viewings for each condition.

| | Built #1 | Built #2 | Test #1 | Test #2 |
|---|---|---|---|---|
| No Face Visible | 28.1 (8.4) | 38.6 (4.8) | 12.9 (2.3) | 15.2 (2.0) |
| Face Visible with Gaze Guidance | 27.2 (9.3) | 39.0 (3.3) | 12.3 (2.3) | 14.2 (2.6) |
| Face Visible without Gaze Guidance[a] | 25.5 (10.9) | 37.7 (4.8) | 12.5 (3.0) | 15.0 (2.2) |

*Note:* The maximum score on the building test was 41 and on the knowledge test 20 (with chance level being 50%).
[a] These data are reported for completeness, but we did not analyze test performance in this condition because it was included for explorative analyses of eye tracking data only.

In addition, the analysis revealed a main effect of event, $F(2,74) = 251.01$, $p < 0.001$, $\eta_p^2 = 0.87$. Follow-up analyses revealed that most time was spent looking at the Face-AoI when the instructor was talking and not performing an action (i.e., event 1) as compared to when he talked *and* performed an action (i.e., event 2) or did not talk but performed an action (i.e., event 3, all paired comparisons: $p$'s $< 0.001$). However, this main effect was qualified by a significant interaction between event and condition, $F(2,74) = 9.87$, $p < 0.001$, $\eta_p^2 = 0.21$. Post hoc tests on this interaction revealed that in the *Face Visible with Gaze Guidance* condition more time was spent looking at the Face-AoI as compared to the *Face Visible without Gaze Guidance* condition for the events in which the instructor talked (i.e., event 1 and 2; $p = 0.011$ and $p = 0.012$, respectively), but that this was reversed for the event in which the instructor did not talk (i.e., event 3; $p = 0.029$). Finally, a significant interaction between event and viewing was revealed, $F(2,74) = 5.21$, $p = 0.013$, $\eta_p^2 = 0.11$, which seemed to indicate that for the event in which the instructor was talking but not performing an action (i.e., event 1), there was an increase in dwell time from first to second viewing on the instructor's face, $p = 0.001$, but not for the other two events (i.e., event 2 or 3, $p = 0.240$ and $p = 0.203$, respectively). There was no three-way interaction ($F < 1$).

In sum, these findings show that participants in the *Face Visible with Gaze Guidance* condition looked more at the instructor's face than participants in the *Face Visible without Gaze Guidance* condition, but only when the instructor talked (event 1 and 2); when he did not talk (event 3) the pattern was reversed, that is, participants in the *Face Visible without Gaze Guidance* condition looked more at the instructor's face than participants in the *Face Visible with Gaze Guidance* condition.

#### 3.3.2. Transitions

A similar 2 × 2 × 3 mixed ANOVA on transition data (presented in Table 4) neither revealed a main effect of condition, $F(1,37) = 2.34$, $p = 0.135$, $\eta_p^2 = 0.06$, nor of viewing, $F(1,37) = 2.06$, $p = 0.160$, $\eta_p^2 = 0.053$, nor an interaction between condition and viewing ($F < 1$). Yet, there was a main effect of event: $F(2,74) = 95.86$, $p < 0.001$, $\eta_p^2 = 0.72$. Post hoc analyses showed that most transitions were made during the events that the instructor was talking (i.e., event 1 and 2) as compared to when the instructor did not talk (i.e., event 3, both comparisons: $p$'s $< 0.001$). There was no difference in number of transitions between the two events in which the instructor was talking (event 1 and event 2; $p = 0.076$). However, this main effect was qualified by an interaction effect between event and condition, $F(2,74) = 14.64$, $p < 0.001$, $\eta_p^2 = 0.28$. Post hoc analyses on this interaction indicated that participants in the *Face Visible with Gaze Guidance* condition performed more transitions than participants in the *Face Visible without Gaze Guidance* condition for the event that involved speech accompanied by an action (event 2: $p = 0.005$). This effect was reversed for the event in which the model was only performing an action, during which participants in the *Face Visible without Gaze Guidance* made significantly more transitions than the *Face Visible with Gaze Guidance* (event 3: $p = 0.020$). There was no difference between conditions in the number of transitions during the events in which the instructor was solely talking (i.e., event 1: $p = 0.251$).

In sum, in the *Face Visible with Gaze Guidance* condition, participants switched attention between the instructor's face and the task more often when the instructor talked and acted upon the object (event 2) than participants in the *Face Visible without Gaze Guidance* condition did. Yet, when there was no need to look at the instructor's face (i.e., event 3), participants in the *Face Visible without Gaze Guidance* condition switched more often between the instructor's face and the task demonstration than participants in the *Face Visible with Gaze Guidance* condition.

## 4. Discussion

The main aim of this study was to attempt to conceptually replicate the findings by Van Gog et al. (2014). That is, we first hypothesized that participants in the *Face Visible with Gaze Guidance* condition would allocate a substantial amount of attention to the instructor's face, meaning that they would pay significantly less attention to the task than students in the *No Face Visible* condition (H1). In line with this hypothesis, we indeed found that participants who saw the instructor's face paid less attention to the task area than participants who did not see the face.

Second, we expected participants in the *Face Visible with Gaze Guidance* condition to outperform participants in the *No Face Visible* condition on the test tasks –at least after having seen the example twice (H2a,b). However, we did not replicate this finding: seeing the instructor's face had neither beneficial nor detrimental effects on learning in the present study. Our third

**Table 4**
Mean (and SD) standardized number of transitions per second for the face visible with gaze guidance and the face visible without gaze guidance condition presented for each event and viewing.

| Event | Viewing | Face Visible with Gaze Guidance | Face Visible without Gaze Guidance |
|---|---|---|---|
| Event 1 – Speech, No Action | 1st | 0.19 (0.08) | 0.15 (0.07) |
| | 2nd | 0.21 (0.08) | 0.19 (0.09) |
| Event 2 – Speech Plus Action | 1st | 0.20 (0.09) | 0.12 (0.08) |
| | 2nd | 0.21 (0.12) | 0.12 (0.08) |
| Event 3 – No Speech, Action | 1st | 0.03 (0.03) | 0.07 (0.04) |
| | 2nd | 0.04 (0.03) | 0.06 (0.05) |

*Note:* The standardized number of transitions was calculated by dividing the number of transitions within an event by the duration of that event (in s) to enable comparison between conditions.

hypothesis that participants would show an improvement in performance on both tests from first to second test moment was confirmed (H3a,b). Moreover, attention allocation changed from first to second viewing (Q1), with more attention being devoted to the instructor's face during the second viewing as compared to the first viewing of the example.

Although this lack of effect on learning outcomes is in contrast to the findings by Van Gog et al. (2014) with demonstration video examples, it does correspond with previous studies that did not show beneficial (or detrimental) effects of displaying an instructor's face when learning from weblectures (Kizilcec et al., 2014) or "lecture-style" video examples (Ouwehand et al., 2015; though due to the low number of participants, the performance data from that study should be interpreted with caution). So, it is possible that the results of Van Gog et al. are an anomaly and that seeing the instructor simply does not matter for learning outcomes.

Nevertheless, there are several differences between the task used in the present study and the task used by Van Gog et al. that are worth highlighting as they might be related to the difference in findings. First, in the current task, the molecule was continuously visible and in case of having been distracted by the instructor's face, participants could catch up with the missed step(s) by attending to the state of the molecule at that moment. Yet, in the study by Van Gog et al. (2014), information was transient, so paying attention to the right place at the right time was more crucial in order to memorize all steps.

Second, the performance scores in the present study did not take into account the extent to which participants followed the exact building procedure demonstrated in the video example. We used a photograph of the end product to score performance, because the order in which the particular steps are completed is not that important for building a molecule (in contrast to the puzzle problem used by Van Gog et al., 2014). However, it is possible that we might have found beneficial effects of seeing the instructor's face (with gaze guidance) on learning if we had recorded and scored the building *process* (i.e., gaze guidance and being able to anticipate on the instructor's actions might have a beneficial effect on remembering the order of the steps). Indeed, in the study of Van Gog et al. (2014), participants were required to reproduce the exact same procedure demonstrated in the example in order to successfully accomplish the task.

Third, the task differences possibly reduced the usefulness of gaze guidance. That is, whereas the task of Van Gog et al. (2014) consisted only of 15 steps, the task in the present study required connecting 41 elements. Consequently, the instructor was looking down at the objects for relatively long stretches of time. This may explain why participants seemed to look somewhat less at the face (13–17%) than in the Van Gog et al. study (17–23%). Indeed, the explorative analyses of the eye tracking data showed that in terms of events, participants attended most to the instructor's face in the events in which he was only explaining (i.e., event 1), as compared to occasions in which he was explaining *and* demonstrating (i.e., event 2) or only demonstrating (i.e., event 3); note that event 2 and 3, in which the instructor was building and therefore looking down at the objects a lot, together made up 78% of the video. As such, cues such as switches in the instructor's attention via his gaze, may have been less strong (more subtle) when he is looking down a lot than when the instructor looks into the camera and then towards objects.

That seeing the instructor's gaze seems to provide a cue for students as to when to switch their attention to the task, was suggested by the detailed event analyses. In contrast to our hypothesis (H4), participants in the condition in which the instructor was seen to switch his gaze from the camera to the task (condition with gaze guidance), looked at the instructor's face when he talked more often than participants in the condition in which the instructor stared straight into the camera. As anticipated (H5), they also switched their attention between the task and the instructor's face more often when the instructor was speaking and acting on the objects than participants in the condition in which the instructor stared into the camera. Interestingly, participants who saw the instructor looking only into the camera, spent more time looking at the instructor's face and made more transitions between the instructor's face and the task when he was *not* speaking (i.e., event 3), even though the instructor's gaze was not functional or meaningful. This suggests that it is difficult to avoid looking at another person's face, especially when that person's gaze is directed at yourself (which is in line with fundamental research on attention to faces and eyes: e.g., Langton et al., 2000). Note however, that sample size in this exploratory condition was rather small and future research should replicate these findings using larger sample sizes.

Seeing the instructor's gaze during example study affected students' attention differently from first to second viewing. More specifically, students were inclined to look more at the instructor's face from first to second viewing, which is in contrast to the findings of Van Gog et al. (2014), who reported a decrease in looking times from first to second viewing. One explanation for this discrepancy might be fact that we also administered a knowledge test as a result of which students paid more attention to the instructor's explanations during the second example study. Note, though, that this did not result in better performance on the knowledge test as compared to the condition in which the instructor's face was not visible.

The present study has two potential limitations. One limitation of the present study is that only psychology students participated in the study of which the majority (75.4%) was female. Hence, future studies should address this question in different samples and with different materials in order to confirm that the (lack of) findings can be generalized to different populations. In addition, the sample size was relatively small, especially for the exploratory *Face Visible without Gaze Guidance* condition. Note, however, that for the comparison of the learning outcomes in the *Face Visible With Gaze Guidance* and *Face Not Visible* conditions, sample size was comparable to Van Gog et al. (2014), who reported a medium-sized effect ($r = 0.354$) for the finding that participants who saw the model's face showed higher learning outcomes after having studied the example twice than participants who did not see the model's face. We had enough power (0.80) to detect a similar effect size with our sample with a one-tailed test and $\alpha = 0.05$.

In sum, our findings contribute to increasing evidence that seeing the instructor in instructional videos does not hamper learning (Kizilcec et al., 2014; Ouwehand et al., 2015; see also; Mayer, 2014), but we were unable to replicate prior research

(Van Gog et al., 2014) showing beneficial effects of seeing the instructor's face when learning from demonstration video examples. The fine-grained analysis of the eye movement data yielded relevant insights into how participants allocated their visual attention: when the instructor's gaze provided guidance regarding what he was attending to, participants paid more attention to the instructor's face when he was speaking and adaptively switched their attention between the instructor's face and the task area in the episodes in which the instructor was speaking and acting. This underlines the suggestion that the instructor's gaze may be a powerful cue for students that may help them to switch their attention timely from the instructor to the task (see also Ouwehand et al., 2015). Moreover, these findings suggest that seeing the instructor's face in video examples, may perhaps have a beneficial effect on learning when the cues provided by the instructor's gaze are necessary to ensure that students understand what s/he is referring to or to ensure that they attend timely to what s/he is doing. Future research should address this question, by systematically manipulating gaze guidance across different types of tasks.

## Acknowledgments

This research was supported by a Vidi grant (# 452-11-006) from the Netherlands Organization for Scientific Research (NWO) awarded to Tamara van Gog. The authors would like to thank Chenella Dewnarain, Susan Ravensbergen, Kirsten Versijde and Marcel Vielvoije for their help with creating the materials and running the experiment. We also thank Marcel Vielvoije for acting as model in the video examples and for his permission to include a recognizable photo in the manuscript.

## References

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.011.

Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology, 29*, 194–213. http://dx.doi.org/10.1177/0261927X09359589.

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121. http://dx.doi.org/10.1016/j.compedu.2014.08.015.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction, 28*, 35–47. http://dx.doi.org/10.1016/j.learninstr.2013.05.001.

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. http://dx.doi.org/10.1075/pc.14.1.05gul.

Gunawardena, C. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications, 1*, 147–166.

Homer, B., Plass, J., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*, 786–797. http://dx.doi.org/10.1016/j.chb.2007.02.009.

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction, 44*, 22–30. http://dx.doi.org/10.1016/j.learninstruc.2016.02.004.

Kay, R., & Kletskin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education, 59*, 619–627. http://dx.doi.org/10.1016/j.compedu.2012.03.007.

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, 107*, 724–739. http://dx.doi.org/10.1037/edu0000013.

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 annual SIGCHI conference on human factors in computing systems (pp. 2095–2102)*. New York: Association for Computing Machinery. http://dx.doi.org/10.1145/2556288.2557207.

Korving, H., Hernández, M., & De Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education, 94*, 151–161. http://dx.doi.org/10.1016/j.compedu.2015.11.011.

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*, 121–132. http://dx.doi.org/10.1016/j.learninstruc.2011.08.004.

Kuhn, G., & Martinez, M. (2012). Misdirection – past, present, and the future. *Frontiers in Human Neuroscience, 5*(172), 1–7. http://dx.doi.org/10.3389/fnhum.2011.00172.

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. http://dx.doi.org/10.1016/S1364-6613(99)01436-9.

Levy, J., Foulsham, T., & Kingstone, A. (2012). Monsters are people too. *Biology Letters, 9*, 1–4. http://dx.doi.org/10.1098/rsbl.2012.0850.

Louwerse, M. M., Graesser, A. C., McNamara, D. S., & Lu, S. (2009). Embodied conversational agents as conversational partners. *Applied Cognitive Psychology, 23*, 1244–1255. http://dx.doi.org/10.1002/acp.1527.

Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369, 2nd, rev. ed.

McLaren, B. M., Van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior, 55*, 87–99. http://dx.doi.org/10.1016/j.chb.2015.08.038.

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

Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. http://dx.doi.org/10.1111/cogs.12086.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155–174. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.compedu.2013.12.004.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89. http://dx.doi.org/10.1016/j.chb.2017.01.049.