Computers in Human Behavior 65 (2016) 448–459

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Showing a model's eye movements in examples does not improve learning of problem-solving tasks

Tim van Marlen [a, b, *], Margot van Wermeskerken [a, b], Halszka Jarodzka [c, d], Tamara van Gog [a, b]

[a] Institute of Psychology, Erasmus University Rotterdam, The Netherlands
[b] Department of Education, Utrecht University, The Netherlands
[c] Welten Institute, Open University of the Netherlands, The Netherlands
[d] Humanities Laboratory, Lund University, Sweden

ARTICLE INFO

Article history:
Received 3 March 2016
Received in revised form
8 June 2016
Accepted 27 August 2016
Available online 12 September 2016

Keywords:
Example-based learning
Eye tracking
Attention cueing
Multimedia learning

ABSTRACT

Eye movement modeling examples (EMME) are demonstrations of a computer-based task by a human model (e.g., a teacher), with the model's eye movements superimposed on the task to guide learners' attention. EMME have been shown to enhance learning of perceptual classification tasks; however, it is an open question whether EMME would also improve learning of procedural problem-solving tasks. We investigated this question in two experiments. In Experiment 1 (72 university students, $M_{age} = 19.94$), the effectiveness of EMME for learning simple geometry problems was addressed, in which the eye movements cued the underlying principle for calculating an angle. The only significant difference between the EMME and a no eye movement control condition was that participants in the EMME condition required less time for solving the transfer test problems. In Experiment 2 (68 university students, $M_{age} = 21.12$), we investigated the effectiveness of EMME for more complex geometry problems. Again, we found no significant effects on performance except for time spent on transfer test problems, although it was now in the opposite direction: participants who had studied EMME took longer to solve those items. These findings suggest that EMME may not be more effective than regular video examples for teaching procedural problem-solving skills.

1. Introduction

Worked examples or modeling examples, in which it is demonstrated how to perform a task, are an effective way to promote learning, especially when learners have no or limited prior knowledge (for reviews, see Renkl, 2014; Van Gog & Rummel, 2010). Indeed, *video* modeling examples have never been more prominent than they are today, thanks to technological advancements, such as digital cameras to record them, online learning environments to store and deliver them, and the availability of digital devices with internet connections (e.g., smartboards, laptops, and tablet PC's) in classrooms and at home to replay them. Video modeling examples come in many forms; for instance, showing the model (partly) who is manipulating objects as part of the demonstration of the task (Braaksma, Rijlaarsdam, & van den Bergh, 2002; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013; Hoogerheide, Loyens, & Van Gog, 2014; Van Gog, Verveer, & Verveer, 2014); showing the model in a lecture-style situation next to a screen on which a slideshow is projected that shows the steps needed to complete the task (Ouwehand, van Gog, & Paas, 2015) or on which the model is writing out those steps (Fiorella & Mayer, 2015; Exp. 1); showing only the slides or the model's writing in the form of a computer screen-recording with a voice-over explanation (Fiorella & Mayer, 2015; Exp. 3; see also www.khanacademy.org); or showing a screen-recording of the model working on a computer-based task, with or without a voice-over explaining the procedure (McLaren, Van Gog, Ganoe, Karabinos, & Yaron, 2016; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009).

It has been suggested that the effectiveness of the latter type of screen recording examples, in which the model is demonstrating a computer-based task, may be enhanced by showing the model's eye movements overlaid on the screen recording (Van Gog et al., 2009). In such *Eye Movement Modeling Examples* (Jarodzka et al., 2012; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Mason,

* Corresponding author. Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands.
E-mail address: t.v.a.vanmarlen@uu.nl (T. van Marlen).

http://dx.doi.org/10.1016/j.chb.2016.08.041
0747-5632/© 2016 Elsevier Ltd. All rights reserved.

Pluchino, & Tornatora, 2015a; Van Gog et al., 2009; see for a review Van Gog & Jarodzka, 2013) the model's eye movements are visualized by, for instance, a colored dot. It is expected that by showing the model's eye movements, learners' visual attention is synchronized with that of the model; in other words, that learners are attending to the relevant information at the right time.

That such guidance might be necessary, is suggested by research showing that novices attend to task irrelevant information (i.e., information that is high in visual contrast and therefore more salient), whereas experts attend to task relevant information faster and more often and are able to ignore irrelevant information (Charness, Reingold, Pomplun, & Stampe, 2001; Haider & Frensch, 1999; Jarodzka, Scheiter, Gerjets, & Van Gog, 2010; Van Gog, Paas, & Van Merriënboer, 2005; Wolff, Jarodzka, Van den Bogert, & Boshuizen, 2016). Hence, when novice learners are observing an expert's demonstration of the task, it is likely that their attention is not directed at the information the expert is attending to or referring to at the same time. Especially in cases in which the visual or verbal information in the video modeling example is transient, this might result in the learner missing out on the relevant information, which might hamper learning (see Ayres & Paas, 2007, for a discussion of transience and need for attention guidance in animations). By displaying the models' eye movements in the example, however, the learner not only sees *what* the model is doing on the computer, but also *where* the model is looking, which is hypothesized to guide learners' attention and to improve their learning outcomes by helping them to optimally process the video example (e.g., Jarodzka et al., 2012, 2013; Van Gog et al., 2009).

1.1. Attention guidance based on eye movement displays

Several different approaches have been taken to designing attention guidance based on the differences in attention allocation between experts and novices or successful and unsuccessful problem solvers. First, the observation that successful problem solvers allocate their visual attention to other information than unsuccessful problem solvers, has been used to design visual cues to guide visual attention to the information successful problem solvers attended to (Grant & Spivey, 2003; Groen & Noyes, 2010). And indeed, such cues resulted in higher solution rates on an insight problem-solving task (i.e., Duncker's radiation problem; Grant & Spivey, 2003; Thomas & Lleras, 2007).

Second, the eye movements themselves can be displayed to function as a visual cue. Also using Duncker's radiation problem, Litchfield and Ball (2011) investigated whether dynamically displaying a solution-related sequence of eye movements for 30 s would increase performance. In Duncker's radiation problem a schematic drawing of a tumor is presented surrounded by healthy tissue and skin. The goal is to destroy the tumor without damaging healthy surrounding tissue by means of converging low intensity lasers from multiple sides. Litchfield and Ball (2011) showed that a didactic (very deliberate, 'clean') or a natural (more chaotic) sequence of eye movements related to the solution (i.e., crossing the skin area from different angles), led to enhanced solution rates compared to eye movements focused on other areas of the task. Similar results of displaying another person's eye movements to guide attention and improve performance were obtained in studies with visual search tasks, in which people had to search for faults in software code (Stein & Brennan, 2004), faults on printed circuit boards (Nalanagula, Greenstein, & Gramopadhye, 2006), or lung-nodules on X-ray scans (Litchfield, Ball, Donovan, Manning, & Crawford, 2010). These studies show that attention guidance by displaying eye movements improved *performance*. However, they did not consider potential effects on *learning* (i.e., later performance in the absence of such guidance), which is the objective of displaying eye movements in modeling examples.

1.2. Learning from eye movement modeling examples

Research on *eye movement modeling examples* has found mixed support for the usefulness of displaying eye movements to guide attention and enhance learning. It seems that this kind of guidance is effective for learning tasks relying on visual inspection in order to classify or diagnose motion patterns from dynamic and visually rich stimuli. For instance, in the study by Jarodzka et al. (2013), participants had to learn how to classify fish locomotion patterns and were shown either only the video of the fish with the expert model's explanation, or they additionally saw the expert's eye movements. Consequently, when the expert verbally explained which fins the fish used for locomotion, the learners knew which fins he was referring to because they saw what he was looking at. The expert model's eye movements (i.e., fixations) were either visualized as a solid dot or as a 'spotlight' by means of blurring the video except for the part where the expert was fixating. After the video modeling examples participants were shown novel videos, without the expert's eye movements and verbal explanations, displaying fish locomotion patterns that they had to classify. Participants who had seen the model's eye movements showed marginally better performance on this classification task, with the dot condition outperforming the spotlight condition. In a similar vein, Jarodzka et al. (2012) showed that attention guidance by means of displaying the expert's eye movements in modeling examples, yielded superior learning outcomes. Participants had to learn to interpret symptoms of epileptic seizures in infants, either being shown only the video of the infant along with the expert model's verbal explanation, or they additionally saw the expert's eye movements being displayed either as a circle or as a spotlight. The spotlight condition outperformed the condition that did not receive attention guidance.

Eye movement modeling examples were also shown to be effective in learning a text-picture processing strategy (Mason et al., 2015a). Children who were presented with an example that showed a model's eye movements, with the model deliberately making transitions between corresponding elements of the text and picture in order to emphasize integration, showed more text picture integration (i.e., number of transitions between text and picture) on a novel text and recalled more information units and performed better at the transfer test about that novel text than children in the control condition who did not receive such an example. Recently, these results were replicated and extended by showing that children with lower reading comprehension skills benefitted more from eye movement modeling examples regarding factual knowledge and the transfer of knowledge, compared to children with high reading comprehension skills (Mason, Pluchino, & Tornatora, 2015b). Thus, EMME are not only effective for learning a domain-specific task, but also for learning general processing strategies.

In contrast, when it comes to learning procedural problem-solving tasks, guiding the learners' attention by displaying the model's eye movements did not yield beneficial effects on learning, and even had a negative effect on learning when the modeling example also contained a verbal explanation (Van Gog et al., 2009). In this study, participants were shown an example of how to solve an animated puzzle problem (i.e., frog leap) with or without a verbal explanation and with or without the model's eye movements being displayed. All examples showed a screen recording of the solution steps, which were executed by the model clicking on a frog to move it forward. The verbal information (when present) explained the different choice options at each step, and indicated which options were incorrect and why. The displayed eye movements also showed the model considering the various choice options. In the conditions in which no verbal explanation was given, seeing the model's eye movements did not affect learning of the puzzle problem (although there seemed a slight advantage of attention guidance with regard to transfer, i.e., completing the problem in a different order). In the conditions in which a verbal explanation was present, seeing the model's eye movements had a *negative* effect on subsequent test performance. A possible explanation offered by the authors for this negative effect, was that the verbal explanation might have been sufficient to guide attention to the right location at the right time. As a result, the cues provided by the eye movements may have been redundant (and research on the redundancy effect shows that displaying redundant information does not help and can even hinder learning; Kalyuga & Sweller, 2014).

Note though, that the procedural problems also differ from the 'classification' (Jarodzka et al., 2012, 2013) and 'strategy' (Mason et al., 2015a, 2015b) examples –and for that matter, from the insight and search problems discussed above-in terms of the model's interaction with the display of the task. In contrast to the other tasks, where the model is inspecting a (complex) visual display but not acting upon it, a procedural problem-solving task usually requires the model to act upon the objects in the problem (e.g., by moving the mouse or typing), and such overt actions will also automatically draw the learners' attention. So although the eye movements, like the verbal explanations, provide additional information on covert cognitive actions (e.g., the choice processes) that are relevant for understanding the overt actions, the executed steps in the solution procedure will automatically draw the learners' attention.

1.3. The present study

Based on the literature reviewed above, it seems possible that attention guidance based on the model's eye movements would be less effective, or even ineffective, for procedural problem solving tasks. If that would be the case, this would be relevant for educational research and practice, as it would provide insight into the conditions under which attention guidance based on eye movement displays is or is not effective. However, since one cannot draw a conclusion about a whole category of tasks based on one single study, the present study aimed to investigate whether attention guidance based on eye movements can be effective for learning to solve procedural geometry problems (i.e., higher performance and faster response times). In Experiment 1, learners were presented with *simple* geometry problems without verbal explanations, whereas in Experiment 2, learners were presented with more *complex* geometry problems that did include a verbal explanation.

2. Experiment 1

Experiment 1 investigated whether participants would benefit from seeing the model's eye movements in examples of simple, geometry problems that only required solving one angle (from hereon: one-angle problems). Solving that angle in a task involving the F-rule, for instance, required the following simple steps to be (mentally) performed: (1) searching for the asked unknown angle; (2) locating the parallel sign; (3) searching for the second parallel sign; (4) identifying the parallel lines; (5) applying the corresponding angle principle (i.e., F-rule) and solving the problem. Whereas participants in the control condition only saw the model eventually typing in the answer, participants who additionally received meaningful guidance additionally saw the model's eye movements, which signaled to what information the model was attending, as well as the underlying principle (corresponding/alternating angles based on the 'F'/'Z' rule).

It is possible that seeing eye movements might raise participants' overall attention to the task, which in itself might result in learning benefits, irrespective of the usefulness/meaningfulness of the eye movements (cf. Litchfield et al., 2010). Therefore, we included a condition in which meaningless eye movements were displayed. If, despite prior research with a puzzle problem (Van Gog et al., 2009), attention guidance based on the model's eye movements would be effective for learning procedural problem-solving tasks, then the meaningful eye movement display condition should show better learning outcomes (i.e., higher accuracy and faster response times on the learning and transfer problems) compared to the meaningless eye movements and no eye movements condition.

2.1. Methods

2.1.1. Participants and design

Seventy-two students of a Social Sciences Faculty of a Dutch university (the majority from the Psychology program) volunteered to participate in this study ($M_{age} = 19.94$, $SD = 2.10$, 18–29 years; 19 male). They were randomly assigned to one of three conditions ($n = 24$ in each condition): meaningful eye movement modeling examples (EMME), meaningless EMME, or control (i.e., modeling example only). Participants received either a monetary reward or course credit for participating. All participants had normal or corrected to normal vision.

2.1.2. Materials and apparatus

2.1.2.1. Eye tracking equipment. Eye movements were recorded with a SMI RED250 eye tracker with a sampling rate of 250 Hz (SensoMotoric Instruments, GmbH). The experiment was created in SMI *Experiment Center 3.34* software and presented on a monitor with 1680 × 1050 pixels resolution with a refresh rate of 60 Hz.

2.1.2.2. Pretest. A pretest was administered to check whether the level of prior knowledge among conditions was equal. The pretest consisted of four open questions regarding the geometry problems presented with *Experiment Center* (e.g., "What is a triangle?", "What is a straight line?", "What is an alternating angle?" and "What is a corresponding angle?").

2.1.2.3. Geometry problems. Four types of geometry problems were created: triangle problem, straight line problem, alternating angle problem (i.e., Z-rule; see Fig. 1), and corresponding angle problem (i.e., F-rule). Each problem consisted of line drawings of triangles and parallel lines. All angles were coded A, B, C etc. Values were given for some of the angles and unknown values of angles were marked with a question mark. On top of each line drawing the problem statement was provided (e.g., "How many degrees is angle B?"). For each type of problem a modeling example (see below) and two isomorphic problems were created. One isomorphic problem of each type was identical to the modeling examples in terms of layout, but differed in terms of the numbers used for the angles. The other isomorphic problem had a comparable (but not identical) layout.

In addition, two transfer problems were created, which had a visually more complex layout than the example and isomorphic problems, as they combined more angles, parallel lines, triangles and straight lines. The geometry problems were created and presented with the program *Geogebra* (www.geogebra.org). See Fig. 2 for an example of a transfer problem. The width of the rectangular area containing both the geometry figure and problem statement ranged between 747 and 1478 pixels across problems, and the height ranged between 467 and 780 pixels.

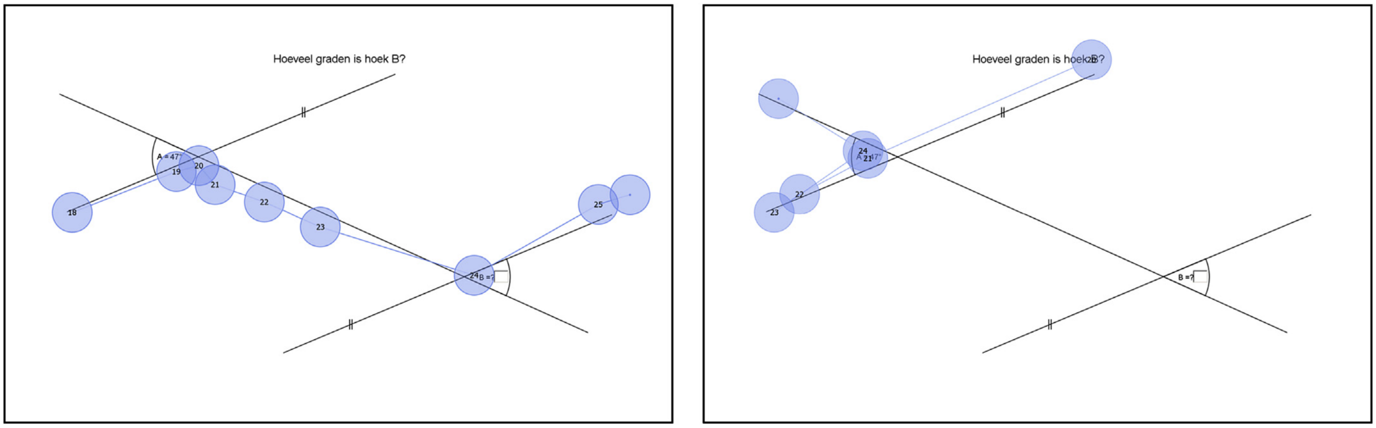


Fig. 1. Static representation of the dynamic scan path shown in a meaningful EMME (left) and a meaningless EMME (right) in Experiment 1 for the alternating angle problem. On top is the problem statement stating (translated from Dutch) "How many degrees is angle B?". The blue dots represent the location of the model's gaze and the numbers represent the order of the fixations number. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

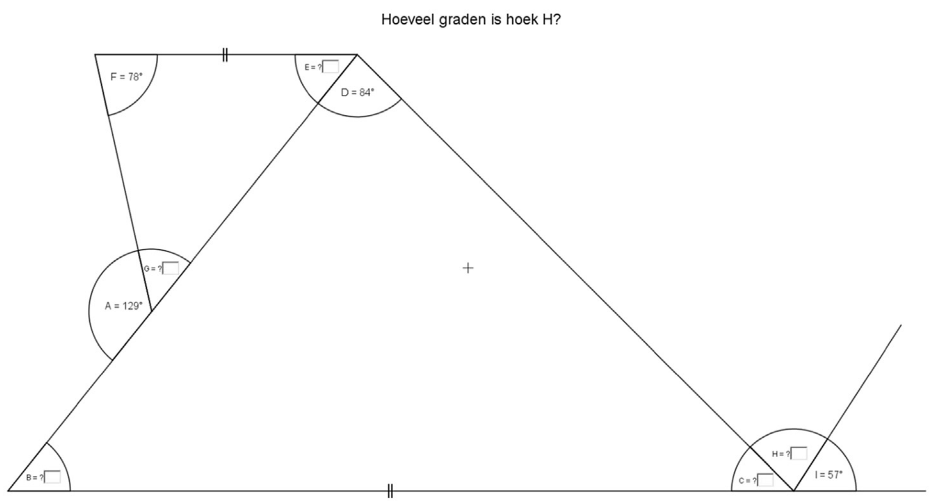


Fig. 2. Example of a transfer problem in Experiment 1 with the problem statement (translated from Dutch) stating "How many degrees is angle H?".

2.1.2.4. Modeling examples. One example was created for each problem type, using SMI *Experiment Center 3.34* for recording the eye movements. SMI *BeGaze 3.3* was then used for superimposing the eye movements onto the example problem with the Bee Swarm function. The model's eye movements were represented as a blue translucent dot with a diameter of 30 pixels. In the meaningful EMME condition, the eye movement modeling example showed how to solve the problem (e.g., for the alternating angle problem the model's gaze followed a Z-shaped pattern, see the left side of Fig. 1). In the meaningless EMME condition, the eye movement modeling example focused on all features of the problem and on regions between the features but not in a meaningful order (e.g., for the alternating angle problem, the model's gaze did not move in a Z-shaped pattern, see the right side of Fig. 1). In the control condition, participants saw the problem being solved without a model's gaze. In both the meaningful and meaningless modeling examples, the model's started with reading the problem statement. All modeling examples, regardless of condition, ended by showing the solution to the problem statement for 2 s. In the meaningful condition the model's gaze fixated on the solution when it was shown, whereas the model's gaze in the meaningless condition fixated on the problem statement at that time (which would make students slower to notice the problem solution if they would indeed follow the meaningless gaze).

The length of the modeling videos for the triangle problem was between 20.2 and 20.8 s depending on condition, for the straight line problem it was between 12.5 and 12.9 s, for the alternating angle problem it was between 32.7 and 33.2 s, and for the corresponding angle problem it was between 33.9 and 34.3 s. The small differences in video length (up to a maximum of 0.6 s) were caused by the different eye movement patterns in the two EMME conditions.

2.1.3. Procedure

The experiment was run in individual sessions of approximately 25 min duration. Participants were briefly instructed about the general overview of the experiment, when they entered the lab. Then participants answered the pretest questions. After the pretest, they were seated properly in front of the eye tracker with the help of a forehead and chin-rest, which was positioned 57 cm in front of the monitor.[1] After the five-point calibration (plus four-point

[1] Note that we collected eye movement data to be able to explain whether possible effects of EMME would indeed arise via gaze following behavior during example study; in the absence of such effects, however, these data were not analyzed.

validation) procedure, participants were told that they would be presented with video examples in which they saw how to solve geometry problems and that each video showed the solution of the problem. Participants in both EMME conditions were additionally informed that they would see the eye movements made by the model, represented as blue translucent dots. Then, participants were instructed that after each video example they had to solve a similar geometry problem themselves as accurately and as fast as possible. This example-problem sequence was repeated four times, after which participants completed the remaining geometry problems (i.e., four isomorphic and two transfer problems). The order of the example-problem pairs was counterbalanced across participants and conditions, as was the order of the remaining four isomorphic test problems. The transfer problems always came last. Finally, participants were asked to indicate in how many of the modeling examples they knew the solution before it was shown (ranging from zero to four) to obtain additional information on their prior knowledge and an impression of the usefulness of the modeling example.

2.1.4. Data analysis

On the pretest, one point was given for each correctly answered question and a half point was given for partially correct answers, resulting in a maximum score of 4 points. Because these were open questions, a second independent rater scored a randomly selected subset (16.67%) of answers. The percentage absolute agreement between the raters was 93.75% with a linear weighted Cohen's kappa (Cohen, 1968), which takes into account the agreement based on chance and the degree of disagreement, of $\kappa\omega = 0.712$, indicating substantial agreement (Landis & Koch, 1977).

For each correctly solved isomorphic and transfer problem, one point was given, resulting in a maximum score of 8 for the isomorphic problems, and a maximum score of 2 for the transfer problems. A half point was given when participants made a minor calculation error (i.e., the given answer deviated 1, 10 or 100° from the correct answer; $n = 19$) but showed they knew how to solve the problem. Performance scores on isomorphic and transfer problems were summed and then converted to a proportion of the maximum score. Besides task performance, mean response times (RT) for correctly completed problems were computed for the eight isomorphic and two transfer problems separately. Only correctly solved items were included as it is uninformative to know how much time is needed to incorrectly 'solve' a problem and the RT for incorrectly answered problems can affect the average RT in several ways. On the one hand participants who struggle from the start might 'give-up' early on, which results in a lower average RT. On the other hand struggling participants might keep on trying to solve the problem resulting in higher RT's. To avoid any kind of bias, only the RT for correct solved problems were used.

For several reasons, 5 problems had to be excluded from the performance and RT data analysis (i.e., technical issues: $n = 2$, participant asking question during problem solving: $n = 3$) and 8 problems were excluded from the RT analysis only (i.e., technical issues: $n = 4$, mouse cursor lost: $n = 1$, software bug: $n = 3$). In those cases the average performance or RT were calculated based on the data of the remaining problems.

2.2. Results

The data were analyzed with one-way ANOVAs, except when the assumption of normality was violated or the group sizes were small (in case of the response time analysis of the transfer problems, see below) or unequal, in which case the more conservative non-parametric Kruskal-Wallis test was conducted. In case the assumption of homogeneity was violated we analyzed the data with a one-way ANOVA using Welch's corrected F value (Field, 2009). See Table 1 for the complete overview of the number of participants and items for each analysis. There were no significant differences in pre-test performance among conditions ($M = 0.44$ points; $SD = 0.66$), $F(2, 69) = 1.20$, $p = 0.309$, nor did they differ significantly in terms of how many modeling examples they knew the answer to before it was shown ($M = 2.76$; $SD = 1.16$), $F(2, 69) = 0.75$, $p = 0.475$.

2.2.1. Performance

To address our research question of whether seeing a meaningful EMME resulted in enhanced performance as compared to the meaningless EMME and control condition, non-parametric Kruskal-Wallis tests were conducted with condition as between-subjects variable and performance on the isomorphic and transfer problems as dependent variables. Neither performance on the isomorphic problems, $H(2) = 0.34$, $p = 0.842$, nor performance on the transfer problems, $H(2) = 1.35$, $p = 0.509$, differed significantly among the conditions (see Fig. 3).

2.2.2. Response times

We then examined the question of whether the meaningful EMME condition needed less time to correctly solve the isomorphic and transfer problems as compared to the meaningless and control condition. On the isomorphic problems, a one-way ANOVA was conducted using Welch's corrected F value with condition as between-subject variable and RT as dependent variable. This analysis did not reach statistical significance, $F(2, 43.507) = 2.67$, $p = 0.081$, $\eta_p^2 = 0.10$ (see Fig. 4), yet we decided to exploratively conduct Games-Howell *post hoc* tests (more suited for data with unequal variances; Field, 2009), revealing: meaningful vs. meaningless EMME ($p = 0.127$), meaningful vs. control condition ($p = 0.836$), and meaningless vs. control ($p = 0.068$, $r = 0.32$). Despite not reaching statistical significance, this last, medium effect size, suggests that it took participants in the meaningless EMME condition more time to correctly solve the isomorphic problems than participants in the control condition.

For the transfer problems the non-parametric Kruskal-Wallis test was conducted with condition as between-subject variable and RT as dependent variable (see Table 1). The difference in RT among the modeling conditions did not reach statistical significance, $H(2) = 5.16$, $p = 0.076$ (see Fig. 4), yet we decided to exploratively conduct Hochberg's GT2 *post-hoc* test (more suited for data with unequal group sizes; Field, 2009), revealing: meaningful vs. meaningless EMME ($p = 0.826$), meaningless vs. control ($p = 0.153$), and meaningful vs. control ($p = 0.038$, $r = 0.42$). This last, medium effect size suggests that it took participants in the meaningful EMME condition less time to correctly solve the transfer problems than participants in the control condition.

2.3. Discussion

The hypothesis that meaningful EMME would yield higher performance was not confirmed, as there were no significant performance differences among conditions. Regarding response times, our hypothesis was not confirmed either, although there were trends suggesting that participants who had observed meaningful EMME were faster in solving the transfer problems than participants in the control condition and that participants in the meaningless EMME condition were slower in solving the isomorphic problems than participants in the control condition, but this difference did not reach statistical significance. This suggests that participants were somewhat hindered (meaningless EMME) or helped (meaningful EMME) by the displayed eye movements, but this was not sufficient to affect their performance —possibly

Table 1
Mean (and SD) of performance, number of participants, and number of items excluded/included in each analysis in Experiment 1.

	Isomorphic problems		Transfer problems	
	Proportion correct	Response time (s)	Proportion correct	Response time (s)
Meaningful EMME	0.90 (0.13)	11.91 (2.97)	0.38 (0.42)	106.86 (34.80)
N	24	24	24	12
Excluded items	0	18 (9.38%)[a]	1 (2.08%)	29 (60.42%)[a]
Included items	192	174	47	19
Meaningless EMME	0.90 (0.13)	14.64 (5.98)	0.34 (0.37)	119.63 (43.01)
N	24	24	24	13
Excluded items	1 (0.52%)	18 (9.38%)[a]	0	30 (62.5%)[a]
Included items	191	174	48	18
Control	0.86 (0.19)	11.37 (3.59)	0.26 (0.40)	156.88 (47.94)
N	24	24	24	8
Excluded items	3 (1.56%)	33 (17.19%)[a]	0	35 (72.91%)[a]
Included items	189	159	48	13

[a] Note that only correctly solved problems were included in the response times analyses. Thus, the number of excluded items represents both items excluded due to technical difficulties and items excluded due to performance errors.

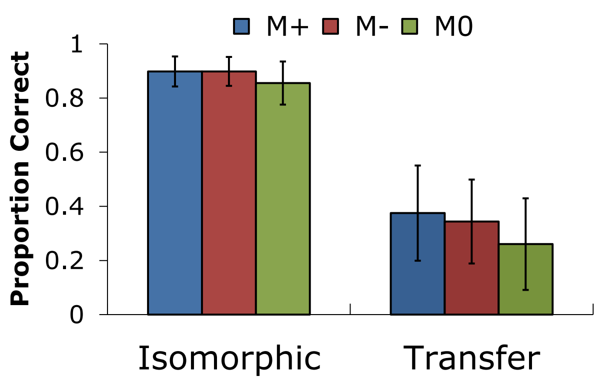


Fig. 3. Proportion correct for the isomorphic and transfer problems in Experiment 1 for the meaningful eye movements modeling condition (M+), meaningless eye movements modeling condition (M-), and control condition (M0). Error bars represent the 95% confidence interval of the mean.

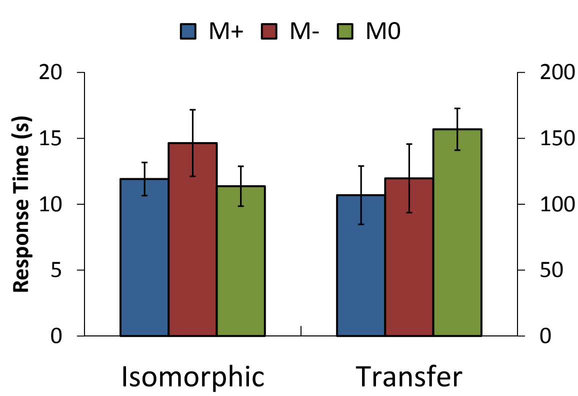


Fig. 4. Response times expressed in seconds for correctly answered isomorphic and transfer items for the meaningful eye movements modeling condition (M+), meaningless eye movements modeling condition (M−), and control condition (M0). Errorbars represent the 95% confidence interval.

because of ceiling effects on the isomorphic problems.

In addition, the geometry problems presented in Experiment 1 might not have been complex enough to require attention guidance: the problems consisted of simple geometrical shapes that could be solved almost on-sight when the principle was understood. Indeed, participants indicated they knew the answer to most of the examples before it was shown. The materials used in previous studies that found positive effects of displaying another person's eye movements prior to problem solving (Litchfield & Ball, 2011; Litchfield et al., 2010) or during example study (Jarodzka et al., 2012, 2013; Mason et al., 2015b, 2015a) were more visually complex. Seeing a model's eye movements is arguably most useful when a student cannot automatically infer what the model is doing. Therefore, a second experiment was conducted with more procedurally and visually complex geometry problems that required more solution steps.

3. Experiment 2

The geometry problems in Experiment 2 were more complex than the one-angle problems from Experiment 1, as they required four angles to be solved and included known and unknown angles that were irrelevant for the solution procedure. A pilot study showed very poor performance on such problems. Because of the increase in complexity, verbal explanations were added to the modeling examples. The meaningless EMME condition was omitted from Experiment 2.

As in Experiment 1, we predicted that if seeing an EMME successfully guides students' attention to the right location at the right time, this would result in higher learning outcomes (i.e., performance) and faster problem solving (i.e., faster response times) as compared to the control condition. To establish whether EMME indeed guide attention, eye tracking was used to explore whether participants in the EMME condition would fixate the relevant areas mentioned in the verbal explanation more often, faster, and longer compared to participants in the control condition.

3.1. Methods

3.1.1. Participants and design

Sixty-eight students of a Social Sciences Faculty of a Dutch university (the majority from the Psychology program) volunteered to participate in this experiment ($M_{age} = 21.12$, $SD = 1.93$; 18–26 years 23 male). They were randomly assigned to the EMME or the control condition. Two participants had to be excluded from all analyses because they accidentally skipped one of the video modeling examples. In addition, four participants had to be excluded as they already participated in Experiment 1 (this was not noticed beforehand due to an error in the registration system). This left $n = 30$ in the control condition and $n = 32$ in the EMME condition. Participants received either a monetary reward or course credit for participating. All participants had normal or corrected to normal vision.

3.1.2. Materials and apparatus

3.1.2.1. Eye tracking equipment. The eye tracking equipment used in Experiment 2 was the same as in Experiment 1. Experiment Center 3.4.165 (SensoMotoric Instruments, GmbH) software was used for the creation and presentation of the experiment. The screen recording function of the Experiment Center software was used to record participants' eye movements. The geometry problems were presented with the program Flash Adobe CC and programmed with ActionScript 3.0.

3.1.2.2. Pretest. A pretest was administered to check whether the amount of prior knowledge among conditions was equal. The pretest consisted of three multiple-choice questions and two open short-answer questions. The questions tested the geometry knowledge regarding triangles, straight lines, corresponding angles, and alternating angles.

3.1.2.3. Geometry problems. Geometry problems were created in a similar way as in Experiment 1, but the problems were made more complex in that they combined different geometrical principles within one problem/example, required four angles to be solved, and contained a total of thirteen angles. Two modeling examples were created along with two isomorphic problems (i.e., an identical layout but different numbers) and four transfer problems with different visual layouts and numbers. To prevent participants from making calculation errors a digital calculator was added at the top-right corner of the screen. The width of the rectangular area containing the geometry figure, the problem statement, and the calculator ranged between 1092 and 1351 pixels across problems and the height ranged between 787 and 847 pixels (see Fig. 5).

3.1.2.4. Modeling examples. The modeling examples were recorded in a similar fashion as in experiment 1, using SMI Experiment Center 3.4.165 to record the eye movements and SMI BeGaze 3.4.52 for creating the video examples. The EMME and control versions of the modeling examples both contained a male model's narration explaining the different steps (see Appendix for a screenshot of a modeling example and for the corresponding transcript of the verbal explanation steps). The model first identified the location of the to-be-solved angle. Once the angle was identified, the model started looking for a starting point for the solution by working backwards. Then, the model began solving the geometry problem and explained each solution step until the problem was solved. In the control condition, participants heard the verbal explanation and saw the answers to each step appear. In addition, participants in the EMME condition, saw the model's eye movements superimposed onto the modeling example. As in Experiment 1, the eye movements were displayed as a blue translucent dot with a diameter of 30 pixels. The duration of the two modeling example videos was identical across conditions (with the first lasting 122 and the second lasting 131 s).

3.1.3. Procedure

The procedure was similar as in Experiment 1 with the exception of the number of video examples and transfer problems presented. In addition, a short EMME demonstration video of a model looking at a picture of a living room was added in Experiment 2 as part of the general instruction. This was done to familiarize participants in the EMME condition with the representation of the model's eye movements. After this instruction they were presented with two example-problem pairs. Subsequently, the eye tracker was re-calibrated and participants were presented with the four transfer problems. The order of the example-problem pairs was counterbalanced across participants and the four transfer problems were presented in a fixed order. For each problem, participants' performance, response times, and eye movements were recorded.

3.1.4. Data analysis

3.1.4.1. Prior knowledge. For each question on the pretest only one answer was correct and for each correct answer one point was given (i.e., max. score = 5 points).

3.1.4.2. Performance. Two participants were excluded from the performance analyses due to very poor performance (z-score < -2.5). In addition, due to a technical issue, two participants received one geometry problem less (i.e., their proportion scores on

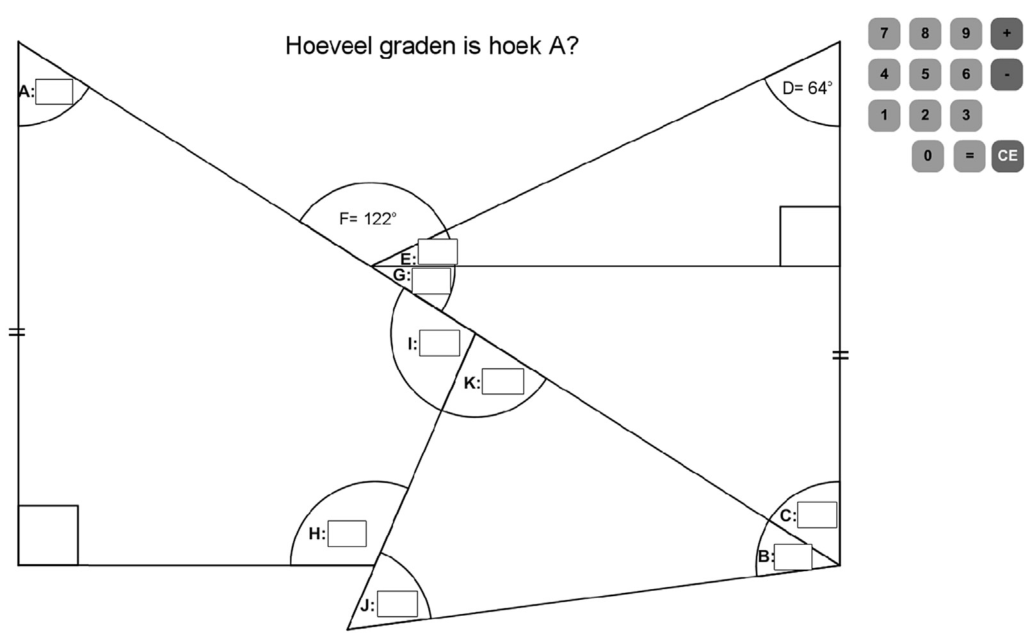


Fig. 5. An example of a transfer problem in Experiment 2 including the calculator. On top is the problem statement, (translated from Dutch) “How many degrees is angle A?”.

the transfer problems are based on three instead of four problems).

3.1.4.3. Response times. For the response times analyses, two additional participants were excluded due to very high response times (z-score > −2.5).

3.1.4.4. Eye tracking measures. To explore whether EMME helped participants to attend to the right visual information at the right time, we first determined the onset of the verbal referents in the narration (i.e., in the sentence "Now that you know angle *A*, you can calculate angle *C*." the letters A and C are the verbal referents; there were 15 verbal referents in each modeling example) and the corresponding areas of interest (AOI) in the geometry problem (e.g., the degree of angle A or C). We then determined the proportion of verbal referents fixated (i.e., number of verbal referents fixated divided by total number of referents; *proportion fixated)*, how long it took participants to fixate the referent (*time lag*) after it was mentioned in the verbal explanation, and how long the verbal referent was fixated (*fixation duration)*, with fixations defined as yielding a peak velocity $\leq 40^{\circ}$/s and fixation duration $\geq$ 100 ms (cf.Jarodzka et al., 2013; Litchfield et al., 2010). Only fixations on a verbal referent's corresponding AOI that occurred within a time window of 1500 ms after the onset of the verbal referent were included in the analyses (cf. Dahan & Tanenhaus, 2005). In addition, fixations occurring within the first 100 ms were excluded from the analysis, as research indicates that initiating eye movements based on language input takes approximately 100 ms or longer (Altmann, 2011).

Three participants were excluded from all eye tracking analyses due to low tracking ratio (i.e., number of gaze points on the screen recorded by the eye tracker divided by the total duration of the experiment; z-score < −2.5; $n = 2$) or bad calibration measures (i.e., deviation > 1 deg; 1st calibration deviation: M = 0.42 deg; SD = 0.12 deg; 2nd calibration deviation: M = 0.41 deg; SD = 0.13 deg; $n = 1$). One additional participant was excluded from the Time Lag analysis only due to very short time lags (z-score < −2.5). Two additional participants were excluded from the Fixation Duration analysis only due to very long fixation durations (z-score > 2.5).

3.2. Results

The data were analyzed with independent samples t-test's. In the case the assumption of normality was violated the non-parametric Mann-Whitney U test was conducted and reported. See Table 2 for the complete overview for the number of participants and items for the performance and response times analyses and see Table 3 for a similar overview for the eye-track measures analyses. There were no significant differences between conditions in participants' prior knowledge (pretest: M = 3.77; SD = 1.21), t (60) = 0.162, p = 0.872 (this did not change when excluding the outliers re. posttest performance: M = 3.83; SD = 1.18, t (58) = −0.22, p = 0.829).

3.2.1. Performance

To address our research question of whether seeing an EMME resulted in enhanced performance as compared to the control condition, non-parametric Mann-Whitney U tests were conducted. These revealed that neither performance on the isomorphic problems, U = 435.00, z = −0.46, p = 0.643, r = 0.06, nor performance on the transfer problems, U = 448.50, z = −0.02, p = 0.981, r = −0.003, differed significantly between conditions (see Fig. 6).

3.2.2. Response times

In order to examine whether the EMME condition needed less time to correctly solve the isomorphic and transfer problems as compared to the control condition, Mann-Whitney U tests were conducted on response times (in s) for correctly solved isomorphic and transfer problems. Participants in the EMME and control condition did not differ significantly in the time they took to solve the isomorphic problems, U = 345.00, z = −1.17, p = 0.243, r = −0.15. However, on correctly solved transfer problems, the response times were *higher* in the EMME condition (M = 128.15, SD = 30.73) than the control condition (M = 110.02, SD = 29.50), U = 238.00, z = −2.68, p = 0.007, r = −0.35, meaning that participants in the EMME condition were *slower* at solving the geometry problems (see Fig. 7).

3.2.3. Eye tracking measures

To explore whether EMME were beneficial for guiding attention to the relevant information at the right time in the video modeling examples, independent samples t-tests were conducted. This revealed no significant differences in the proportion of timely fixated AoIs mentioned in the verbal explanation (i.e., within 1500 ms after the onset of the referent), between the EMME (M = 0.43, SD = 0.13) and control condition (M = 0.46, SD = 0.18), t (57) = 0.64, p = 0.527. In terms of the time (in ms) required to first fixate the verbal referents after onset, participants in the EMME condition (M = 663.99, SD = 127.83) were significantly faster than participants in the control condition (M = 729.63, SD = 110.04), t (56) = 2.09, p = 0.041, r = 0.27. Finally, there was a significant difference in fixation duration (in ms), t (55) = −2.41, p = 0.019, r = 0.31, with a medium effect size showing that participants in the EMME condition (M = 524.27, SD = 149.02) fixated the verbal referents longer than the control condition (M = 434.62, SD = 129.39).

3.3. Discussion

The first part of our hypothesis, that EMME would help guide students' attention to the right information at the right time, was partly confirmed: participants in the EMME condition fixated the verbal referents significantly faster and longer; however, there was no difference between conditions in the proportion of timely fixated verbal referents. In contrast to our expectation though, the attention guidance did not result in better learning outcomes (which is in line with some studies on visual cueing: e.g., De Koning, Tabbers, Rikers, & Paas, 2010; Jarodzka et al., 2013; Kriz & Hegarty, 2007): performance on the isomorphic and transfer problems did not differ between conditions and students in the EMME condition were not faster at problem solving than students in the control condition. In fact, they were even *slower* at solving the transfer problems compared to the control condition.

In this respect the findings of Experiment 1 and 2 are seemingly contradictory. That is, in Experiment 1, participants in the meaningful EMME condition were *faster* at solving the transfer problems than participants in the control condition, whereas in Experiment 2, they were *slower*. One (speculative) explanation for this finding might be that the geometry problems demonstrated in the video modeling examples were more complex in Experiment 2 than in Experiment 1, whilst the transfer problems of both experiments were comparable in terms of complexity and number of steps. Having seen the eye movements related to the F or Z rule in the EMME in Experiment 1 might have allowed participants to locate the corresponding/alternating angles in the transfer problems faster, resulting in faster solving speeds, compared to participants who did not receive any guidance. The EMME in Experiment 2, on the other hand, showed a much longer and multi-step search and solution procedure. It is possible that participants in the EMME condition were attempting to mentally simulate how the model in the EMME would solve the problem, and since the model explained and solved the problem in a didactic manner this might have

Table 2
Mean (and SD) of performance, number of participants, and number of items excluded/included in each analysis in Experiment 2.

	Isomorphic problems		Transfer problems	
	EMME	Control	EMME	Control
Proportion correct	0.95 (0.15)	0.97 (0.13)	0.78 (0.22)	0.77 (0.28)
N	30	30	30	30
Excluded items	4 (6.67%)	0	13 (10.83%)	1 (0.83%)
Included items	56	60	107	119
Response time (s)	79.44 (15.14)	89.26 (34.01)	128.16 (30.73)	110.02 (29.50)
N	28	30	28	29
Excluded items	9 (16.07%)[a]	2 (3.33%)[a]	34 (28.33%)[a]	29 (24.17%)[a]
Included items	47	58	86	91

[a] Note that only correctly solved problems were included in the response times analyses. Thus, the number of excluded items represents both items excluded due to technical difficulties and items excluded due to performance errors.

Table 3
Mean (and SD) of eye tracking measures, number of participants, and number of items excluded/included in each analysis in Experiment 2.

	EMME	Control
Proportion fixations	0.43 (0.13)	0.46 (0.18)
N	31	28
Excluded items	0	0
Included items	62	56
Fixation duration (ms)	524.27 (149.02)	455.04 (166.75)
N	30	27
Excluded items	2 (3.23%)	2 (3.70%)
Included items	60	54
Time lag (ms)	663.99 (127.83)	729.63 (110.04)
N	30	28
Excluded items	2 (3.23%)	0
Included items	60	56

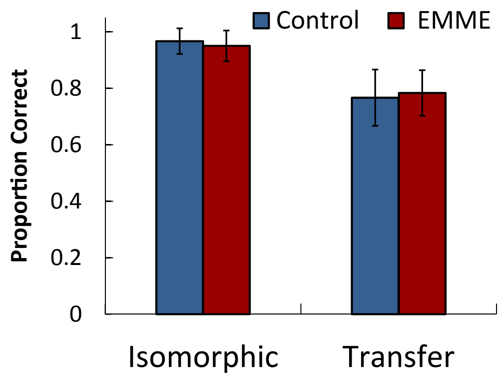


Fig. 6. Proportion correct for the isomorphic and transfer problems in Experiment 2 for the EMME and control condition. Error bars represent the 95% confidence interval of the mean.

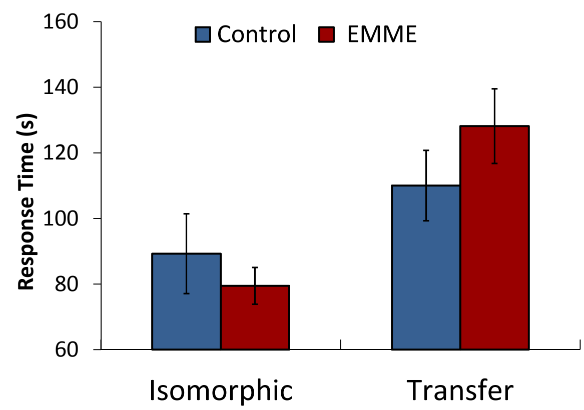


Fig. 7. Response times (in seconds) for the isomorphic and transfer problems in Experiment 2 for the EMME and control condition. Error bars represent the 95% confidence interval of the mean.

resulted in longer response times compared to the control condition. Yet, this explanation is rather speculative and it is also possible that the explanation might simply lie in the smaller number of participants and transfer problems in Experiment 1 than Experiment 2 (and the associated power/error issues). Nevertheless, future research should try to shed more light onto the question of whether participants attempt to "copy" the modeled eye movements on novel problems.

4. General discussion

In two experiments, we aimed to investigate whether studying eye movement modeling examples (EMME), in which students not only see the model performing the problem-solving steps in simple (Experiment 1) and complex (Experiment 2) geometry problems, but also see what the model is looking at, would enhance learning outcomes compared to regular examples in which the model's eye movements are not displayed. Neither experiment revealed benefits of EMME on learning outcomes compared to the control condition. This is in contrast to recent studies that have found EMME to be more effective than no EMME in enhancing learning of classification tasks (Jarodzka et al., 2012, 2013) and text-picture integration strategies (Mason et al., 2015a, 2015b). However, it is in line with prior research that failed to find a beneficial effect of EMME on procedural problem solving tasks (Van Gog et al., 2009).

So do our findings imply that EMME might not be effective for learning procedural problem-solving tasks compared to examples that show only the problem-solving steps? As mentioned in the introduction, this might very well be the case, since the procedural problems differ from the 'classification' (Jarodzka et al., 2012, 2013) and 'strategy' (Mason et al., 2015a, 2015b) examples in terms of the model's interaction with the task. That is, most procedural problem-solving tasks require the model to interact with objects presented on the visual display (e.g., making calculations, typing in answers, moving or dragging objects with the mouse cursor), which in itself guides a learner's attention to the right place at the right time. Although the displayed eye movements provide additional information on covert cognitive actions (e.g., the choice processes) that are relevant for understanding the overt actions, the executed steps in the solution procedure will automatically draw the learners' attention and are arguably most important for learning the procedure. When the model does not interact with the display (cf. Jarodzka et al., 2012; Mason et al., 2015b, 2015a), there are no overt actions and consequently, the displayed eye movements (which make covert processes visible) might make a more important contribution to students' attention and learning.

However, there may be another explanation. Whereas the tasks in Experiment 1 may have been too easy to establish any effects of EMME, the verbal explanations that were present in the examples in Experiment 2, may have further served to guide students' attention. Although the eye tracking data from Experiment 2 did show that students who studied EMME fixated on relevant information referred to in the verbal explanation significantly faster and longer, there were no differences among conditions in the amount of relevant information that was fixated following the verbal referents. Thus, even though the students who saw the model's eye movements got there faster, it seems that the verbal explanation were also sufficient to guide students' attention to the right location at the right time. As such, the verbal explanation may have made the guidance provided in the EMME redundant (see also Van Gog et al., 2009), and it is known that presentation of redundant information does not contribute to, or may even hamper learning (Kalyuga & Sweller, 2014). While studying effects of EMME without verbal information might also be interesting in future research (if the examples can be understood without explanation), note that the findings by Van Gog et al. (2009), who included an EMME and no-EMME condition without verbal explanations, do not give reason to think that this would have led to significant differences between EMME and regular examples.

Hence, before we can say with certainty whether EMME are not effective for learning procedural problem-solving tasks, future research should investigate whether the model's interactions with the task making the guidance provided by EMME redundant, or whether the verbal explanations provided by the model do so. This could be investigated, for instance, by manipulating the extent to which the verbal explanation can be unambiguously interpreted by the participant. That is, when the verbal explanation clearly states what to look for (e.g., "angle B") and there is only one element of the visual display that fits the description (i.e., only one angle labeled 'B'), as was the case in the present study, then additional attention guidance may not be required. If, on the other hand, the verbal explanation is ambiguous for a student, either because of their knowledge of the referents (e.g., in the study by Jarodzka et al., 2013, one would have had to know what a dorsolateral fin is in order to attend to the part of the fish that the model was talking about), or because of characteristics of the display (e.g., the referent may refer to one of several locations), then attention guidance might be necessary to attend to the right information at the right time, having a stronger effect on learning.

Eye movement research provides support for this idea that verbal information influences a listener's eye movements and that the coupling of eye movements between speaker and listener may affect memory. For instance, in one study by Richardson and Dale (2005), people listened to someone describing a video-clip of a TV show while watching pictures of the characters from that show. It was found that the eye movements of the speakers and listeners to the various characters referred to by the speaker, were very similar (with some delay, given that the listener first had to process the information and then move the eyes to the same location) but also that the amount of correspondence between the eye movements predicted the score of a comprehension test (Richardson & Dale, 2005). In a follow-up study on dialogues, participants' prior knowledge about a painting was manipulated. Interestingly, it was found that the eye movements of dyads engaged in dialogue showed greater correspondence when they had previously heard the same information than when they had heard different information about the painting (Richardson, Dale, & Kirkham, 2007; Experiment 2). This suggests that when there is ambiguity in a verbal description because of prior knowledge, listeners might not timely attend to the same information as speakers.

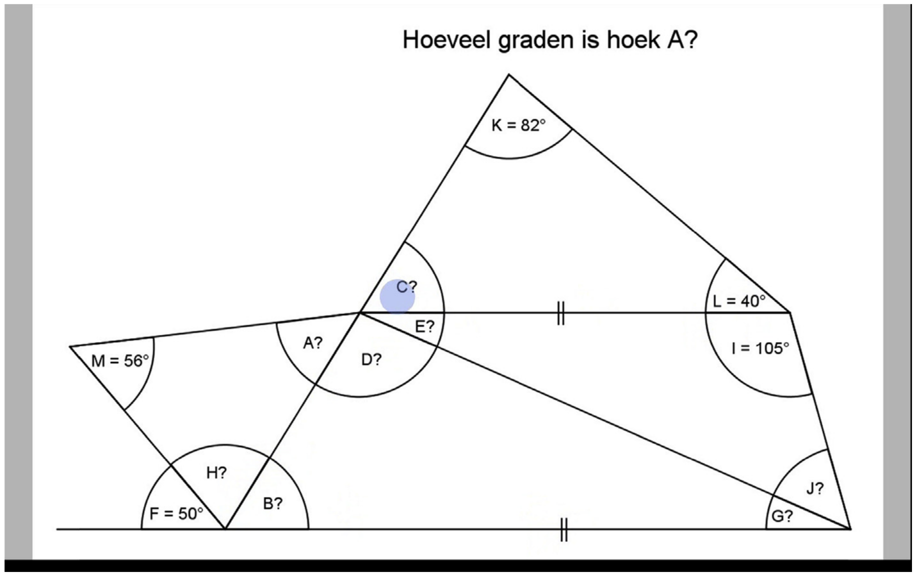
Moreover, ambiguity resulting from the interplay between the verbal description and visual stimuli can affect eye movements, resulting in a lower percentage of fixations on a described object and in more search behavior. For example, in a study by Louwerse and Bangerter (2010) participants heard ambiguous descriptions of cartoon faces in a 4×3 grid and found that the more ambiguous descriptions (i.e., containing less specific information about the location of the cartoon face) not only resulted in less fixations on the described cartoon face, but participants were also slower in fixating the described face. In addition, studies using the visual world paradigm, in which participants are presented with an image depicting several distinct objects, have shown that participants attend to the objects that they hear being described in a sentence. When hearing a verbal description that is ambiguous with respect to which one of two objects is meant, viewing behavior is affected (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; for a review see; Huettig, Rommers, & Meyer, 2011). For instance, in one study an image would display a bowl, an envelope, an envelope with a saltshaker on top, and a pencil (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Then participants heard the sentence *"Put the saltshaker on the envelope in the bowl"*. Because the display contained two pictures with envelopes it is temporally ambiguous where the saltshaker should be put until the last part of the sentence (i.e., *the bowl*) is heard, and eye tracking data revealed that this made participants fixate both pictures with envelopes. In contrast, if participants heard the sentence "*Put the saltshaker that's on the envelope in the bowl*" they did not fixate the envelope without the saltshaker. Thus, ambiguous descriptions lead to more visual search. These studies – albeit conducted in very different paradigms and not investigating learning-suggest that it is possible that ambiguous verbal explanations in an example, might result in the learner being too late in attending the relevant visual information shortly after being mention, which might hinder integration of the verbal and visual information, and thereby, learning. Under such conditions, guidance in the form of EMME could be expected to be effective in guiding attention and improving learning. Hence, for future research it would be interesting to address this issue by investigating how task ambiguity affects the effectiveness of EMME.

In conclusion, in line with prior research (Van Gog et al., 2009), we found no evidence that eye movement modeling examples would enhance learning of procedural problem-solving tasks compared to regular modeling examples showing only the model's actions. Future research should examine conditions that may affect whether displaying the model's eye movements is effective for learning, such as the model's interaction with the task or the ambiguity of the verbal explanations. This would contribute to the development of guidelines for when to use eye movement modeling examples.

Acknowledgements

This research was funded by a Vidi grant (# 452-11-006) from the Netherlands Organisation for Scientific Research awarded to Tamara van Gog. The authors would like to thank Susan Ravensbergen and Kirsten Versijde for assistance with data collection, and Charly Eielts for programming the tasks in Flash.

Appendix. Example of a transcript of the verbal explanation in a modeling example in Experiment 2.



Screenshot of a modeling example (EMME condition) in Experiment 2 with the blue dot representing the location of the model's gaze. The following verbal instruction was used during the modeling example (translated from Dutch): *"The question is, how many degrees is angle A? You start by searching for angle A. Angle A is part of a triangle. A triangle contains a total of 180 degrees. If two of the three angles are known in a triangle, you can calculate the third angle. You calculate the third angle by subtracting the two angles from 180 degrees. You cannot calculate angle A right now, because angle H is unknown. Angle H is part of a straight line. A straight line contains a total of 180 degrees. You can calculate the unknown angle in a straight line by subtracting all known angle from 180 degrees. However, besides angle H angle B is also unknown, so for now it is not possible to calculate angle H. You cannot calculate angle B directly, but it can be derived from angle C, because these are equal. This can be seen by the tilted equal sign, which indicates that the lines are parallel and thus have the same angle. Because of the parallel lines, you can derive by means of the corresponding angle principle that the angle B and C are equal. Angle C is unknown for now but can be calculated. Angle C equals 180 degrees minus the known angles, equals 58 degrees. Now angle C is known, you know that angle B, by means of the corresponding angle principle, also equals 58 degrees. With angle B known, you can now calculate angle H. Angle H equals 180 degrees minus the known angles, equals 72 degrees. With angle H known, you can now calculate angle A. Angle A equals 180 degrees minus the known angles, equals 52 degrees. So angle A is 52 degrees."*

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439. http://dx.doi.org/10.1006/jmla.1997.2558.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137*, 190–200. http://dx.doi.org/10.1016/j.actpsy.2010.09.009.

Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology, 21*, 811–820. http://dx.doi.org/10.1002/acp.1351.

Braaksma, M. A. H., Rijlaarsdam, G., & van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*, 405–415. http://dx.doi.org/10.1037/0022-0663.94.2.405.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29*, 1146–1152. http://dx.doi.org/10.3758/BF03206384.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213–220. http://dx.doi.org/10.1037/h0026256.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*, 453–459. http://dx.doi.org/10.3758/BF03193787.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction, 20*, 111–122. http://dx.doi.org/10.1016/j.learninstruc.2009.02.010.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409–436. http://dx.doi.org/10.1007/BF02143160.

Field, A. (2009). *Discovering statistics using SPSS*. London, England: Sage.

Fiorella, L., & Mayer, R. E. (2015). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/edu0000065. Advance online publication.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*, 462–466. http://dx.doi.org/10.1111/1467-9280.02454.

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.learninstruc.2013.05.001.

Groen, M., & Noyes, J. (2010). Solving problems: How can guidance concerning task-relevancy be provided? *Computers in Human Behavior, 26*, 1318–1326. http://dx.doi.org/10.1016/j.chb.2010.04.004.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25*, 172–190.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80–91. http://dx.doi.org/10.1016/j.chb.2014.09.013.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171. http://dx.doi.org/10.1016/j.actpsy.2010.11.003.

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., et al. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science, 40*, 813–827. http://dx.doi.org/10.1007/s11251-012-9218-5.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*, 146–154. http://dx.doi.org/10.1016/j.learninstruc.2009.02.019.

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. http://dx.doi.org/10.1016/j.learninstruc.2012.11.004.

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 247–262). New York: Cambridge University Press.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies, 65*, 911–930.

http://dx.doi.org/10.1016/j.ijhcs.2007.06.005.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. http://dx.doi.org/10.2307/2529310.
Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology, 64*, 649–656. http://dx.doi.org/10.1080/17470218.2011.558628.
Litchfield, D., Ball, L., Donovan, T., Manning, D., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*, 251–262. http://dx.doi.org/10.1037/a0020082.
Louwerse, M. M., & Bangerter, A. (2010). Effects of ambiguous gestures and language on the time course of reference resolution. *Cognitive Science, 34*, 1517–1529. http://dx.doi.org/10.1111/j.1551-6709.2010.01135.x.
Mason, L., Pluchino, P., & Tornatora, M. C. (2015a). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology, 41*, 172–187. http://dx.doi.org/10.1016/j.cedpsych.2015.01.004.
Mason, L., Pluchino, P., & Tornatora, M. C. (2015b). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*. http://dx.doi.org/10.1111/bjet.12271. Advance online publication.
McLaren, B. M., Van Gog, T., Ganoe, C., & Karabinos, M. (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.
Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*, 289–300. http://dx.doi.org/10.1016/j.ergon.2005.11.008.
Ouwehand, K., van Gog, T., & Paas, F. (2015). Effects of gestures on older adults' learning from video-based models. *Applied Cognitive Psychology, 29*, 115–128. http://dx.doi.org/10.1002/acp.3097.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. http://dx.doi.org/10.1111/cogs.12086.
Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045–1060. http://dx.doi.org/10.1207/s15516709cog0000_29.
Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science, 18*, 407–413. http://dx.doi.org/10.1111/j.1467-9280.2007.01914.x.
Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. *Proceedings of the 6th International Conference on Multimodal Interfaces*, 9–15. http://dx.doi.org/10.1145/1027933.1027936.
Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review, 14*, 663–668. http://dx.doi.org/10.3758/BF03196818.
Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo, & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). New York, NY: Springer.
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. http://dx.doi.org/10.1016/j.chb.2009.02.007.
Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205–221. http://dx.doi.org/10.1002/acp.1112.
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 and Education, 72*, 323–327. http://dx.doi.org/10.1016/j.compedu.2013.12.004.
Wolff, C. E., Jarodzka, H., Van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*, 243–265. http://dx.doi.org/10.1007/s11251-016-9367-z.