Received: 11 September 2019 | Revised: 19 December 2019 | Accepted: 20 January 2020

DOI: 10.1111/jcal.12428

ARTICLE

Journal of Computer Assisted Learning WILEY

Effects of prior knowledge and joint attention on learning from eye movement modelling examples

Lucia B. Chisari | Akvilė Mockevičiūtė | Sterre K. Ruitenburg | Lian van Vemde | Ellen M. Kok | Tamara van Gog

Department of Education, Utrecht University, The Netherlands

Correspondence
Ellen Kok, Department of Education, Utrecht University, PO Box 80140, 3508 CS, Utrecht, The Netherlands.
Email: e.m.kok@uu.nl

Funding information
NRO PROO, Grant/Award Number: 405-17-301

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12428.

Abstract

Eye movement modelling examples (EMMEs) are instructional videos of a model's demonstration and explanation of a task that also show where the model is looking. EMMEs are expected to synchronize students' visual attention with the model's, leading to better learning than regular video modelling examples (MEs). However, synchronization is seldom directly tested. Moreover, recent research suggests that EMMEs might be more effective than ME for low prior knowledge learners. We therefore used a 2 × 2 between-subjects design to investigate if the effectiveness of EMMEs (EMMEs/ME) is moderated by prior knowledge (high/low, manipulated by pretraining), applying eye tracking to assess synchronization. Contrary to expectations, EMMEs did not lead to higher learning outcomes than ME, and no interaction with prior knowledge was found. Structural equation modelling shows the mechanism through which EMMEs affect learning: Seeing the model's eye movements helped learners to look faster at referenced information, which was associated with higher learning outcomes.

KEYWORDS

eye movement modelling examples, instructional videos, joint attention, observational learning, prior knowledge

1 | INTRODUCTION

The present study addresses the effects of prior knowledge (PK) and joint attention on learning from eye movement modelling examples (EMMEs). EMMEs are digital instructional videos that can be used in online or blended learning environments, which not only show a model's (e.g., a task expert's or teacher's) demonstration and explanation of a task by means of a screen-recording (with or without voice-over) but also show where the model was looking by means of a gaze cursor (e.g., a cross, circle, or dot) overlaid on the screen-recording. EMMEs were designed based on the assumption that following the model's gaze would synchronize students' visual attention with the model's, establishing joint attention, which would lead to better learning outcomes than regular screen-recording video examples that do not show the model's eye movements (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). However, whether this joint attention indeed occurs is seldom directly tested, and thus far, studies have not addressed the relationship between joint attention and learning outcomes directly. Moreover, as we will discuss below, evidence for the effectiveness of EMMEs compared with regular video examples is

The first four authors contributed equally to this manuscript.

mixed. Recent research suggests PK might influence effectiveness, with EMMEs being more effective for low PK learners (i.e., if the learner does not know the name of symbols used in the task), but effects of PK have not yet been tested within a single study.

1.1 | Learning from video examples

Observing how another person performs a task is a very natural way of learning (Bandura, 1977; Gergely, 2013), and the effectiveness of example-based learning has been shown in a wide variety of domains (for reviews: Van Gog & Rummel, 2010; Van Gog, Rummel, & Renkl, 2019). With the advent of modern technology, it has become very easy to create and share *video* modelling examples (MEs; i.e., "how to" videos), in which a "model" (e.g., a task expert, teacher, or more knowledgeable peer) demonstrates and explains how to perform a task. Thus, it is not surprising that video MEs are a common ingredient in contemporary online or blended learning environments. However, it is known that example-based learning is only effective when the examples are well designed (as shown early on by Tarmizi & Sweller, 1988), and a specific challenge in video MEs lies in the fact that information is often transient (Ayres & Paas, 2007). To learn effectively from multimedia materials (like video MEs), learners need to select (i.e., attend to) the corresponding elements in the verbal and pictorial information, organize these into a coherent mental representation, and integrate this representation with PK (Mayer, 2014). Because the model's verbal explanation is transient (and the pictorial information might be too), learners need to attend to the pictorial information that the model is referring to at the right time. If the timely selection of the right information is hampered, information is not available for organizing and integrating, and learners will have difficulties understanding the model's demonstration and explanation. In other words, one major issue when it comes to video example design is how to ensure that students pay attention to the same information that the model is referring to, at the same time; that is, how to establish *joint attention*.

1.2 | Joint attention in video examples and the role of PK

Joint attention is fundamental to social communication and learning (Butterworth, 1995). It underlies early (language) learning (Bloom, 2002; Meltzoff & Brooks, 2013; Scaife & Bruner, 1975) and leads to better understanding of what another person is saying (e.g., Richardson & Dale, 2005). In face-to-face modelling situations, the model and learner can regularly monitor (by visually checking-back) whether joint attention to an object is established and maintained; however, this is impossible when studying video examples in asynchronous computer-based learning environments. This can pose a problem for learning from video examples, because in complex and information-rich tasks (e.g., learning to troubleshoot electronic circuits), joint attention is not a given.

That is, even though the model will verbally refer to elements of the pictorial information in the explanation, this may not suffice to quickly guide students' attention to that information. The verbal referents used by the model may be ambiguous for students if they cannot see what the model is looking at, especially in visually complex tasks. For instance, suppose the model talks about a "variable resistor" in an electronic circuit. If students do not yet know the symbol for variable resistor and/or the circuit contains multiple variable resistors, then the model's verbal explanation may not suffice to guide students' attention to the right location at the right time (i.e., to establish joint attention), and—because the video is moving on—students' understanding may be hampered.

In video examples in which the model is visible (e.g., standing next to the slides: Van Wermeskerken, Ravensbergen, & van Gog, 2018; manipulating objects on a table: Van Gog, Verveer, & Verveer, 2014), joint attention can be established by means of gaze or gesture cues (i.e., the model looking or pointing at the information she or he is referring to; e.g., Mayer & DaPra, 2012; Ouwehand, van Gog, & Paas, 2015). In screen-recording examples (in which the model's computer screen is recorded on which she or he demonstrates the task; e.g., Van Gog et al., 2009), this is more difficult, but eye-tracking technology provides a unique tool to establish joint attention: The model's eye fixations (i.e., centre of visual attention) can be recorded during the demonstration and displayed in the video ME by means of a gaze cursor (e.g., a cross, circle, or dot). These EMMEs were designed based on the assumption that following the model's gaze would synchronize students' visual attention with the model's, establishing joint attention, which would lead to better learning outcomes than regular screen-recording video examples that do not show the model's eye movements (Van Gog et al., 2009).

Yet the assumption that EMMEs establish joint attention is seldom directly tested. The studies that did address how EMMEs affect students' attention during example study indeed found evidence of beneficial effects of EMMEs on joint attention: EMMEs were shown to help learners to adopt the model's viewing pattern (i.e., smaller distance between model's and learners' viewing location; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013) and look at the pictorial information that was verbally referenced by the model faster (i.e., Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016, Experiment 2, 2018, Experiment 1) than regular video examples. However, the relation between the effects of EMMEs on students' viewing behaviour and their learning outcomes was not tested directly.

Moreover, evidence regarding the effectiveness of EMMEs for learning is mixed; some studies did find beneficial effects of EMMEs compared with regular video MEs (Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017; Jarodzka et al., 2012; Jarodzka et al., 2013),[1] whereas others did not (Van Gog et al., 2009; Van Marlen, et al., 2016, 2018, Experiment 1). Recent findings suggest that the explanation for the mixed results might lie in the PK of the learners, as EMMEs were found to be more effective than regular video examples for low PK learners (Van Marlen et al., 2018, Experiment 2; this is also in line with findings from a meta-analysis on the effectiveness of visual cueing in multimedia learning: Richter, Scheiter, & Eitel, 2016). One particular aspect of PK is likely to moderate the effectiveness of EMMEs: When learners do not know the verbal labels for visual information (name–symbol correspondence), they might benefit from the guidance of the EMMEs to know what information to attend to. For example, when learners have more PK (e.g., know what a variable transistor symbol looks like), the verbal explanation might be sufficient to guide their attention to the right pictorial information at the right time, and they do not need the visual attention guidance provided by the EMMEs to learn from the example. However, although the combined findings of two studies by Van Marlen et al. (2016, Experiment 2, and 2018, Experiment 2) suggest that PK is an important factor in the effectiveness of EMMEs, the specific effect of PK of name–symbol correspondences has not been directly tested.

1.3 | The present study

Therefore, in the present study, we used a 2 × 2 design to investigate the effectiveness of an EMME (vs. a regular video ME) about troubleshooting electronic circuits for high and low PK learners. Here, high PK refers to participants' PK of the correspondence between the names and symbols of electronic components, which was manipulated through pretraining. We also recorded students' eye movements during example study to investigate the effects of EMMEs on joint attention (i.e., the distance from the model's viewing location and the speed of looking at referenced information) and the relation with learning outcomes.

We hypothesized that EMMEs would have a beneficial effect on joint attention during example study, with the distance from the model's viewing location being smaller and the speed of looking at referenced information being higher in the EMMEs than in the ME condition (Hypothesis 1: main effect of example type on eye movement data). We further expected that high PK learners would learn more from the videos (i.e., perform better on the posttest) than low PK learners (Hypothesis 2: main effect of PK on learning outcome data) and that EMMEs would be more effective for learning (as assessed by posttest performance) than ME for low PK learners but not for high PK learners (Hypothesis 3: interaction effect of example type and PK on learning outcome data). Using structural equation modelling, we also investigated the hypothesis that the relation between example type and learning would be mediated by joint attention (Hypothesis 4).

2 | METHOD

2.1 | Participants

Participants were Dutch university students who were fluent in English as a first or second language and had normal or corrected to normal vision. Of the 71 participants who originally enrolled in the study, 12 had to be excluded because (a) they had taken physics classes in the last 3 years of secondary education or in higher education (n = 1), (b) they had a score of five or higher on the pretest (n = 1), (c) they did not pass the PK test (n = 1), (d) calibration of the eye tracking system could not be successfully completed after several attempts (n = 7), or (e) they had a co-occurrence of two of these issues (n = 2). The final sample therefore consisted of 59 participants for the analyses of the eye movement data (age M = 23.46 years, SD = 3.04; 47 females). Because the performance data were not affected by calibration of the eye tracking system, participants with unsuccessful calibration could be included in the analyses of these data, yielding a final sample of 66 participants (age M = 23.35 years, SD = 2.98; 53 females). The study was approved by the research ethics committee of the institute where this study was conducted; all participants provided written informed consent. The data that support the findings of this study are openly available in Dataverse at https://hdl.handle.net/10411/HY8IBJ, (Kok, 2020).

2.2 | Design

The experiment had a 2 × 2 design with between-subjects factors example type (EMMEs vs. ME) and PK (Low vs. High). Participants were randomly assigned to one of four conditions: (a) *Low PK + ME* (n = 14), (b) *Low PK + EMME* (n = 16), (c) *High PK + ME* (n = 15), and (d) *High PK + EMME* (n = 14).

2.3 | Apparatus and materials

2.3.1 | Eye-tracking equipment

The experiment was developed using SMI Experiment Center (Version 3.7; SensoMotoric Instruments GmbH, 2017) and presented on a 22-in. monitor (1680 × 1050 pixels). Eye movements were recorded using a SMI RED250 eye tracker with a sampling rate of 250 Hz (SensoMotoric Instruments GmbH, 2017). The screen subtended 44° of visual angle horizontally and 28° of visual angle vertically at a viewing distance of 59 cm. A velocity-based, high-speed event-detection algorithm was used with a velocity threshold of 40°/s. Speakers were used as an audio outlet, and a chin rest was utilized to stabilize participants' head position while their eye movements were being recorded.

2.3.2 | Pretest

A pretest, filled out on paper, was administered to check whether participants did not have an excessive amount of PK on symbols of electronic circuits. The pretest consisted of an open question, asking participants to list all the electronic components and their corresponding symbols they knew (time was not limited). The final score was the total number of correctly listed name–symbol correspondences. Participants were excluded from the study if they listed five or more components correctly, indicating their names in English or Dutch and their corresponding symbols according to the convention used in Hewes (2018).

2.3.3 | PK training

The PK training (PKT) consisted of teaching participants in the *high PK* conditions the names and functions of 10 basic components of electronic circuits and their corresponding symbols, and the definitions of series and parallel circuits. A static image depicting these components with an audio explanation that lasted approximately 4.5 min was used. Participants assigned to the *low PK* conditions heard the same audio explanation on the names and functions of the 10 components and the definitions of series and parallel circuits but were shown pictures of the physical images of those components instead of symbols. This subtle manipulation was necessary to ensure we investigated the effect of an important aspect of PK, namely, knowing the name–symbol correspondences. The static image used in the PKT for the high PK conditions is displayed in Figure 1.

Afterwards, participants were tested on their acquired PK using a multiple-choice test with 30 questions (each component was presented three times), which was created digitally in Qualtrics survey software (Version 2018; Smith, 2002). For each component name, participants in the *high PK* conditions were required to select the correct symbol and participants in the *low PK* conditions were required to select the correct picture from a list of all 10 symbols or pictures. After each question, feedback indicating the correct answer was displayed: the name–symbol correspondence in high PK conditions and the name–picture correspondence in low PK conditions. The goal of the questions and feedback was twofold. On the one hand, it served as part of the training (cf. retrieval practice, Roediger & Butler, 2011; Roediger and Karpicke, 2006). At the same time, it served as a manipulation check.

2.3.4 | Video ME on the working of electronic circuits

The video ME in the ME condition showed a static image of an electronic circuit that included all the components presented in the PKT (see Figure 2), with a voice-over in which a female model explained the steps needed to troubleshoot the circuit (i.e., to diagnose and repair its faults). In order for the circuit to work, four faults needed to be corrected: (a) the voltmeter had to be turned around; (b) the ammeter had to be connected in series with respect to its corresponding source of light; (c) the diode had to be turned around; and (d) the LED had to be connected in series with the push-button switch or a resistor had to be placed in the path of the switch to assure that the LED was no longer short-circuited by the push-button switch. In the verbal explanation, the model did not give any clues regarding name–symbol correspondences. The video lasted 3 min 41 s.

In the EMME conditions, the video ME additionally showed where the model was looking (i.e., the model's fixations) as a translucent blue dot superimposed on the electronic circuit (see Figure 2). The model's eye movements and voice-over were recorded simultaneously with the eye tracking equipment described in Section 2.3.1, and the EMME was created with SMI BeGaze software (Version 3.7; SensoMotoric Instruments GmbH, 2017). The model was instructed to behave didactically and align the eye movements with the verbal explanation and could re-record the video as many times as necessary (cf. procedure by Jarodzka et al., 2013).

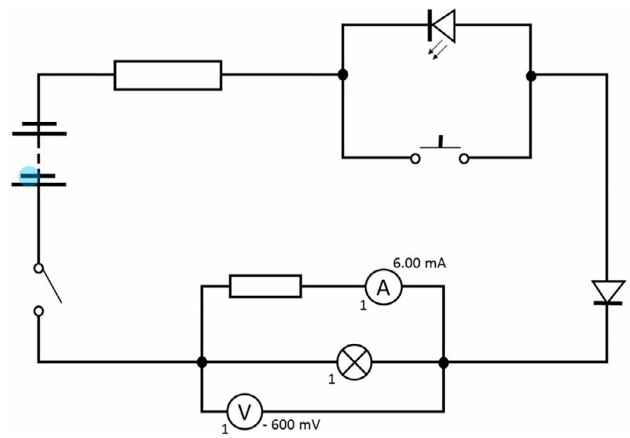

FIGURE 2 Screenshot of the eye movement modelling examples (EMMEs) with the translucent blue dot (here located on the battery on the left) indicating where the model was looking at any given moment [Colour figure can be viewed at wileyonlinelibrary.com]

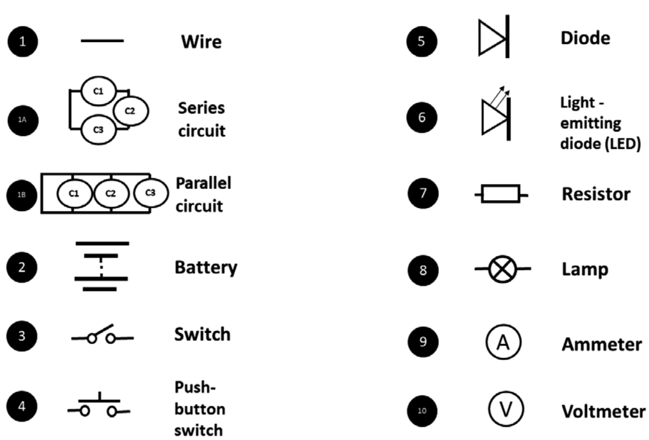

FIGURE 1 The static image used in the prior knowledge training for the high prior knowledge (PK) conditions

2.3.5 | Knowledge posttest

The knowledge posttest, administered via Qualtrics survey software (Version 2018; Smith, 2002), consisted of 15 items. The first five items were focused on only diagnosing the faults made in the circuit and had all the same nine answer options (provided alphabetically; see Figure A1 for example item). The last 10 items were focused on also repairing the faults made and all had the same 10 answer options (again provided alphabetically; see Figure A2 for example item). Multiple answers could be selected as correct. There was no time limit for the test, but participants were encouraged to answer as fast as possible.

All circuits presented in the knowledge posttest were variations of the circuit shown in the video ME containing a selection of or all components but had a different surface structure (i.e., a different layout). The test was developed in several steps. First, the types of connections were identified (i.e., circuits could be connected in series, in parallel, or in series *and* in parallel). Second, all faults that could possibly be made with the used components were divided into categories (viz., position of the meter with respect to the corresponding source of light; direction of component with regard to the battery; and short-circuited). All circuits contained two or three of these faults. To make some items harder, up to two distractors per circuit were included. This led to 18 possible combinations of circuits, which were listed. Fifteen of these were randomly selected and developed into test items. Five of these 15 combinations were then randomly selected as diagnosing items and the other 10 as repairing items.

Each item in the test was scored separately, granting participants 1 point if they performed the correct action pertaining to an answer option and no points if they performed the incorrect action (e.g., if an answer option to an item was supposed to be selected and the participant did, she or he was given 1 point; if an answer option to an item was not supposed to be selected, but the participant did, she or he was given no points). Therefore, the maximum score per question was the number of answer options it contained. The total score was the sum of the scores for all 15 questions, resulting in a maximum score of 145 points. Participants' total score was used as learning outcome measure. Cronbach's alpha for the five diagnosing questions was .70, and for the 10 repairing questions 0.88, and the whole knowledge posttest had an excellent reliability of .90 (no items could be removed to improve the reliability).

2.4 | Procedure

First, participants completed the pretest and a short questionnaire asking for demographics, both on paper. They were then seated in front of the monitor and completed the PKT according to their assigned experimental condition. After the PKT, all participants were tested on their level of PK with a multiple-choice test. To make sure that participants in the *high PK* conditions had gained the required PK, they were only allowed to proceed to the video ME after (a) having an 80% overall correct performance on one round of the test and (b) making no mistakes in the last 10 items. If this was not the case after the first round (n = 5), participants were automatically directed into the second (n = 5) and third (n = 3) rounds without watching the video again until they met the two requirements (feedback was provided in the first round only). We considered 80% to be the lower boundary for a score to be called high, which would mean that participants would know the names of at least eight out of the 10 components in this round. With the additional requirement of making no mistakes in the last 10 items, this should ensure that participants knew all name–symbol correspondences. Then, the eye tracking equipment was calibrated using a 9-point calibration with a 4-point validation procedure, which was only considered to be successful if the error on the x-axis and y-axis was not larger than 1.0° of the visual angle. Next, participants were instructed that they would study a video example (without a reference to whom the model was) and subsequently studied the video example (ME or EMMEs depending on their assigned condition) while their eye movements were recorded. Finally, participants completed the knowledge posttest. The total experiment had an average duration of 40 min.

2.5 | Data analysis

2.5.1 | Joint attention

Two measures of joint attention were computed. First, the alignment between the model's and each participant's eye gaze was measured using Euclidean distances (i.e., the distances in pixels between the model's gaze location and the learners' gaze location; see also Jarodzka et al., 2012). Average Euclidean distances were calculated for each participant using MATLAB (Version R2017a; Moler, 1984). This measure corresponds to the average distance between the expert's gaze location and the participant's gaze location throughout the video, in both EMME and ME conditions, and was calculated based on raw gaze data. The average Euclidean distance does not account for the fact that participants need time to plan eye movements that follow the EMMEs or (in the case of ME conditions) the verbal explanation. The time needed to follow the dot of the EMMEs is likely to be less than 0.5 s (the time needed to program saccades that follow the dot), whereas the time needed to follow the verbal explanation for ME conditions is likely to be around 2–3 s (the time needed to interpret verbal data and program saccades; cf. Richardson & Dale, 2005). Thus, the average Euclidean distance is likely to favour the EMME conditions. Therefore, we calculated each individual's Euclidean distance for a range of 46,200-ms interval lags, starting at −3,000 and ending at +6,000 ms (see Richardson & Dale, 2005, for similar lags). We then calculated each individual's optimal lag (i.e., that lag with the lowest Euclidean distance between expert's gaze location and the participant's gaze location), and we used the Euclidean distance for each individuals' optimal lag (minimal distance) as a measure of joint attention. The measure "minimal distance" is thus a conservative measure of joint attention.

Second, the *entry time* (measured in ms) was calculated as the mean time a participant needed to fixate on a referent after it was mentioned in the verbal explanation. A lower entry time (i.e., the participant fixates the referent soon after it was mentioned) reflects higher joint attention. For each circuit component that was verbally referred to by the model, an area of interest (AOI) was created using BeGaze 3.7 software (SensoMotoric Instruments, GmbH, 2017). Those AOIs only became "active" when the component name was used in the model's explanation and finished when the next component was being mentioned. The entry time was calculated from the moment an AOI would become active. In total, the video example contained 30 active AOIs (identical for EMME and ME videos). The average active time of AOIs was 7,553.47 ms (range 1,467–33,007 ms). The average size of AOIs was 37,156.9 px (range 15,930–51,986 px). The active time of an AOI was used as entry time if the AOI was not looked at by a participant. Again, this was done to make this a conservative measure of entry time: If, in some conditions most AOIs would never be looked at, but a few AOIs would be looked at quickly, the average entry time of only those AOIs would be an overestimation of joint attention.

2.5.2 | Analyses of variance

Hypotheses 1–3 were tested with 2 × 2 analyses of variance (ANOVAs) with example type (EMMEs vs. ME) and PK (Low vs. High) as between-subjects factors. The ANOVAs were based on $N = 59$ for the eye movement data and $N = 66$ for the performance data (see Section 2.1). Partial eta-squared ($\eta_p{}^2$) is reported as a measure of effect size with values of .01, .06, and .14 corresponding to small, medium, and large effect sizes, respectively (Cohen, 1988). Bonferroni correction was applied to control familywise error due to multiple testing (see Field, 2013, pp. 68–69), resulting in a significance level of $\alpha = .017$.

Before analysis, the assumptions of (a) normality (assessed with histograms, skewness and kurtosis values, and the Shapiro–Wilk test), (b) homogeneity of variance, and (c) absence of outliers (participants were considered an outlier with $z \geq |2.5|$) were checked. For entry time, one univariate outlier was identified ($z = 2.81$) and therefore removed from the ANOVA of this variable.

2.5.3 | Structural equation modelling

Hypothesis 4 was tested using structural equation modelling in Mplus 8 software (Muthén & Muthén, 2018). Structural equation modelling was used because it allows for testing both mediators in one model, incorporates measurement error in the model, and offers greater power than hierarchical regression to detect effects (Cheung & Lau, 2008; Sardesmukh & Vandenberg, 2017). The manifest path model displayed in Figure 3 was specified. In order to test the hypothesis that EMMEs would foster joint attention, which, in turn, would have a beneficial effect on learning outcomes, it was tested whether (a) $\beta_{31} * \beta_{23} > 0$, (b) $\beta_{41} * \beta_{24} > 0$, and (c) $\beta_{21} = 0$ (indicating that joint attention perfectly mediates the effect of example type on learning). The model was based on $N = 59$ for both eye movement data and performance data.

The total effect and all indirect effects were evaluated using bootstrapped confidence intervals (using 1,000 bootstraps); all direct effects were evaluated at $\alpha = .05$. For the path model, model fit was evaluated using the following five model fit indices: (a) chi square test of model fit ($p > .05$ indicates good fit), (b) root mean square error of approximation (RMSEA; <0.08 indicates adequate fit; <0.05 good fit), (c) comparative fit index (CFI; >0.90 indicates adequate fit; >0.95 good fit), (d) Tucker–Lewis index (TLI; >0.90 indicates adequate fit; >0.95 good fit), and (e) standardized root mean square residual (SRMR; <0.08 indicates good fit; Hu & Bentler, 1999). Besides that, R^2 effect size was reported for each endogenous (i.e., being predicted) variable in the model. For R^2, values of .25, .50, and .75 correspond to weak, moderate, and substantial effect sizes, respectively (Hair, Ringle, & Sarstedt, 2011).

Before analysis, the assumptions for structural equation models were checked. For all variables used in the path model (i.e., knowledge posttest score, minimal distance, and entry time), multivariate normality could be assumed according to Q–Q plots. According to the Mahalanobis distance test for the eye-tracking measures, defined as $\chi^2(2) = 9.21$, $p < .001$, one multivariate outlier was present ($\chi^2 = 17.02$), but Cook's values were close to zero (maximum value of 0.12), indicating that the outlier did not exert large influence. According to the Mahalanobis distance test for the knowledge posttest scores, defined as $\chi^2(1) = 6.64$, $p < .001$, no multivariate outliers were present (maximum values were below 6.64).

3 | RESULTS

The means and standard deviations for pretest score, knowledge posttest score, minimal distance, and AOI entry time per condition are provided in Table 1. As can be seen in the first column, scores on the pretest

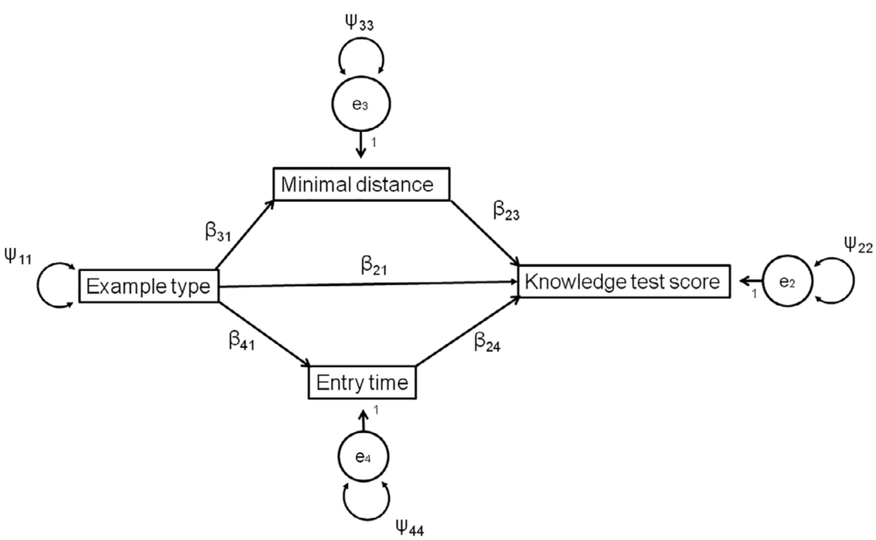

FIGURE 3 Structural equation model for testing Hypothesis 4. e indicates the error term, ψ indicates the residual variances, and β indicates parameter estimates

TABLE 1 Means and standard deviations for pretest score, knowledge posttest score, minimal distance, and entry time for each condition separately

	Pretest score			Knowledge posttest score			Minimal distance			Entry time		
Condition	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Low PK + ME	17	0.12	0.33	17	105.76	13.30	14	413.58	77.70	14	3,179.80	819.68
Low PK + EMME	17	0.06	0.24	17	107.29	13.22	16	187.63	55.21	15[a]	2,255.91	743.28
High PK + ME	17	0.35	0.70	17	114.00	12.17	15	356.90	78.75	15	2,701.81	826.87
High PK + EMME	15	0.07	0.26	15	114.80	10.47	14	234.46	96.16	14	2,048.50	669.32

Abbreviations: EMME, eye movement modelling example; ME, modelling example; PK, prior knowledge.
[a]*n* = 15 instead of 16, because one participant was removed for being an outlier on this variable.

were low for all conditions: On average, participants knew less than one name–symbol correspondence before the start of the experiment.

3.1 | Hypothesis 1: Joint attention

3.1.1 | Minimal distance

In line with our first hypothesis, the 2 × 2 ANOVA on minimal distance showed a large, significant main effect of example type, $F(1, 55) = 74.08$, $p < .001$, $\eta_p^2 = 0.574$, with participants in the EMME conditions ($M = 209.49$, $SD = 79.29$) fixating more closely to the model's viewing location than participants in the ME conditions ($M = 384.26$, $SD = 82.06$). There was no significant main effect of PK, $F(1, 55) = 0.059$, $p = .808$, $\eta_p^2 < 0.001$, but there was a significant interaction between example type and PK, with a medium effect size, $F(1, 55) = 6.54$, $p = .013$, $\eta_p^2 = 0.106$. This interaction indicates that the effect of example type on minimal distance differed as a function of PK: EMMEs decreased minimal distance in both the high PK and low PK conditions, but this decrease was much larger for participants in the low PK condition (cf. Table 1 and Figure 4).

3.1.2 | AOI entry time

In line with our first hypothesis, there was a large, significant main effect of example type, $F(1, 54) = 15.26$, $p < .001$, $\eta_p^2 = 0.220$, with participants in the EMME conditions ($M = 2{,}155.78$, $SD = 703.82$) having a faster entry time (i.e., displaying *more* joint attention) than participants in the ME conditions ($M = 2{,}932.56$, $SD = 844.32$). There was no significant main effect of PK, $F(1, 54) = 2.88$, $p = .095$, $\eta_p^2 = 0.051$, nor a significant interaction between example type and PK, $F(1, 54) = 0.45$, $p = .506$, $\eta_p^2 = 0.008$.

3.2 | Hypotheses 2 and 3: Posttest performance

The 2 × 2 ANOVA on knowledge posttest performance showed no significant main effect of example type, $F(1, 62) = 0.145$, $p = .704$,

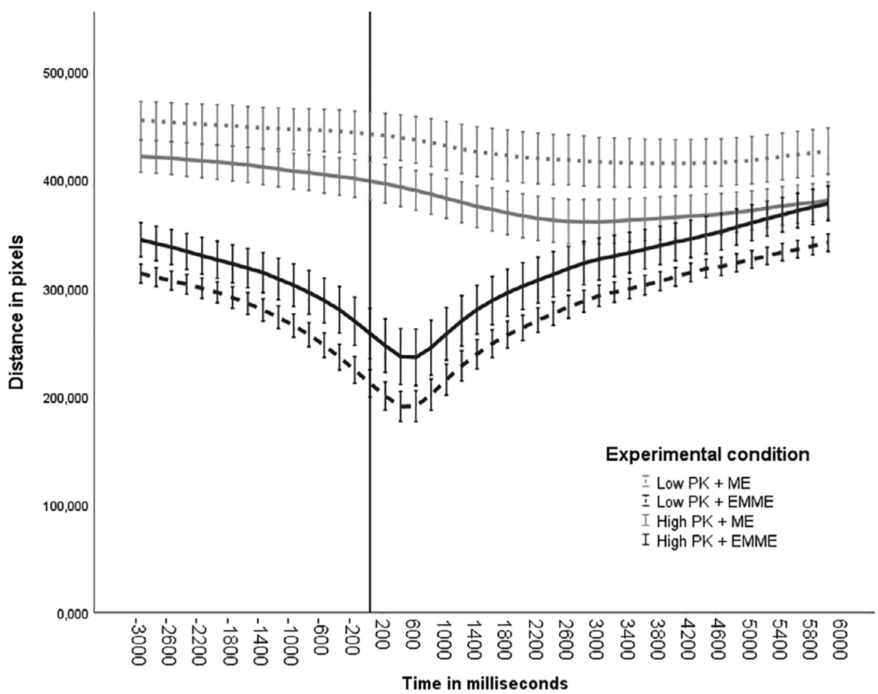

FIGURE 4 Mean Euclidean distance over the course of the instruction video computed for 45 lags with an interval of 200 ms, starting at −3,000 and ending at +6,000 ms, for each condition separately. Error bars display a ±1 standard error. The vertical line indicates a lag of 0 ms between the model's and participants' gaze locations. EMME, eye movement modelling example; ME, modelling example; PK, prior knowledge

η_p^2 = 0.002. In line with our second hypothesis, there was a significant main effect of PK, with a medium effect size, $F(1, 62) = 6.629$, $p = .012$, $\eta_p^2 = 0.097$, with participants in the high PK conditions ($M = 114.38$, $SD = 11.23$) outperforming participants in the low PK conditions ($M = 106.53$, $SD = 13.08$). In contrast with our third hypothesis that EMMEs would be more effective for learning than ME for low PK students but not for high PK students, there was no significant interaction between example type and PK, $F(1, 62) = 0.014$, $p = .905$, $\eta_p^2 < 0.000$.

3.3 | Hypothesis 4: Relation between joint attention and learning

Structural equation modelling was used to test whether the distance from the model's viewing location and the entry time mediated the effect of example type on the knowledge posttest score (see Figure 3). The specified model fitted the data quite well, $\chi^2(1) = 1.08$, $p = .300$; RMSEA = 0.04; CFI = 1.00; TLI = 0.99; SRMR = 0.03. The indirect effect of example type on learning through *minimal distance* was not significantly different from zero, $\beta_{31} * \beta_{23} = 3.68$ (standardized $\beta_{31} * \beta_{23} = 0.15$), 95% CI [−6.71, 14.07]. However, the indirect effect of example type on learning through *entry time* was significantly different from zero, $\beta_{41} * \beta_{24} = 4.40$ (standardized $\beta_{41} * \beta_{24} = 0.18$), 95% CI [0.78, 8.02]. There was no evidence that the EMMEs fostered learning directly in addition to the indirect effect of the EMMEs on learning through distance and entry time, $\beta_{21} = -7.16$ (standardized $\beta_{21} = -0.29$), $z = -1.13$, $p = .257$. This indicates that entry time (but not distance to the model's viewing location) perfectly mediated the effect of example type on learning. The model explained 21%, 55%, and 14% of the variability in knowledge posttest performance, minimal distance, and AOI entry time, respectively. These effect sizes can be considered weak for knowledge test score and entry time, and moderate for minimal distance.

In order to explore whether PK moderated the above-mentioned model, again, the manifest path model displayed in Figure 3 was specified, but now for low PK and high PK learners separately. Stepwise constraining direct effects to be equal across PK conditions showed that only β_{31} was not the same for low PK learners and high PK learners. This implies that the level of domain-specific PK moderated the effect of example type on minimal distance. More specifically, for low PK learners, $\beta_{31} = -234.62$ (standardized $\beta_{31} = -0.87$), $z = -9.56$, $p < 0.001$, whereas for high PK learners, $\beta_{31} = -122.44$ (standardized $\beta_{31} = -0.59$), $z = -3.71$, $p < 0.001$, indicating that the effect of the condition on minimal distance is negative in both groups, but the effect is much stronger (i.e., more negative) for low PK learners. The specified model fitted the data quite well, $\chi^2(6) = 9.51$, $p = .147$; RMSEA = 0.14; CFI = 0.96; TLI = 0.91; SRMR = 0.21. For low PK learners, the model explained 27%, 75%, and 13% of the variability in, respectively, knowledge test score, minimal distance, and entry time. These effect sizes are respectively, weak, moderate and weak. For high PK learners, these percentages were 17%, 34%, and 20%, respectively indicating weak, moderate, and weak effect sizes.

4 | DISCUSSION

We investigated the effectiveness of EMMEs as a function of learners' PK, manipulating PK via pretraining. We also applied eye tracking during example study to directly test the assumption that by showing students where the model is looking, EMMEs improve joint attention, and that it is this joint attention that improves learning.

We found support for our first hypothesis that studying an EMME would have a beneficial effect on joint attention during example study compared with studying a regular video ME that does not show where the model is looking. Students' gaze location was closer to the model's gaze location during example study, and students looked faster at the pictorial information that the model verbally referred to when studying EMMEs. These results are in line with findings from the few prior studies that investigated students' attention during example study (Jarodzka et al., 2013; Van Marlen et al., 2016, Experiment 2, 2018, Experiment 1) and show that EMMEs indeed improve joint attention between the model and the learner. But does joint attention lead to better learning?

As expected (Hypothesis 2), students with high PK showed better learning outcomes after video example study than students with low PK, which shows that our pretraining manipulation of PK was successful. Yet in contrast to our third hypothesis, we did not find the expected interaction effect indicating that EMMEs would be more effective than regular examples for low (but not for high) PK learners (nor was there a main effect of EMMEs on learning). Interestingly, however, structural equation modelling showed that in line with our fourth hypothesis, the relation between example type and learning outcomes was mediated by one of the joint attention measures (AOI entry time). Furthermore, this mediation was moderated by PK: The attention of participants in the low PK condition was guided more by the EMMEs than the attention of high PK condition. Thus, although we did not find a direct effect of EMMEs on learning outcomes, our findings do provide evidence for an indirect effect, indicating that a higher degree of joint attention (and in particular, looking faster at the information that the model is referring to) was associated with better learning outcomes (i.e., "indirect-only mediation"; Zhao et al., 2010).

4.1 | Limitations and future research

This study has some limitations that may provide potential explanations for why we failed to find an interaction effect between example type and PK. First, we manipulated PK through pretraining in which students in the low PK conditions did not acquire knowledge on the symbols used in the video examples, but they did receive information on the function of the components. This specific operationalization allowed us to investigate whether knowing name–symbol correspondences moderates the effectiveness of EMMEs. Thus, our findings might have been different if we had used a control condition in which students had no PK whatsoever (although in that case, if differences would have been found, it would have been unclear whether they arose from not knowing the function, the name–symbol correspondences, or both). The name–symbol correspondences in this study represent the use of task-specific vocabulary, or jargon, in general: Experts might not always be aware that they use technical language to refer to specific aspects of a task, and this might impact the effectiveness of video examples (Hinds, Patterson, & Pfeffer, 2001). Indeed, we found that only knowing the name–symbol correspondences (without the underlying extensive schemata that experts possess) already helped participants to establish joint attention with the model, in such a way that participants in the low PK group were much more reliant on the EMMEs to guide attention whereas those who knew the name–symbol correspondences could follow the verbal explanation without following the eye movements (cf. the ability-as-compensator hypothesis; Richter et al., 2016). Second, our sample size would not have been large enough to detect a small effect. Third, our hypothesis regarding PK was based on findings by Van Marlen et al. showing that EMMEs on learning to solve a geometry problem were not effective for university students (2016, Experiment 2) but were effective for secondary education students who had less PK (Van Marlen et al., 2018, Experiment 2). As we cannot rule out that there may be other relevant differences between these two groups of learners than their PK and between our (conceptual) electronic circuits task and their geometry problem solving task, it would be interesting in future research to replicate the design of the present experiment using the exact same tasks as Van Marlen et al., with secondary education students as participants.

The fact that AOI entry time but not the distance to the model's viewing location was found to mediate the effect of example type on learning outcomes suggests that the mechanism through which EMMEs may contribute to learning lies mainly in improving the selection of the right pictorial information at the right time (i.e., at the moment it is mentioned in the model's verbal explanation). If that is the case, this could in all likelihood also be established through other ways of visual cueing as well (Richter et al., 2016; Van Gog et al., 2014). Nevertheless, it is interesting to know that attention and learning can be guided by visualizing the model's eye movements, because other forms of visual cueing in videos would require postproduction (e.g., adding arrows or highlighting), which would rely on deliberate decisions by the model (e.g., pointing with a mouse cursor), who—because of his or her expertise—may not always know what the students do not know (i.e., when to provide and when to withhold attention guidance).

However, it is important to keep in mind that not all EMMEs necessarily contain verbal information; future research would have to show whether the distance to the model's viewing location may be an important mechanism for learning when EMMEs are used to convey perceptual strategies (e.g., Mason et al., 2015, 2016; Mason et al., 2017; Salmerón & Llorens, 2018; Scheiter et al., 2018). Moreover, there may be other tasks for which more continuous gaze following (i.e., minimal distance to the model's viewing location) is important (e.g., visual search tasks). Further research could explore the working mechanisms of EMMEs in different types of tasks using the measures of entry time and/or minimal distance.

Finally, it is important to note that including the online measures of joint attention was important to shine additional light on the behaviour-level outcomes: If we had not included these measures, we would have incorrectly concluded that EMMEs had no effect on learning.

5 | CONCLUSION

To conclude, to the best of our knowledge, the present study was the first to directly test and confirm the key assumption that one aspect of joint attention (i.e., entry time) is the mechanism through which EMMEs affect learning. Seeing the model's eye movements helped learners to look faster at the pictorial information that the model referred to in her verbal explanation, and looking faster at referenced information was associated with higher learning outcomes.

ACKNOWLEDGEMENTS

The authors would like to thank Tim van Marlen and Margot van Wermeskerken for their advice and assistance regarding the construction and recording of the EMMEs. In addition, the authors would like to thank Mirjam Moerbeek and Milica Miočević for their statistical advice on the structural equation modelling analysis. During the realization of this study, Ellen Kok was funded by an NRO PROO grant (Project 405-17-301).

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Dataverse at [link blinded for review].

ORCID

Ellen M. Kok https://orcid.org/0000-0001-9752-2531

ENDNOTE

[1] Note that positive effects of EMMEs in comparison with no examples or written examples were also observed for a different use of EMMEs, namely, using the model's eye movements to convey study strategies, for example, text–picture integration (Mason, Pluchino, & Tornatora, 2015, 2016; Mason, Scheiter, & Tornatora, 2017; Salmerón & Llorens, 2018; Scheiter, Schubert, & Schüler, 2018).

REFERENCES

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology*, *21*, 695–700. https://doi.org/10.1002/acp.1343

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

BeGaze (Version 3.7) [Computer software]. Teltow, Germany: SensoMotoric Instruments GmbH. 2019

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge and London: The MIT Press.

Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 29–40). New York and London: Psychology Press.

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, *11*(2), 296–325. https://doi.org/10.1177/1094428107300343

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

Experiment Center (Version 3.7) [Apparatus and software]. Teltow, Germany: SensoMotoric Instruments GmbH. 2019

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (Fourth ed.). Thousand Oaks, CA: SAGE Publications Inc.

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, *113*, 212–225. https://doi.org/10.1016/j.compedu.2017.06.001

Gergely, G. (2013). Ostensive communication and cultural learning: The natural pedagogy hypothesis. In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 139-151). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199988341.003.0008

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Hewes, J. (2018). Circuit symbols [Information on a webpage]. Retrieved from www.electronicsclub.info.

Hinds, P. J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, *86*(6), 1232–1243. https://doi.org/10.1037/0021-9010.86.6.1232

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40*, 813–827. https://doi.org/10.1007/s11251-012-9218

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, *25*, 62–70. https://doi.org/10.1016/j.learninstruc.2012.11.004

Kok, E. (2020). *Effects of prior knowledge and joint attention on learning from eye movement modelling examples*, https://hdl.handle.net/10411/-HY8IBJ, DataverseNL, V1.

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*, 172–187. https://doi.org/10.1016/j.cedpsych.2015.01.004

Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, *47*, 1083–1095. https://doi.org/10.1111/bjet.12271

Mason, L., Scheiter, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text–picture processing for multimedia comprehension. *Journal of Computer Assisted Learning*, *33*, 443–460. https://doi.org/10.1111/jcal.12191

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press.

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

Meltzoff, A. N., & Brooks, R. (2013). Gaze following and agency in human infancy. In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 125–138). New York, NY: Oxford University Press.

Moler, C. (1984). *MatLab (Version R2017a) [Computer software]*. Natick, MA: The MathWorks inc.

Muthén & Muthén. (2018). *Mplus (Version 8) [Computer software]*. Los Angeles, CA: Muthén & Muthén.

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. Retrieved from https://www.jstor.org/stable/jeductechsoci.18.4.78

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Sciences*, *29*, 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Sardeshmukh, S. R., & Vandenberg, R. J. (2017). Integrating moderation and mediation: A structural equation modeling approach. *Organizational Research Methods*, *20*(4), 721–745. https://doi.org/10.1177/1094428115621609

Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*, *0*, 1–17. https://doi.org/10.1177/0735633117751605

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266. https://doi.org/10.1038/253265a0

Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning from illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology*, *88*, 80–94. https://doi.org/10.1111/bjep.12175

Smith, R. (2002). Qualtrics [Computer software]. (2018). Retrieved from https://www.qualtrics.com/

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*(4), 424–436. https://doi.org/10.1037/0022-0663.80.4.424

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*, 785–791. https://doi.org/10.1037/edu0000065

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

Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108235631.009

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. https://doi.org/10.1016/j.compedu.2013.12.004

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, *65*, 448–459. https://doi.org/10.1016/j.chb.2016.08.041

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction*, *58*, 274–283. https://doi.org/10.1016/j.learninstruc.2018.07.005

Van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior*, *89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*, 197–206. https://doi.org/10.1086/651257

How to cite this article: Chisari LB, Mockevičiūtė A, Ruitenburg SK, van Vemde L, Kok EM, van Gog T. Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *J Comput Assist Learn*. 2020; 36:569–579. https://doi.org/10.1111/jcal.12428

A. APPENDIX

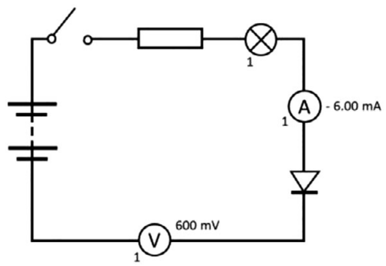

FIGURE A1 Example of a knowledge posttest item focused on only diagnosing the faults made in the circuit. In this circuit, the voltmeter and its corresponding source of light are connected in series and the ammeter is connected in the wrong direction. Hence, answer options 2 and 8 should be selected as correct

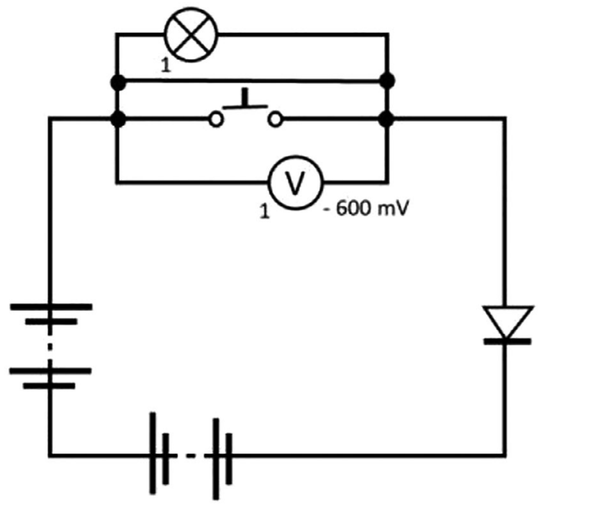

FIGURE A2 Example of a knowledge posttest item focused on also repairing the faults made in the circuit. In this circuit, (a) the voltmeter is connected in the wrong direction, (b) the lamp is short-circuited by the (push-button) switch, and (c) the lamp is short-circuited by bare wire. To repair these faults, one should (a) turn the voltmeter around, (b) connect the lamp and the (push-button) switch in series or place a resistor in the path the (push-button) switch is in, and (c) place a resistor in the path made of bare wire. Hence, answer options 2, 5, 6, and 10 should be selected as correct