Looking Through the Teacher's Eyes: Effects of Eye Movement Modeling Examples on Learning to Solve Procedural Problems

Tim van Marlen

Leden beoordelingscommissie:

Dr. Ignace Hooge
Prof. dr. Marieke van der Schaaf
Prof. dr. Liesbeth Kester
Prof. dr. Saskia Brand-Gruwel
Prof. dr. Katharina Scheiter

© 2019 T.V.A. van Marlen

Cover design by Ridderprint
Layout by Tim van Marlen
Printed by Ridderprint BV., Ridderkerk

ISBN: 978-90-393-7124-4

The research presented in this dissertation was funded by a Vidi grant (# 452-11-006) from the Netherlands Organization for Scientific Research (NWO).

All rights reserved. No part of this dissertation may be reproduced or transmitted in any form, by any means, electronic or mechanical, without the prior permission of the author, or where appropriate, of the publisher of the articles.

Looking Through the Teacher's Eyes: Effects of Eye Movement Modeling Examples on Learning to Solve Procedural Problems

Kijkend door de ogen van de leraar: Effecten van oogbewegingsmodelvoorbeelden op het leren oplossen van procedurele probleemoplostaken
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties in het openbaar te
verdedigen op donderdag 16 mei 2019 des middags te 12.45 uur

door

Tim Viktor Alexander van Marlen

geboren op 6 januari 1988
te Rotterdam

promotor:
Prof. dr. T.A.J.M. Van Gog

copromotoren:
Dr. M.M. Van Wermeskerken
Dr. H.M. Jarodzka

Contents

Chapter 1
Introduction

Imagine you want to edit a digital photo and you have to use unfamiliar computer software to do so. You look around in the program and are overwhelmed by all sorts of dropdown menu's and options. After having spent some time trying to edit your photo without success, you will probably look for a guide to help you. A quick online search yields many different YouTube tutorial videos, in which you see a screen recording of an expert's screen with the expert showing and verbally explaining (in a voice over) to you how to use the computer program.

The use of such video tutorials, also known as video modeling examples, in which the model (e.g., a teacher or expert) demonstrates how to perform a task, is becoming more popular as indicated by the massive amount of tutorial videos on YouTube and the popularity of learning platforms like KhanAcademy (which reached over 60 million registered users and 12 million monthly users in 2017; www.khanacademyannualreport.org). The popularity of video modeling examples is not surprising, as example-based learning is a very natural and very effective way of acquiring new knowledge and skills (Renkl, 2014; Van Gog & Rummel, 2010). With technological advancements the creation of video modeling examples has become much easier, and it is nowadays also possible to create video modeling examples of a model working on digital content, by capturing the model's actions on the computer via a screen recording, along with a voice-over of the model's verbal explanations of what s/he is doing. While there is a lot of research showing that video modeling examples are an effective tool to foster learning, the design of the examples is a crucial factor in their effectiveness (Van Gog, Rummel, & Renkl, in press), and often there is still room for improvement.

For instance, in most video examples the information is transient, meaning that the visual information (e.g., the screen content of a computer program) and the verbal information (e.g., verbal explanation of the model) are only temporarily available. The transience of information can pose a risk for learners, in the sense that when they do not attend to the right visual information at the right time (e.g., when it is mentioned by the model), they will miss out on certain information, which may hamper their learning (Ayres & Paas, 2007). Learning from video examples requires the selection, organization, and integration of the learning content (Mayer, 2014). These learning processes are required to process the visual information (i.e., on screen learning content with the visible actions of the model), and the visual-verbal information (i.e., on screen learning content with the

model's verbal explanation). Disruptions in any of these processes may hamper learning.

To continue with the example about photo editing, when viewing a tutorial video about editing a photo with an unfamiliar program, there are all sorts of unknown menu's and buttons, so the learner has to process a large amount of visual information. Given that the learner is a novice, s/he might find it hard to understand the verbal explanation because s/he is having trouble to select (i.e., attend to) the information the model is referring to (e.g., an editing tool like 'blur'). The difficulty in selecting the correct information might be due to several reasons. For instance, due to the competition of similar looking menu's and buttons (i.e., high visual complexity), the learner might not be able to spot the correct button the model is referring to. Another reason might be that the learner does not know exactly what the model is referring to; the verbal explanation might be ambiguous due to the use of technical terms by the model which the learner with limited expertise is not yet familiar with. Lastly, the verbal explanation of the model might be ambiguous because the term used by the model could be referring to multiple possible buttons in the (visual) context. In such situations, the learner will probably fail to understand and learn from the video example.

This could have been prevented by showing the learner what the model who was demonstrating the use of the photo editing software, was looking at. This would have aligned the learners' attention with the model's attention, and by guiding learners' attention to the right visual information at the right time, the selection of visual information and integration of visual and verbal information would be facilitated (Jarodzka et al., 2013; Richardson & Dale, 2005; Van Gog et al., 2009). Video modeling examples that offer such attentional guidance by displaying the model's eye movements are known as eye movement modeling examples (EMME). In EMME, a visualization of the model's eye movements, for instance in the form of a colored circle or dot is overlaid on the screen-recording, showing the learner where the model was fixating (i.e., what was at the center of attention) at any given moment (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). See Figure 1 for an example screenshot of an EMME.

A first study by Van Gog et al. (2009), comparing the effects of learning from EMME with learning from regular video modeling

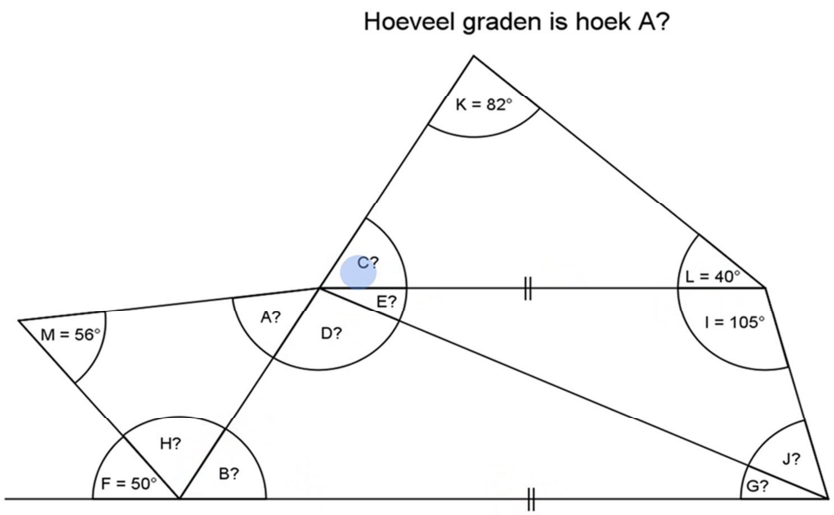


Figure 1. Screenshot of an eye movement modeling example with the blue dot representing the location of the model's gaze

examples (i.e., the same video examples but without the visualization of the model's eye movements), found no positive effect of EMME on learning to solve a procedural puzzle problem. Since then, however, other studies on learning perceptual classification tasks or learning reading strategies aimed at fostering text and picture integration have shown positive effects of EMME (e.g., Jarodzka et al., 2012; 2013; Mason, Pluchino, & Tornatora, 2015; 2016; Mason, Scheiter, & Tornatora, 2017; Salmerón & Llorens, 2018; Scheiter, Schubert, & Schüler, 2018; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012). Considering the positive findings of these other studies, as well as the fact that students have to acquire procedural problem-solving skills in many subjects in education (e.g., science or math problems), it is important to examine whether and under which conditions EMME can also be used to foster learning of procedural tasks.

Therefore, the central question addressed in this dissertation is: Do EMME enhance learning of procedural problem-solving tasks? In addition, the research presented in this dissertation made a first step towards shedding light on several possible boundary conditions that could affect learning procedural problem-solving tasks from EMME, for instance by using tasks of different complexity levels, verbal explanations that are clear versus ambiguous, and by studying effects of EMME on learners of different ability levels.

Eye Movement Modeling Examples (EMME)

By visualizing where the model is looking during task performance, EMME guide the learner's visual attention to the visual information that the model is processing at that moment. By synchronizing the learner's attention with that of the model a state of 'joint attention' is created. Joint attention is the phenomenon of automatically attending to an object someone else is attending to (Brennan, Chen, Dickinson, Neider & Zelinsky, 2008; Frischen, Bayliss, & Tipper, 2007). The attentional guidance offered by EMME can be especially useful for novice learners as research has shown that novices compared to experts fixate relevant information less often and also need more time to start fixating this information (e.g., Charness, Reingold, Pomplun, & Stampe, 2001; Van Gog, Paas, & Van Merriënboer, 2005; Van Meeuwen, Jarodzka, Brand-Gruwel, Kirschner, De Bock, & Van Merriënboer, 2014; Wolff, Jarodzka, Van den Bogert, & Boshuizen, 2016). Creating a state of joint attention between the (novice) learner and the (expert) model can then help learners to attend and select the relevant visual information (i.e., on screen learning content with the visible actions of the model), and the visual-verbal information (i.e., on screen learning content with the model's verbal explanation), thereby enhancing learning.

Thus, the first prediction would be that EMME reduce differences in attention allocation between the (novice) learners and (expert) model, which should be reflected in eye tracking data indicating that EMME compared to regular video examples guide the learner's attention faster and more often towards the relevant information. This, in turn, would be expected to lead to better learning outcomes. Previous research on EMME has indeed shown that EMME guide the learner's visual attention and enhance learning (Jarodzka et al., 2012; 2013; Mason et al., 2015; 2016; 2017; Scheiter et al., 2018). For instance, the studies using EMME to enhance learning to perform classification tasks found that the learner's eye movements were more similar to those of the model compared to regular video examples as evidenced in a higher scanpath similarity (Jarodzka et al., 2013) or smaller Euclidean distances between the model's gaze position and that of the learner (Jarodzka et al., 2012). Other studies focusing on text-picture integration used other eye tracking measures like the number of transitions between text and pictures and the summed fixation times on the text and picture to indicate that EMME guide the

learner's attention (Mason et al., 2015; 2016; 2017; Scheiter et al., 2018).

Whereas the former studies measured the learner's eye movements both during and after watching the EMME, other studies used eye tracking to examine whether EMME changed the learner's attention allocation only after watching EMME. For instance, one study found that EMME enhanced learning to classify histological slides and also showed that learners in the EMME condition needed fewer fixations to inspect the histological slides (Vitak et al., 2012). Similarly, a study regarding classifying medical images showed that learners in the EMME condition had higher performance and also made more task relevant fixations after having watched EMME compared to the control condition (Gegenfurtner, Lehtinen, Jarodza, & Säljö, 2017).

Finally, there are also studies that did not include eye tracking to measure changes in learners' attention allocation but focused primarily on whether EMME enhance learning (Salmerón & Llorens, 2018; Van Gog et al., 2009). The study by Salmerón et al. (2018) demonstrated that EMME enhance learning of reading strategies of digital hyperlinked texts. In contrast, the study by Van Gog et al. (2009) did not find beneficial effects on learning outcomes for learning to solve a procedural puzzle problem.

In sum, EMME seem successful at guiding learners' attention and often this guidance also improves their learning. That learning is not always enhanced as a consequence of the attention guidance provided by EMME, mirrors research on (other forms of) cueing (i.e., visually highlighting relevant elements in the learning material by means of visual cues like color coding or pointing arrows; Van Gog, 2014). The use of visual cues is typically found to guide the learner's attention, but this does not always necessarily result in higher learning outcomes (De Koning, Tabbers, Rikers, & Paas, 2010; Kriz & Hegarty, 2007). A recent meta-analysis showed that visual cueing might be especially effective for learners with limited prior knowledge of the learning content (Richter, Scheiter & Eitel, 2016). Due to the limited prior knowledge learners may have difficulty selecting and integrating the multimedia learning content without the support of visual cues, whereas learners with high prior knowledge –even if they find the relevant information faster with cues- do not seem to need the cues as their prior knowledge enables them to (timely) process and integrate the multimedia content. Considering that in EMME the visualization of the model's eye movements can be considered a form of visual

cueing, it is possible that the effect of prior knowledge might also apply to EMME.

Functions of EMME in Learning to Solve Procedural Problems

The overview presented in the previous section shows that the majority of EMME studies reached similar conclusions stating that EMME affect the learner's attention allocation and enhance learning. However, when looking at what function EMME serve in relation with the learning task the studies are less similar. As stated earlier, one function of EMME is the creation of joint attention between the model and the learner to help guide attention towards the relevant information. Examples of studies in which the main role of EMME was to establish joint attention, were studies aimed at fostering learning in classification tasks: classifying locomotion patterns of fish (Jarodzka et al., 2013), classifying epileptic seizure symptoms in infants (Jarodzka et al., 2012), and classifying cells in histological slides (Vitak et al., 2012). For instance, in the study by Jarodzka et al. (2013), students watched video examples in which an expert demonstrated and verbally explained how to classify locomotion patterns of fish. Students either watched EMME or regular video examples (i.e., without the model's eye movements superimposed). Results revealed that students' attention was guided more towards the relevant information in the EMME conditions compared to the regular video example condition, and that students in the EMME condition had higher learning outcomes on later classification tasks.

Besides using EMME to establish joint attention between the learner and the model, another function of EMME is the ability to illustrate perceptual and cognitive strategies of the model that might otherwise not be observable. An example of a perceptual strategy is the visualization of a person's visual search strategy, and there is evidence that observing a visualization of a model's eye movements improves performance on a visual search task (Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Nalanagula, Greenstein, & Gramopadhye, 2006; Stein & Brennan, 2004). For instance, the visualization of the model's eye movements improved performance in visual search tasks in which people searched for errors on printed circuit boards (Nalanagula et al., 2006), errors in software code (Stein & Brennan, 2004), or lung-nodules on X-ray scans (Litchfield et al., 2010). In fact, Litchfield et al. (2010) showed that it did not matter whether the participants were shown eye movements of an expert or a

naïve searcher as long as the eye movements were task related (i.e., depicting searching for lung-nodules).

Other studies used the visualization of the model's eye movements to convey a cognitive strategy to either improve current task performance on an insight problem-solving task (Litchfield & Ball, 2011) or to enhance study strategies (i.e., integrating text and picture information: Mason et al., 2015; 2016; 2017; Scheiter et al., 2018; navigating digital hyperlinked texts: Salmerón & Llorens, 2018). More specifically, in the studies by Mason et al. (2015; 2016) 7th grade students watched EMME in which the model demonstrated how to integrate a text with a supporting picture, without any verbal explanation. The students watched how the model made deliberate transitions between key concepts in the text and the corresponding elements in the picture, hereby signaling that these information sources needed to be integrated. Results revealed that the EMME condition not only affected how the students read a subsequent text (as indexed by eye-tracking the students) but also that students in the EMME condition showed better text comprehension compared to students in a control condition that did not include a modeling example.

Both of these functions of EMME (joint attention and perceptual/cognitive strategy visualization) would be relevant when learning to solve procedural problems. However, Van Gog et al. (2009) found a negative effect of EMME on learning to solve a procedural puzzle problem. In this study, learners watched a video modeling example about solving a procedural puzzle with or without the model's eye movements superimposed (which illustrated the consideration of the different possible actions at each step) and with or without the model's verbal explanation (which explained the different possible actions at each step). Subsequently, learners attempted to solve the puzzle problem themselves. The results revealed that on the transfer problem, learners of the EMME condition with verbal instructions performed worse than learners of the EMME condition without verbal instructions. One possible explanation given by the authors was that perhaps the verbal instruction was already sufficient for the learners to guide the attention to the object the model was referring to, thereby making the visualization of the model's eye movements redundant. This suggests that the model's verbalizations might play an important role in learning from EMME. EMME might not be effective (or even harmful) when the model's eye movements are

redundant with the model's explanations, and might be most effective when the model's eye movements disambiguate or complement the model's explanations.

Besides the role of verbalizations in EMME, the model's visible interaction with the learning material might also have played a role in the study by Van Gog et al. (2009). That is, the model was demonstrating how to solve a procedural puzzle problem, meaning that at every step the model would interact with the task (i.e., clicking on a puzzle piece to move it to another location). Considering that at every step there were only two or three options to choose from, it is possible that seeing the steps the model took was already sufficient to grasp the underlying strategy involved in the task, in other words, that being able to observe the choice process (through the model's eye movements) was less important for learning than observing the outcome of that process. Thus, the effectiveness of learning with EMME might depend on what the eye movements illustrate in relation to the task being demonstrated.

Research Question and Overview of the Dissertation

Considering the fact that EMME have proven effective for several different kinds of tasks, and that the lack of evidence on effectiveness of EMME for learning procedural problem-solving tasks was limited to a single study, in which there were other factors that might have affected the results (Van Gog et al., 2009), it would be premature to conclude that EMME are not suitable to be used to enhance learning in procedural tasks. Therefore, the central question addressed in this dissertation is whether EMME guide the learner's attention and enhance learning of procedural problem-solving tasks.

Five empirical studies, with a total of eight experiments, were conducted to answer the main research question: *Do EMME enhance learning of procedural problem-solving tasks?* Together, these studies also make a first step towards shedding light on several possible boundary conditions that could affect learning procedural problem-solving tasks from EMME, by using tasks of different complexity levels, tasks in which the visualizations of the model's eye movements have a different function, tasks in which the model's verbal explanations are clear versus ambiguous, and by studying effects of EMME with learners of different ability levels.

In the study reported in **Chapter 2**, it was investigated whether EMME affected attention and enhanced learning how to solve

a procedural geometry problem-solving task. In Experiment 1, university students watched video examples in which a model demonstrated how to solve single-step geometry problems, either without the model's eye movements, with the model's eye movements visualized, or with a random pattern of eye movements (this last condition was added to examine whether we could conceptually replicate the findings of Litchfield et al. 2010, showing that only task relevant eye movements improved concurrent task performance, to rule out that the effect occurred because the mere presence of visualized eye movements itself would make students more attentive). Experiment 2 used more complex four-step geometry problems and given the complexity of the problems, the video examples included the model's verbal explanations. Students were either shown EMME or regular video examples (i.e., without the model's eye movements superimposed). In addition, the student's eye movements were recorded while watching the video examples, to investigate how EMME affected attention. Learning outcomes were assessed with isomorphic problem-solving tasks and transfer problems.

A fundamental proof of concept study on effects of verbal ambiguity and visual complexity on visual search (i.e., selection of information) was conducted, reported in **Chapter 3**. In two experiments, this study examined whether visual search would be affected by verbal ambiguity and visual complexity. The verbal descriptions indicated what had to be searched on the visual search images. Visual complexity was taken into account as it not only affects visual search by itself but might also interact with the verbal ambiguity of a task explanation (e.g., verbal ambiguity might not be as much of a problem when visual complexity is low). A 2 (verbal ambiguity: high vs. low) x 2 (visual complexity: high vs. low) within-subjects design was used in which both the verbal ambiguity of the description and visual complexity of the search image was manipulated. It was expected that both high verbal ambiguity and high visual complexity would result in slower and less accurate visual search with a possible additive effect of both factors. In case verbal ambiguity would indeed negatively affect visual search, this would provide indirect evidence that in situations in which verbal explanations are ambiguous for the learners, EMME might be helpful to disambiguate the model's explanation.

The study reported in **Chapter 4** examined whether the ambiguity of the model's verbalizations would affect learning from EMME. Two experiments were conducted using the same geometry

task as in Experiment 2 of Chapter 2, but with a more ambiguous verbal explanation. In Experiment 1, university students either watched regular video examples (i.e., without the model's eye movements visualized) or EMME. In both conditions the model's verbalizations contained ambiguous language due to the use of deictic terms and sentences like "now you know this angle, you can calculate the other angle". Students' eye movements were recorded while they watched the video examples, and learning outcomes were assessed by means of isomorphic and transfer problems. As research on cueing has shown that attentional guidance is most effective for learners with only limited prior knowledge (Richter et al., 2016). Therefore, participants in Experiment 2 were secondary education students. Experiment 2 had a 2 (modeling example: EMME vs. regular video example) x 2 (verbal explanation: ambiguous vs. unambiguous) between-subjects design. Because this study was conducted in regular classrooms, eye tracking could not be applied to measure students' attention, and only their learning outcomes were measured.

Because research on gaze following behavior (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Gobel, Tufft, & Richardson, 2018) as well as research on learning from modeling examples (Braaksma, Rijlaarsdam, & van den Bergh, 2002; Sonnenschein & Whitehurst, 1980) suggested that attention and learning could be affected by the model's (perceived) social status, the study reported in **Chapter 5** examined whether the purported expertise of the model would affect learning from EMME. Secondary education students watched EMME with unambiguous verbal explanations in which the model was explaining and demonstrating how to solve geometry problems (same as used in Chapters 2 and 4). The purported expertise of the model in the EMME was manipulated by means of a cover story preceding the EMME. In one condition, the model was introduced as a math teacher confident in her ability to explain the geometry problem (Expert condition) in the other condition as a Dutch teacher insecure about her ability to explain the geometry problem (No-Expert condition). Subsequently, students in both conditions watched the exact same EMME (i.e., only the purported expertise was different).

The study described in **Chapter 6** used a problem-solving task that capitalized more on the perceptual and cognitive strategy-conveying function of EMME and addressed the effects of EMME in the presence or absence of the model's (unambiguous) verbal

explanations. An experiment was conducted that was a conceptual replication of the study by Van Gog et al. (2009) described above. In a 2 (modeling example: EMME vs regular modeling example) x 2 (verbal explanation: present vs. absent) between-subject design, secondary education students watched video modeling examples demonstrating how to solve deductive reasoning problems, and their learning outcomes were assessed.

Finally, **Chapter 7** presents a summary and general discussion of the main findings of this dissertation.

Chapter 2

Showing a Model's Eye Movements in Examples Does not Improve Learning of Problem-Solving Tasks

This chapter was published as:
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. doi.org/10.1016/j.chb.2016.08.041

Acknowledgement of author contributions: TM, MW, HJ, and TG designed the experiments, TM recruited participants and collected the data, TM analyzed the data, MW checked the data package, TM drafted the manuscript, all authors contributed to critical revision of the manuscript, MW, HJ, and TG supervised the experiments.

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.

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, Pluchino, & Tornatora, 2015; 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; Jarodzka et al., 2013; Van Gog et al., 2009).

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.

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., 2015). 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, 2016). 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; Jarodzka et al., 2013) and 'strategy' (Mason et al., 2015; Mason et al., 2016) 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.

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.

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.

Methods

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). All participants had normal or corrected to normal vision. Participants received either a monetary reward or course credit for participating.

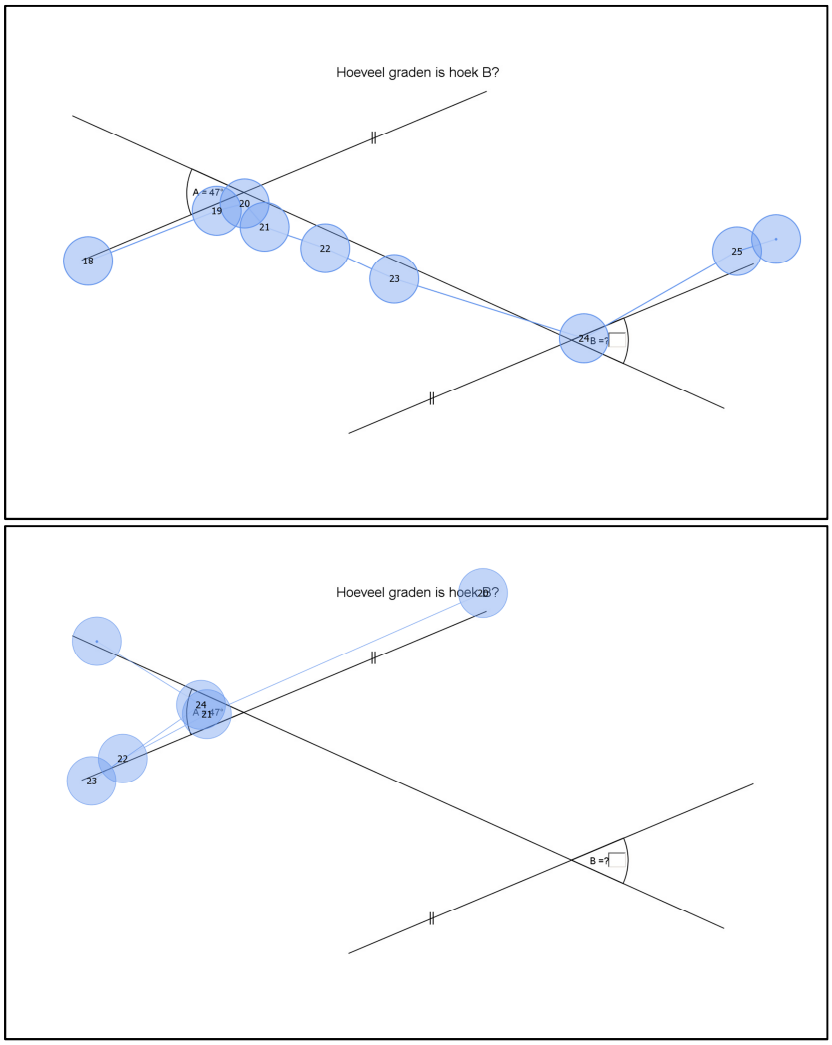


Figure 1. Static representation of the dynamic scan path shown in a meaningful EMME (left) and a meaningless EMME (right) in Experiment 1 for the alternating angle problem. On top is the problem statement stating (translated from Dutch) “How many degrees is angle B?”. The blue dots represent the location of the model’s gaze and the numbers represent the order of the fixations number.

Materials and Apparatus

Eye tracking equipment. Eye movements were recorded with a SMI RED250 eye tracker with a sampling rate of 250Hz (SensoMotoric Instruments, GmbH). The experiment was created in

SMI *Experiment Center 3.34* software and presented on a monitor with 1680 x 1050 pixels resolution with a refresh rate of 60 Hz.

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?").

Geometry problems. Four types of geometry problems were created: triangle problem, straight line problem, alternating angle problem (i.e., Z-rule; see Figure 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 Figure 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.

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 Figure 1). In the meaningless EMME condition, the eye movement modeling example focused on all features of the problem and on regions between the

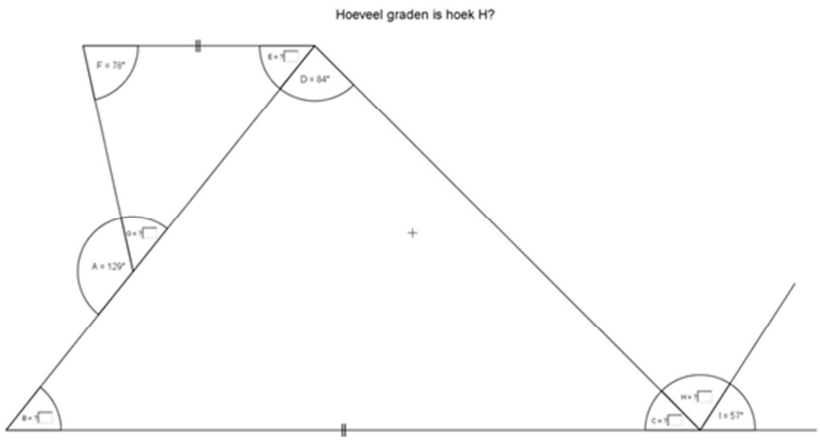


Figure 2. Example of a transfer problem in Experiment 1 with the problem statement (translated from Dutch) stating “How many degrees is angle H?”.

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 Figure 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 seconds. 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.

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 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.

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 $\kappa\omega$ = .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 degrees from the correct answer; n = 19) but showed they knew how to solve

[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.

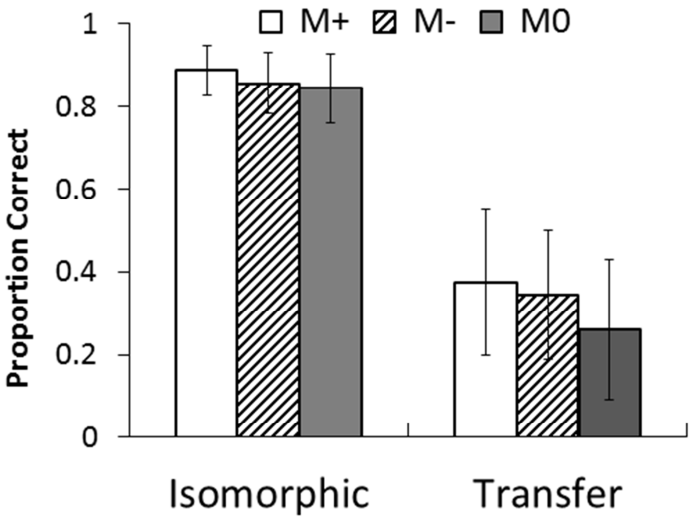


Figure 3. Proportion correct for the isomorphic and transfer problems in Experiment 1 for the meaningful eye movements modeling condition (M+), meaningless eye movements modeling condition (M-), and control condition (M0). Error bars represent the 95% confidence interval of the mean.

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.

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), $F(2, 69) = 1.20$, $p = .309$, 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), $F(2, 69) = .75$, $p = .475$.

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, $H(2) = .34$, $p = .842$, nor performance on the transfer problems, $H(2) = 1.35$, $p = .509$, differed significantly among the conditions (see Figure 3).

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, $F(2, 43.507) = 2.67$, $p = .081$, $\eta_p^2 = .10$ (see Figure 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 ($p = .127$), meaningful vs. control condition ($p = .836$), and meaningless vs. control ($p = .068$, $r = .32$). Despite not reaching statistical significance, this last, medium effect size, suggests that it took participants in the

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.

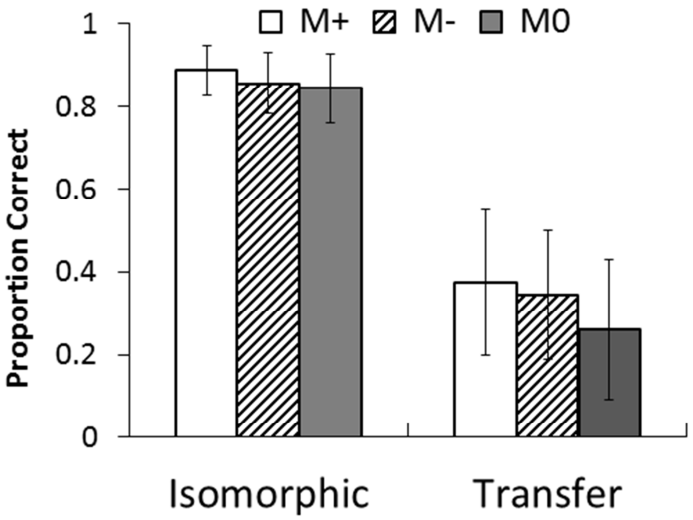


Figure 4. Response times expressed in seconds for correctly answered isomorphic and transfer items for the meaningful eye movements modeling condition (M+), meaningless eye movements modeling condition (M-), and control condition (M0). Error-bars represent the 95% confidence interval.

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, $H(2) = 5.16$, $p = .076$ (see Figure 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 ($p = .826$), meaningless vs. control ($p = .153$), and meaningful vs. control ($p = .038$, $r = .42$). 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.

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 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 et al., 2010; Litchfield & Ball, 2011) or during example study (Jarodzka et al., 2012; Jarodzka et al., 2013; Mason et al., 2015, 2016) 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.

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.

Methods

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.

Materials and Apparatus

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.

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.

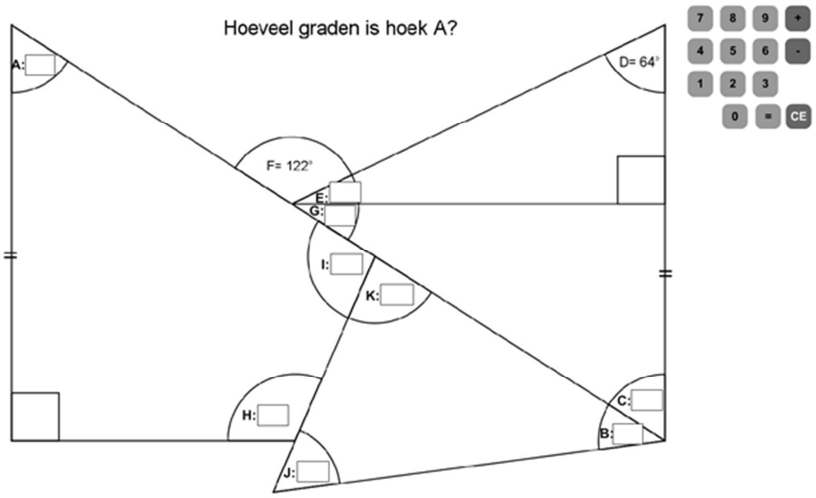


Figure 5. An example of a transfer problem in Experiment 2 including the calculator. On top is the problem statement, (translated from Dutch) “How many degrees is angle A?”.

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 Figure 5).

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).

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.

Data Analysis

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).

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

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

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 ≤ 40°/s and fixation duration ≥ 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 1500ms after the onset of the verbal referent were included in the analyses (cf. Dahan & Tanenhaus, 2005). In addition, fixations occurring within the first 100ms were excluded from the analysis, as research indicates that initiating eye movements based on language input takes approximately 100ms 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).

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

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)	128.15 (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)	434.62 (129.39)
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

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: $M = 3.77$; $SD = 1.21$), $t(60) = 0.162$, $p = .872$ (this did not change when excluding the outliers re. posttest performance: $M = 3.83$; $SD = 1.18$, $t(58) = -.22$, $p = .829$).

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, $U = 435.00$, $z = -.46$, $p = .643$, $r = .06$, nor performance on the transfer problems, $U = 448.50$, $z = -.02$, $p = .981$, $r = -.003$, differed significantly between conditions (see Figure 6).

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, $U = 345.00$, $z = -1.17$, $p = .243$, $r = -.15$. However, on correctly solved transfer problems, the response times were *higher* in

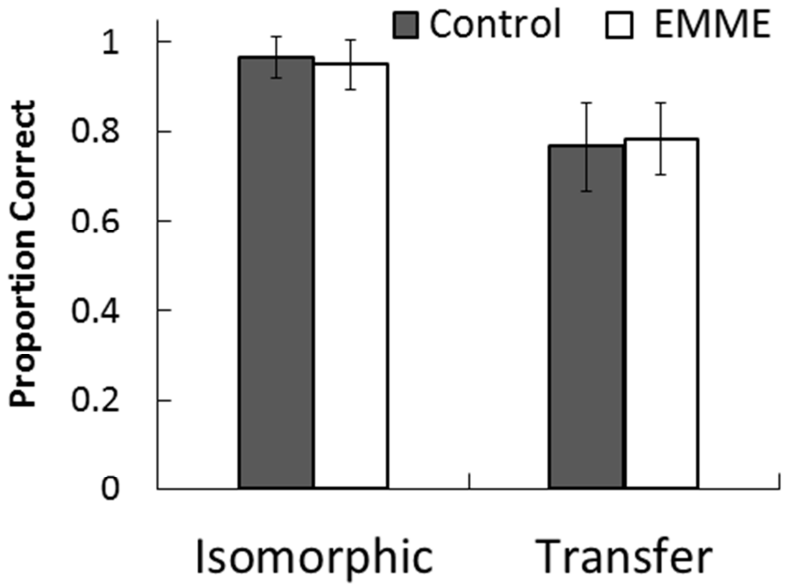


Figure 6. Proportion correct for the isomorphic and transfer problems in Experiment 2 for the EMME and control condition. Error bars represent the 95% confidence interval of the mean.

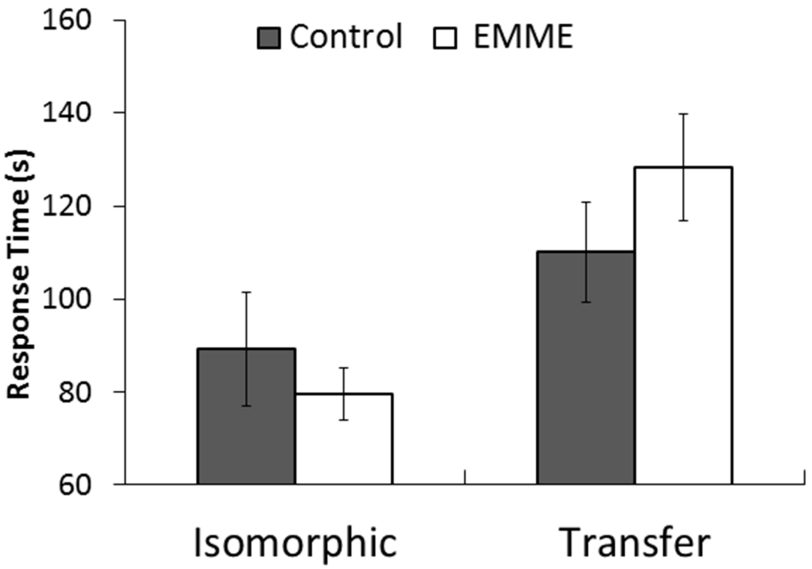


Figure 7. Response times (in seconds) for the isomorphic and transfer problems in Experiment 2 for the EMME and control condition. Error bars represent the 95% confidence interval of the mean.

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 = .007, r = -.35, meaning that participants in the EMME condition were *slower* at solving the geometry problems (see Figure 7).

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 1500ms after the onset of the referent), between the EMME (M = .43, SD = .13) and control condition (M = .46, SD = .18), t (57) = .64 p = .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 = .041, r = .27. Finally, there was a significant difference in fixation duration (in ms), t (55) = -2.41, p = .019, r = .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).

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 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.

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; Jarodzka et al., 2013) and text-picture integration strategies (Mason et al., 2015; 2016). 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; Jarodzka et al., 2013) and 'strategy' (Mason et al., 2015; 2016) 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., 2015a; 2016), 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.

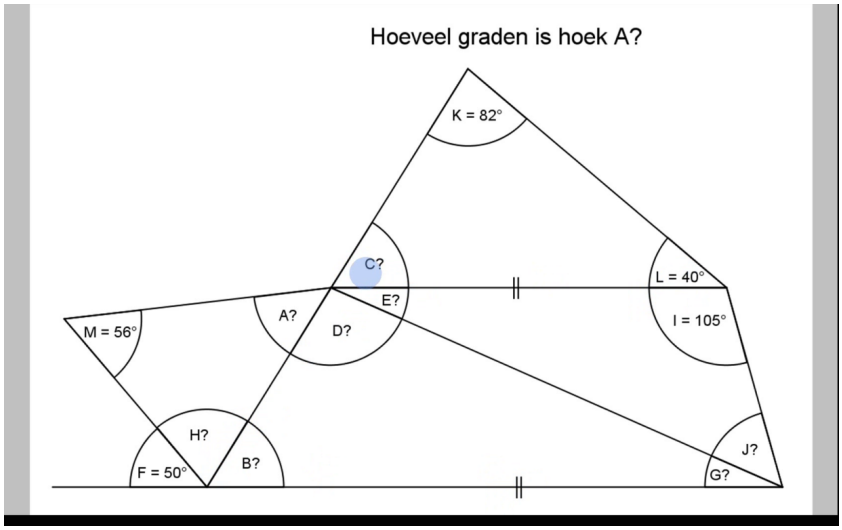
Eye movement research provides support for this idea that verbal information influences a listener's eye movements and that the coupling of eye movements between speaker and listener may affect memory. For instance, in one study by Richardson and Dale (2005), people listened to someone describing a video-clip 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 x 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.

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."*

Chapter 2

Chapter 3

Effects of Visual Complexity and Ambiguity of Verbal Instructions on Target Identification

This chapter is accepted for publication as:
Van Marlen, T., Van Wermeskerken., & Van Gog, T. (in press). Effects of visual complexity and ambiguity of verbal instructions on target identification. *Journal of Cognitive Psychology*. doi.org/10.1080/20445911.2018.1552700

Acknowledgement of author contributions: TM, MW, and TG designed the experiments, TM recruited participants and collected the data, TM analyzed the data, MW checked the data package, TM drafted the manuscript, all authors contributed to critical revision of the manuscript, MW and TG supervised the experiments.

Chapter 3

Abstract

Research has shown that visual complexity and the ambiguity of verbal information affect the speed and accuracy of locating targets during visual search. The higher the visual complexity and description ambiguity, the slower and poorer the target identification performance. Because these factors are seldom studied in combination (even though they regularly co-occur), it is unclear whether they would interact. Therefore, in two experiments, participants viewed images that displayed cartoon-like characters and had to correctly identify a character from a verbal description under conditions of low/high visual complexity and low/high description ambiguity (manipulated within-subjects). Results revealed that high ambiguity descriptions resulted in lower accuracy and slower response times. However, our manipulation of visual complexity did not affect performance or response times either in itself or in interaction with verbal ambiguity. Findings are discussed in terms of theoretical and practical implications, for instance, for multimedia learning.

Introduction

Research on visual search has been concerned with the question of how the visual complexity of the stimuli (Wolfe, 1994) or the ambiguity of the (verbal) task instructions (Louwerse & Bangerter 2010) affect the speed and accuracy of locating a target. However, these factors (i.e., visual complexity and verbal ambiguity) have primarily been studied in isolation, so whether and how they interact to influence visual search is unknown. The potential interaction of visual complexity and verbal ambiguity on visual search is important as it determines the specificity of the target template (i.e., the representation of the features of the target). A clear target template is not always readily available prior to commencing the search. For instance, search can be less effective if the verbal target description is ambiguous, leaving too many potential targets candidates, and the visual display is too complex to effortlessly spot the target.

The present study was designed to test the prediction that both high visual complexity and high description ambiguity would result in poorer target identification performance (lower accuracy and slower response times).

Visual Complexity and Specificity

The visual complexity of an image or task display can be an important characteristic in determining the amount of visual search required to locate and identify a target. The concept of visual complexity has many definitions, discussion of which goes beyond the scope of this article (for reviews see Donderi, 2006b; Forsythe, 2009). Although different definitions exist, most agree with the basic definition that visual complexity refers to the amount of detail present within an image. By visual complexity we mean the number of visual features present, the amount of color differences, and contrast differences. Besides this subjective definition, visual complexity can be quantified by looking at the file size of a compressed digital image (Chikhman, Bondarko, Danilova, Goluzina, & Shelepin, 2012; Donderi, 2006a; Marin & Leder, 2016). That is, the more visually complex an uncompressed image is, the more difficulty the compression algorithm has compressing the image, thus resulting in a larger image file and vice versa. Research has shown that visual search becomes less efficient (Wolfe, 1994) and less accurate (Davis, Shikano, Peterson, & Michel, 2003; Neider & Zelinsky, 2011; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002) with increasing visual complexity. For instance, Neider and Zelinsky (2011) demonstrated that when visual

complexity increases (i.e., clutteredness of an image), individuals became less efficient in searching for a target building embedded in a scene image. However, in such studies, the target is cued beforehand by a specific target image providing the individual with a specific target template to aid visual search, which is not always readily available in everyday life. Therefore, how visual complexity affects visual search might also depend on the specificity of the target template and this specificity might in turn also be affected by the verbal ambiguity.

Indeed there are many studies indicating that the specificity of a target template can influence the efficiency of visual search (Castelhano & Heaven, 2010; Hout &Goldinger, 2015; Malcolm & Henderson, 2009; Schmidt & Zelinksy, 2009; Yang & Zelinksy, 2009). In most of these studies, the target template is either a picture of the target or the target object name or category. Generally, it is found that visual search is most efficient when the target is cued with a picture (e.g., picture of a mug) compared to when it is cued with an object name (e.g., the text 'mug'), because an object name is less specific about the target than the picture (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009; Schmidt & Zelinksy, 2009). However, in most of these studies the visual complexity of the images that are to be searched for the target is either low (e.g., depicting objects with little visual details) or is not manipulated. An exception is the study of Godwin and colleagues (Godwin, Walenchok, Houpt, Hout, & Goldinger, 2015). In this study participants had to classify objects in a dual target search task (i.e., classifying two targets simultaneously) and were given a picture cue of the target before each block of trials. The targets were either simple images (i.e., Landolt C's) or complex images (i.e., teddy bears and butterflies). Contrary to what one might expect, it was found that participants were faster in rejecting the distractors in the case of high complexity as compared to low complexity images, possibly due to the larger number of features present in the complex images that could be used to identify the objects.

In sum, research indicates that visual search can be affected by the visual complexity of the stimuli and that target cue images facilitate visual search (due to the higher specificity of the target template) compared to object names as target cue search templates. Although some studies mentioned above used target templates cues consisting of object or category names, these textual target cues were relatively unambiguous. In the next section we discuss how

ambiguously perceived verbal task descriptions can affect visual attention.

Verbal Ambiguity

Besides the visual complexity of a task, the verbal ambiguity of a task description might also influence the amount of visual search required within a learning task. Support for the idea that the ambiguity of verbal descriptions influences the way people attend visual information comes from research using the visual world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; for a review see Huettig, Rommers, & Meyer, 2011). In this paradigm people see an image depicting several objects while hearing a verbal description that is ambiguous about some of the objects. For instance, in the study by Allopenna et al. (1998) people saw an image displaying a beaker, a beetle, a speaker, and a carriage and heard the word *beaker*. Eye tracking data revealed that just after the onset of the word "beaker" the probability of fixating the image of the beaker did not differ from fixating the images of beetle and speaker, but once the whole word "*beaker*" has been heard and the ambiguity is resolved the probability of fixating the beaker is higher than the probability of fixating the other objects. This finding suggests that the ambiguity of verbal information (in this case arising from the visual context) affects target identification.

Comparable results were found in a study by Louwerse and Bangerter (2010) using an adapted version of the visual world paradigm in which participants were shown images containing a grid of 3 x 4 cartoon faces accompanied by a verbal description about one of the cartoon faces. Half of the verbal descriptions contained an additional ambiguous location description (i.e., *John is in the middle*…), whereas the other half did not. In addition, in half of the trials there was an ambiguous gesture cue (e.g., a pointing hand) pointing at the correct row, while the other half did not. Note that the presence of both types of cues resolves the ambiguity regarding the target location. Both verbal cues and gesture cues guided attention (i.e., were better than no descriptions/cues for target identification), but their combination most improved target identification, as indicated by the higher percentage of fixations on the correct cartoon face for the conditions with the extra location cue(s). In this study, the ambiguity arose from the lack of specificity of the verbal description (which could be resolved by cues in the visual context).

Present Study

The literature discussed above seems to indicate that both conditions of high visual complexity and conditions of high verbal ambiguity negatively affect the accuracy and speed of target identification. Yet, to the best of our knowledge, the effects of visual complexity and verbal description ambiguity have not yet been investigated in a single study, which renders it difficult to draw conclusions on these factors' individual influence on visual search processes and their possible interaction. Therefore, we used a variation of the visual world paradigm (cf. Louwerse and Bangerter, 2010), to explore the effects of visual complexity and verbal description ambiguity on target identification performance and response times. Participants saw images depicting 12 cartoon characters on a 4x3 grid while hearing a verbal description of one of the characters, which they had to identify. Visual complexity (low vs. high) was defined as the number of visual features contained within a display. Verbal description ambiguity (low vs. high) was defined as the number of characters referred to by the first two cues presented in the verbal description. We predicted that both high visual complexity and high verbal description ambiguity would result in lower accuracy and slower response times on accurate trials than low visual complexity and low verbal description ambiguity. An open question is whether the effects would be additive, thus even more pronounced in the combined condition with high visual complexity and high verbal description ambiguity. To investigate this question two experiments were conducted: Experiment 1 was conducted online with Amazon's Mechanical Turk and Experiment 2 was conducted in the lab.

Experiment 1

Method

Participants and Design

Participants were 250 U.S.A. citizens recruited online with Amazon's Mechanical Turk and received $0.80 payment for participation. Ten participants were excluded due to incomplete data, resulting in 239 participants (M_{age} = 38.14, SD_{age} = 12.44; 146 females). A 2x2 factorial within-subjects design was used, with visual complexity (high vs. low) and verbal ambiguity (high vs. low) as factors.

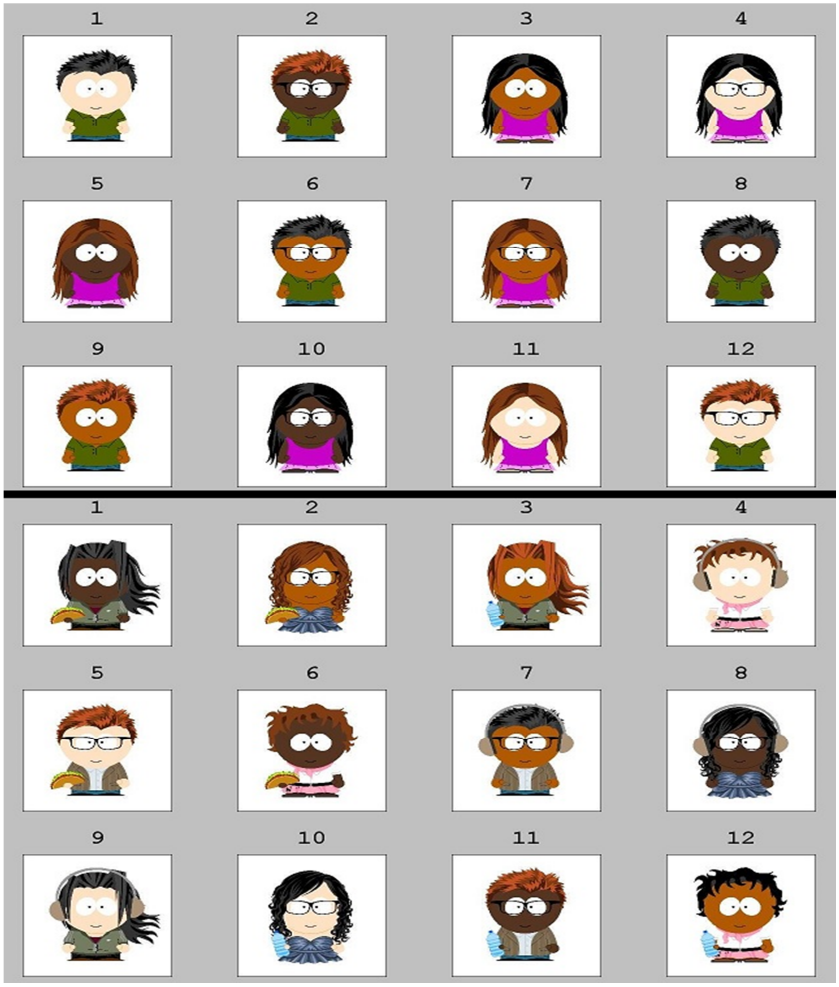


Figure 1. An example image displaying the cartoon characters of the low visual complexity condition (top) and the high visual complexity condition (bottom).

Materials

The experiment was run in Qualtrics software (www.qualtrics.com).

Target displays. Cartoon characters were created with an online avatar creator tool (www.sp-studio.de). The rectangular area covering a cartoon character was 230x230 pixels in size and the total image consisting of 4x3 cartoon characters was 1280x920 pixels. Thus, each display for a target identification trial featured 12 cartoon characters (six males/six females, with/without glasses, with

white/olive/dark-colored skin, and with blond/brown/black/red hair). The *low* and *high* visual complexity conditions differed in the number of variations amongst characters. The low-complexity condition featured characters with one female and one male outfit, one hair style, and two hair colors. The high-complexity condition featured characters with two female and two male outfits (with more visual details than the low-complexity condition), two hair styles, two hair colors, who additionally carried accessories (i.e., headphones, food, bottles). Above each character a number was shown (i.e., 1 – 12; see Figure 1). A play button for the audio was located underneath each image (this was done automatically by Qualtrics).

To verify with an objective measure[2] that the images used in the different conditions indeed differed in visual complexity a paired *t*-test was conducted with the file size of the JPEG images as dependent variable (cf. Chikhman et al., 2012; Donderi, 2006a; Marin & Leder, 2016). The average file size of the images in the low-complexity condition (M = 106.42 KB, SD= 0.67) was indeed lower than that of the images in the high-complexity condition (M = 117.17 KB, SD = 0.94), t (22) = 32.34, p < .001, r = .99.

Twenty-five displays were created: one practice, 12 low visual complexity and 12 high visual complexity displays. The 24 low/high complexity displays were used twice: once with a low verbal ambiguity description and once with a high verbal ambiguity description (i.e., 48 trials: 12 low complexity/low ambiguity, 12 low complexity/high ambiguity, 12 high complexity/low ambiguity, and 12 high complexity/high ambiguity).

For each trial, one of the twelve cartoon characters had to be identified (i.e., target). Within each of the four conditions (high/low complexity x high/low ambiguity), each character number position (1 to 12) was used as a target location only once.

[2] Upon request of a reviewer we additionally collected subjective complexity ratings. Thirty-four participants (M_{age} = 25.00, SD_{age} = 4.53) who did not participate in Experiment 1 or 2, were presented with the low and high visual complexity images one-by-one (in random order), and asked to rate on a 7-point scale (with 1 = 'very simple' and 7 = 'very complex') how visually complex they perceived the images to be. A paired *t*-test on the average ratings of the low (M = 2.93, SD = 0.95) vs. high (M = 4.22, SD = 1.04) complexity images revealed a large effect, t(33) = 7.40, p < .001, r = .78. Thus, both according to an objective (file size) and subjective measure, there was a difference in visual complexity between low and high complexity images.

Verbal descriptions. The verbal descriptions were spoken by a female voice and lasted on average 7.62 seconds (*SD* = 0.43). Each verbal description (e.g., "A person with brown hair, next to a person without glasses, has an olive colored skin.") contained three cues necessary to correctly identify the target character (e.g., hair color, skin color). The verbal descriptions were chosen in such a way that the number of cartoon characters that fit the description decreased along with each of the three cues of the description. In the low-ambiguity condition, after the first, second, and third cue, there were four, two, and one possible cartoon characters left who fit the description, respectively. In the high-ambiguity condition there were six, four, and one possible cartoon characters left who fit the description after the first, second, and, third cue, respectively.

Procedure

After participants agreed to take part in the study on Amazon's Mechanical Turk, they were redirected to the Qualtrics online survey. Participants were instructed that they would see displays featuring 12 cartoon characters, while hearing descriptions of one of the characters, and that their task was to correctly identify the described character and type in the correct character number as fast as possible. After a practice trial, participants received the 48 trials (12 per condition) in random order. After completion of the task, participants provided demographic data. The experiment lasted approximately 20 minutes.

Data Analysis

Within each condition, target identification accuracy was measured as the proportion of correctly identified targets. Average response times per condition on *correct* trials were calculated. 21 participants were excluded from analyses due to low performance on the task (i.e., > 3*SD* below average; *n* = 2) or unfeasible display resolution (i.e., participants had to scroll in order to *view* the full stimulus display, *n* = 19). Moreover, many participants reported that the cue "a person with a *light* colored skin" was often mistaken for the cue "a person with a *white* colored skin". To explore whether this confusion had affected the performance of the participants on these trials, a paired *t*-test on proportion correct was conducted between trials that contained the cue "a person with a light colored skin" and the remaining trials. This analysis confirmed that performance was lower on trials with the "light" colored skin cue (*M* = .53, *SE* = .02) as

Table 1. Mean (and SD) Proportion Correct Identifications and Response Times in Experiment 1 and in Experiment 2.

	Low Visual Complexity		High Visual Complexity	
	Low Verbal Ambiguity	High Verbal Ambiguity	Low Verbal Ambiguity	High Verbal Ambiguity
Experiment 1 (*N*=239)				
Proportion correct (*N*=218)	0.85 (0.21)	0.80 (0.22)	0.82 (0.20)	0.80 (0.21)
Response time (s) (*N*=185)	17.60 (5.35)	20.95 (6.87)	20.82 (9.13)	21.76 (7.75)
Experiment 2 (*N*=34)				
Proportion correct (*N*=33)	0.83 (0.21)	0.74 (0.18)	0.85 (0.16)	0.73 (0.16)
Response time (s) (*N*=29)	14.66 (2.38)	16.39 (2.60)	15.06 (2.04)	16.81 (4.38)

compared to the other trials (*M* = .82, *SE* = .01), *t*(217) = -15.66,001, *r* = .73. Therefore, trials containing this confusing cue (*n* = 16) were excluded from further analyses, leaving eight trials in each 33 additional participants were excluded from the analyses due to very long response times (i.e., > 3*SD* above average; *n* = 4), RT data not being logged (*n* = 1), unfeasible display resolution (i.e., participants had to scroll down to be able to *type* in their answer; *n* = 26) or having no correct responses within a condition (*n* = 2).

Results and Discussion

Data (Table 1) were analysed with 2 (visual complexity: low/high) x 2 (verbal ambiguity: low/high) repeated measures ANOVAs. On target identification accuracy, in contrast to our prediction this analysis revealed no main effect of visual complexity, *F* (1, 217) = 1.75, *p* < .188, η_p^2 = .01. However, in line with our prediction there was a main effect of verbal ambiguity, *F* (1, 217) = 15.77, *p* < .001, η_p^2 = .07, indicating that higher ambiguity resulted in less accurate performance, but only on the low-complexity trials as indicated by an interaction effect, *F* (1, 217) = 4.76, *p* < .030, η_p^2 = .02, with low-complexity/low-ambiguity leading to the highest accuracy, *t* (217) = 2.46, *p* = .015, *r* = .16.

In line with our predictions, the analyses regarding identification speed revealed main effects of visual complexity, *F* (1, 184) = 26.54, *p* < .001, η_p^2 = .13, and verbal description ambiguity, *F* (1, 184) = 28.57, *p* < .001, η_p^2 = .13, with high-complexity and high ambiguity resulting in slower response times. However, this effect was again qualified by an interaction effect, *F* (1, 184) = 8.31, *p* = .004, η_p^2 = .04, indicating that verbal ambiguity only affected response times on low-complexity trials, *t* (184) = 5.50, *p* < .001, *r* = .38, with low-complexity/low-ambiguity leading to the fastest response times.

In sum, the results of Experiment 1 revealed that participants were slower and less accurate in identifying the correct cartoon character for the high ambiguity trials when the visual complexity was low. Although the results are in line with previous research stating that the more specific the target template is (i.e., in this case the low ambiguity conditions) the better the performance, we did not find the expected interaction in which performance would deteriorate even further in high ambiguity trials that were also high in complexity. However, the results have to be interpreted with caution, as compared

to the other trials (M = .82, SE = .01), $t(217) = -15.66$, $p <$ condition. Regarding the response times analyses,

participants were able to replay the verbal descriptions. This could have affected the results on both speed and accuracy (e.g., the lack of effect of verbal ambiguity on high-complexity trials might be due to participants replaying the description). To ensure that participants would only hear the verbal description once, a second experiment was conducted in the lab.

Experiment 2

The set-up and materials in Experiment 2 were the same as in Experiment 1, except for three adjustments: First, participants were tested in a controlled lab environment instead of online. Second, participants were explicitly instructed to listen to the verbal descriptions only once (which was controlled for by the experimenter). Third, due to the confusion caused by the cue concerning skin color, the formulation of this cue was changed. With these changes the aim of Experiment 2 was similar to the aim of Experiment 1, but with the difference that the verbal description was presented only once. We predicted to see similar effects of visual complexity and task ambiguity as seen in Experiment 1 but with somewhat lower overall performance due to the relative increase of the difficulty of the task and shorter response times given that the verbal description was only presented once.

Methods

Participants

Thirty-four International Bachelor in Psychology students from a Dutch university participated for course credit (M_{age} = 19.88, SD_{age} = 1.15; 26 females).

Materials and Apparatus

The same experimental task as in Experiment 1 was used. However, due to the low performance on trials that contained the cue ‘a person with a light colored skin’, this formulation was changed into ‘a person with an olive colored skin’.

Procedure

Participants received the same instructions as in Experiment 1, but with the additional instruction to not replay the verbal description after it had played once. The experimenter was in the same room as the participant and checked whether the participants followed this instruction. The session lasted approximately 20 minutes.

Data analysis

As in Experiment 1, 2 (visual complexity: low/high) x 2 (verbal ambiguity: low/high) repeated measures ANOVAs were conducted on the proportion correct and the average response times of correct trials. One participant failed to follow the instruction to listen to the verbal description only once and was excluded from all analyses. In addition, four participants were excluded from the response times (but not accuracy) analysis due to technical issues. This left 33 participants in the accuracy and 29 in the response times analysis.

Results and Discussion

Data (Table 1) were analysed with 2 (visual complexity: low/high) x 2 (verbal ambiguity: low/high) repeated measures ANOVAs. On target identification accuracy, in contrast to our prediction this analysis did not reveal a main effect of visual complexity, $F(1, 32) < 1$, $p = 1.00$, $\eta_p^2 = .00$. However, in line with our prediction there was a large main effect of verbal ambiguity, $F(1, 32) = 32.65$, $p < .001$, $\eta_p^2 = .51$, indicating that participants had lower performance on high-ambiguity trials than on low-ambiguity trials. There was no interaction effect, $F(1, 32) < 1$, $p = .438$, $\eta_p^2 = .02$.

On the response times the same pattern of results was found. In contrast to our prediction there was no main effect of visual complexity, $F(1, 28) = 1.25$, $p = .273$, $\eta_p^2 = .04$, but a large main effect of verbal ambiguity, $F(1, 28) = 26.27$, $p < .001$, $\eta_p^2 = .48$, indicating that participants were slower to respond correctly on high-ambiguity trials than on low-ambiguity trials. There was no interaction effect, $F(1, 28) < 1$, $p = .983$, $\eta_p^2 < .01$. In sum, the results of Experiment 2 revealed that participants were slower and less accurate in identifying the correct cartoon character for the high ambiguity trials than for the low ambiguity trials. However, the visual complexity of the images did not seem to influence performance. These results are at odds with the visual complexity results of Experiment 1 which showed that participants were slower in identifying the correct cartoon character for high visual complexity trials. One possible explanation for the difference in results regarding visual complexity might be due to the fact that participants in Experiment 2 were instructed not to replay the audio description, whereas participants of Experiment 1 were able to replay the audio description of the target. Other potential explanations for a lack of effect of visual complexity are discussed below.

General Discussion

Research has shown that visual complexity and the ambiguity of verbal information affect the speed and accuracy of locating targets during visual search, but these factors are seldom studied in combination (even though they regularly co-occur, e.g. in multimedia learning materials). Therefore, the present study examined the potential interplay between the visual complexity of the stimuli and (verbal) instruction ambiguity on the performance of a visual search task (i.e., accuracy and speed). As predicted, higher verbal description ambiguity led to lower target identification accuracy and higher response times on accurate trials. This finding is in line with prior research in which verbal ambiguity resulted from non-specific location descriptions (Louwerse & Bangerter, 2010) or the visual contexts (Allopenna et al., 1998).

Contrary to our hypothesis, visual complexity did not affect accuracy or response times. The finding that visual complexity had little or no influence on target identification accuracy and response times, deviates from earlier research and the results of Experiment 1. For instance, visual search became less efficient (Henderson, Chanceaux, & Smith, 2009; Wolfe, 1994) and less accurate (Neider & Zelinsky, 2011; Wolfe et al., 2002) when the target stimulus was embedded in a visually more complex background.

Possibly, the lack of effect of visual complexity on target identification in Experiment 2 results from the presence/nature of our verbal instruction (which was absent in the studies of Wolfe, 1994; Wolfe et al., 2002). Another possible explanation lies in the definition of visual complexity in our study (i.e., number of features contained within a display). High complexity trials contained more features and details than low complexity trials, which resulted in higher digital image file sizes, confirming a visual complexity difference (i.e., the more visually complex an uncompressed image, the more difficult it is to compress, the larger the image file; Chikhman et al., 2012; Donderi, 2006a; Marin & Leder, 2016). Nevertheless, a potential limitation of the present study is that this manipulation may have been too weak to yield complexity effects, although the subjective complexity ratings were higher for the high complexity images. Despite the results of the complexity ratings, some studies have shown discrepancies between objective features of visual complexity and perceived visual complexity (Yoon et al., 2015). Similarly, in our study we cannot rule out that the attributes that were added in the high-complexity displays (e.g.,

bottles, headphones) might have actually made the search process easier, whereas the relative homogeneity of characters in the low-complexity displays might have made it harder than intended. This explanation is also in line with the results of a study in which participants saw images of objects on which they performed both a categorization task, requiring less differentiation of the objects, and an object decision task which required more differentiation of the objects (Gerlach & Marques, 2014). For the object decision task, visual complexity of the object images resulted in slower response times. Yet, for the object classification task, visual complexity of the object images resulted in faster response times, suggesting that visual complexity made the classification task easier rather than more difficult. In addition, Godwin et al. (2015) found that complex distractors were rejected faster than simple distractors, possibly due to the greater number of visual features available in the high complexity images to use to identify the target. Thus, although the high complexity stimuli were more complex as indicated by both an objective (file size) measure and subjective measure (complexity ratings), the lack of effects of the visual complexity manipulation in the current study might have been due to unintended effects of (dis)similarity of the cartoon images on the search process.

Regarding practical implications, our findings are relevant for the design of multimedia learning materials (e.g., instructional animations, videos). To facilitate learning from multimedia material, visual cues (e.g., pointing arrows or highlights) are often applied to guide the learner's attention towards the relevant information at the right time to enhance students' integration of the visual and verbal information (i.e., the signalling or cueing principle; Van Gog, 2014). Research findings regarding the effectiveness of these cues are mixed, however, with some finding beneficial effects of cueing (Jamet, 2014), whereas others do not (De Koning , Tabbers, Rikers, & Paas, 2011). Our findings suggest that verbal ambiguity may be a boundary condition for the effectiveness of cueing for learning. Although the present study used a task that is uncommon in multimedia learning environments, target identification based on verbal instructions is at the heart of many multimedia learning materials. The finding that high verbal ambiguity lowers target identification accuracy and increases response times, suggests that visual cueing will be most effective for learning under high ambiguity conditions. Under high-ambiguity conditions, learners might not locate the referred information timely

when there are no visual cues, which would hamper integration with the verbal description, and thereby, learning. High-ambiguity situations might arise when learners are confronted with new information and they lack the required prior knowledge to fully follow the multimedia learning materials even though visual cues are present. The latter ties in with the meta-analysis by Richter, Scheiter, and Eitel (2016), which showed that prior knowledge was a significant moderator of the cueing effect, with cueing being more helpful for low prior knowledge learners (for whom a verbal description would likely be more ambiguous). It would be interesting for future research to examine the potential role of verbal ambiguity in multimedia materials in combination with factors like visual complexity of the material and the prior knowledge to gain more knowledge on how to make these learning materials more effective.

In conclusion, verbal description ambiguity, but not visual complexity, affected performance on a target identification task. These findings suggest that the ambiguity of verbal instructions can have a detrimental influence on how visual information is processed. These findings have implications, for instance, for the design of multimedia learning materials.

Chapter 4

Effectiveness of Eye Movement Modeling Examples in problem solving: The role of verbal ambiguity and prior knowledge

This chapter was published as:
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. doi.org/10.1016/j.learninstruc.2018.07.005

Acknowledgement of author contributions: TM, MW, HJ, and TG designed the experiments, TM recruited participants and collected the data, TM analyzed the data, MW checked the data package, TM drafted the manuscript, all authors contributed to critical revision of the manuscript, MW, HJ, and TG supervised the experiments.

Abstract

Eye movement modeling examples (EMME) are video modeling examples with the model's eye movements superimposed. Thus far, EMME on problem-solving tasks seem to be effective for guiding students' attention, but this does not translate into higher learning outcomes. We therefore investigated the role of ambiguity of the verbal explanation and prior knowledge in the effectiveness of EMME on geometry problems. In Experiment 1, 57 university students observed EMME or regular video modeling examples (ME) with ambiguous verbal explanations. Eye-tracking data revealed that –as in prior research with unambiguous explanations- EMME successfully guided students' attention but did not improve test performance, possibly due to students' high prior knowledge. Therefore, Experiment 2, was conducted with 108 secondary education students who had less prior knowledge, using a 2 (EMME/ME) x 2 (ambiguous/unambiguous explanations) between-subjects design. Verbal ambiguity did not affect learning, but students in the EMME conditions outperformed those in the ME conditions.

Introduction

Video modeling examples in which a model demonstrates and explains how to perform a task (e.g., “how to” tutorial videos on YouTube), are widely used in formal and informal learning settings. Such videos lie in the tradition of example-based learning, which is an effective and efficient way of learning, provided that the examples are well-designed (Renkl, 2014; Van Gog & Rummel, 2010). It has been proposed that, depending on the task at hand, the design of screen-recording video examples could be further improved by showing learners what the model is looking at, by displaying the model’s eye movements (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). Displaying a visualization of the model’s eye movements (e.g., fixations represented as a circle or dot) is expected to guide learners’ attention to what the model is looking at in that moment, which should make it easier to understand and learn from the demonstration and verbal explanation. Several studies have found beneficial effects of such “eye movement modeling examples” (EMME) on attention guidance and found enhanced learning of classification tasks (Jarodzka et al., 2012; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013) and enhanced integration of text and pictures during reading (Mason, Pluchino, & Tornatora, 2015; 2016). Thus far, however, EMME on problem-solving tasks seem to be effective for guiding learners’ attention, but this does not translate into higher learning outcomes (i.e., higher performance on the test problems) compared to the no EMME control condition (Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016). This discrepancy between the studies regarding the effectiveness of EMME might be related to the extent to which the verbal explanation accompanying the EMME is clear (i.e., unambiguous) to the participants. For instance, studies using classification tasks in which EMME were found to be effective (Jarodzka et al., 2012; 2013), specific jargon was used; when learners do not yet know the jargon, this increases the usefulness of visual guidance. In contrast, in a study on learning to solve a puzzle problem, it was clear from the verbal explanation what object the model was looking at, and the visual guidance provided by EMME was not useful for learning (Van Gog et al., 2009). The present study addressed two potential explanations for this lack of effect of attention guidance on learning procedural problem-solving from EMME: ambiguity of the verbal explanation and prior knowledge.

Eye Movement Modeling Examples

Multimedia materials provide a combination of verbal and pictorial information, which according to the *dual-coding theory* (Clark & Paivio, 1991; Paivio, 1986) are processed in separate auditory and visual channels. According to the *Cognitive Theory of Multimedia Learning* (Mayer, 2014), learners first need to attend the relevant verbal and pictorial information (selection). After selecting the relevant verbal and pictorial information learners *organize* this information into coherent mental representations, and *integrate* the verbal and pictorial mental representations with each other and with available prior knowledge (Mayer, 2014). In dynamic learning materials like video modeling examples, one challenge for the selection of information lies in the transience of the material. If the learner does not attend to the right information at the right moment, it is no longer available for processing (i.e., organization and integration) and learning is hindered (Ayres & Paas, 2007). One reason why learners might not be able to attend to the right information at the right time in a video modeling example, is that it is likely that there is a discrepancy in what the expert model and the novice learner are attending to, and that the verbal explanation provided by the model may not be sufficiently clear to rapidly guide the learner's attention to what the expert is referring to.

The discrepancy between experts' and novices' attention allocation has been shown in different eye-tracking studies. Experts often attend to task-relevant information relatively longer and faster while paying less attention to task-irrelevant information than novices (Charness, Reingold, Pomplun, & Stampe, 2001; Van Gog, Paas, & Van Merriënboer, 2005; Van Meeuwen, Jarodzka, Brand-Gruwel, Kirschner, De Bock, & Van Merriënboer, 2014; Wolff, Jarodzka, Van den Bogert, & Boshuizen, 2016). This expertise effect has also been demonstrated within participants as a result of task experience (Blair, Watson, & Meier, 2009; Canham & Hegarty, 2010; Haider & Frensch, 1999; Hegarty, Canham, & Fabrikant, 2010). This difference in attention allocation might cause learners to miss the information the model is attending to, unless the model's verbal explanation would be sufficiently clear to rapidly guide learners' attention to the right information at the right time.

It has been proposed that one way to reduce the discrepancy between the model's and the learner's attention allocation would be to show the learner what the model is attending to, by displaying a

visualization of the model's eye movements (e.g., as a dot or circle; Van Gog et al., 2009). In such eye movement modeling examples (EMME), the learner is not only shown how the model is performing the task (by means of a screen recording of the model's computer screen), but also where the model was looking while performing the task. By showing the eye movements of the model the visual attention of the learners is guided and synchronized with the model to create a state of *joint attention* (i.e., *joint attention* is the phenomenon characterized as automatically looking at an object someone else is looking at; Brennan, Chen, Dickinson, Neider & Zelinsky, 2008; Frischen, Bayliss, & Tipper, 2007*),* thus helping the learners attend to the relevant information at the right time which, in turn, can be expected to improve learning.

However, findings regarding the effectiveness of EMME for learning are mixed. Whereas some studies have found beneficial effects on learning classification tasks (Jarodzka et al., 2012; 2013) and learning text and picture integration (Mason et al., 2015; 2016), EMME in which problem-solving tasks are demonstrated seem to be effective for guiding learners' attention (Van Marlen et al., 2016), but this does not translate into higher learning outcomes (Van Marlen et al., 2016; see also Van Gog et al., 2009). One possible reason for these mixed findings might lie in the extent to which the model's verbal explanation is sufficient to rapidly guide learners' attention to what the model is referring to. EMME might be most effective for learning when the model's verbalizations are ambiguous.

The Role of Verbal Ambiguity in the Effectiveness of EMME

When the verbal explanation in a modeling example contains ambiguous verbal referents, it will not be immediately clear to the learner what the model is referring to. Ambiguity of verbal referents can originate from different sources. For instance, experiments in cognitive science have shown that ambiguity can arise due to the visual context (e.g., multiple objects that the referent might refer to; Allopenna, Magnuson, & Tanenhaus, 1998; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995), the lack of specificity of the verbal information (e.g., referring to a target without verbal location descriptions; Louwerse & Bangerter, 2010), or a lack of prior knowledge (e.g., lack of background knowledge about a to be inspected and discussed painting; Richardson, Dale, & Kirkham, 2007; Experiment 2). When verbal referents are ambiguous for any of those

reasons, it will take listeners more time to locate the relevant (i.e., target) information, if they are able to locate it at all (Louwerse & Bangerter, 2010; Van Marlen, Van Wermeskerken, & Van Gog, in press)

These studies about the effects of verbal ambiguity on the speed and accuracy with which referents are located, suggest that the attention guidance provided by EMME might be most needed and most effective for learning when the model's verbal explanation is ambiguous for learners. Providing clear verbal explanations might not always be possible depending on the task and source of the ambiguity. For instance, in classification tasks providing a clear verbal description of a complex visual shape denoted by a jargon term, might be quite difficult. In the classroom, teachers/instructors can resolve this problem by using available non-verbal cues (e.g., looking or pointing at the part of the task they are discussing) that will disambiguate their verbal message. However, in digital video instructions these non-verbal cues are not necessarily present. Thus, it is likely that verbal explanations in some circumstances are not sufficient and have to be accompanied by non-verbal cues that align the learners' attention with that of the model. EMME do this by showing the learner what the model is looking at, at any given moment, which may resolve potential ambiguities in the model's verbal explanation. Hence, the discrepancy in results regarding the effectiveness of EMME might be due to whether verbal explanations are perceived as ambiguous without further guidance of an EMME. Indeed, there is some tentative evidence suggesting that this is the case: It is likely that verbal referents were ambiguous for the learners in the studies that found positive effects of EMME on learning classification tasks (Jarodzka et al., 2012; 2013; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012).

For instance, in one study learners had to learn to classify fish locomotion patterns and were shown videos of different fish while an expert gave verbal explanations about their locomotion pattern (Jarodzka et al., 2013). In his verbal explanation, the expert was using terms like 'the dorsolateral fin', which can be ambiguous when the learner does not yet know what that is. In this case, seeing the expert's eye movements (i.e., seeing what he is looking at) would help to attend to the right information at the right time. Similarly, in the study by Vitak et al. (2012) learners had to classify cells on histological slides. The learners were shown video examples with or

without the expert model's eye movements superimposed onto the example slides while listening to a verbal explanation. The expert referred to certain cells with terms like "there's" or "this one". Results indicated that the expert's eye movements were helpful in disambiguating the verbal referents as indicated by fewer classification errors on subsequent test tasks. Also, search behavior was more efficient for learners in the EMME condition as those learners needed less time to classify the cells and made fewer fixations on subsequent test tasks.

In studies on problem-solving tasks, in which the verbal referents were likely unambiguous (e.g., referring to "angle A" of a geometry problem), EMME had no beneficial effects on learning (Van Marlen et al., 2016) or a negative effect (Van Gog et al., 2009). In the study by Van Marlen et al. (2016), participants were shown modeling examples with or without the model's eye movements superimposed while also hearing verbal explanations about how to solve geometry problems (Exp. 2). Although EMME were effective for *more rapidly* guiding attention towards the information the model referred to (i.e., shorter times to first fixations and –probably as a consequence– longer fixation of referents), there was no difference in how many referents were fixated between the EMME and regular modeling example condition. This suggests that the verbal information was already sufficient in guiding the visual attention towards the relevant information, which might explain why EMME did not result in better performance (Van Marlen et al., 2016). That is, when the verbal referents are sufficient to follow and understand the model's demonstration and explanation, the attention guidance provided by the eye movements would be unnecessary, and research regarding the redundancy effect shows that displaying unnecessary information does not enhance learning but instead can be detrimental for learning (Kalyuga & Sweller, 2014).

In sum, the (un)ambiguity of the model's verbal explanation might determine the (in)effectiveness of EMME compared to regular video modeling examples in which the model's eye movements are not visualized. Therefore, Experiment 1 investigated whether verbal ambiguity indeed plays a role in the effectiveness of EMME.

Experiment 1

To investigate whether EMME would not only be effective for attention guidance but also for learning when the verbal explanation is

ambiguous, Experiment 1 replicated the study design by Van Marlen et al. (2016) but using EMME in which the referents were made more ambiguous (i.e., in the sense that they were less specific; e.g., “this angle” instead of “angle A”). A pilot study (Van Marlen, Van Wermeskerken, Boven et al., 2016) showed that participants studying ambiguous EMME fixated the referred information more often than participants studying ambiguous modeling examples (ME) without the model’s eye movements and there seemed to be a (non-significant) trend in test performance in favor of the ambiguous EMME condition. Based on the eye-tracking results of both the pilot study and previous studies demonstrating that EMME can guide attention (Jarodzka et al., 2013; Van Marlen et al., 2016, Mason et al., 2015; 2016), we hypothesized that ambiguous EMME would guide learners’ attention (H1), resulting in verbal referents being fixated more often (H1a), faster (H1b), and longer (H1c) than in the ambiguous ME condition. In addition, we hypothesized that the attentional guidance would also foster learning, resulting in higher accuracy (H2a) and faster solving speeds (H2b) on test problems in the ambiguous EMME than in the ambiguous ME condition.

Methods

Participants and Design

Participants were 57 Dutch university students, enrolled at a Faculty of Social Sciences (M_{age} = 21.98, SD = 2.82; 18 – 33 years, 21 male). They were assigned to either the EMME (n = 28) or ME condition (n = 29) at random. Participants had normal or corrected to normal vision and received a €5.00 reward for their participation.

Materials and Apparatus

Eye tracking equipment. A SMI RED250 eye tracker (SensoMotoric Instruments, GmbH) with a sampling rate of 250Hz was used to record the eye movements of the participants. The experiment was created and presented with the software package Experiment Center 3.4.165 (SensoMotoric Instruments, GmbH). The experiment was run on a monitor with a resolution of 1680 x 1050 pixels and a refresh rate of 60 Hz. Within the Experiment Center software the screen recording function was used for measuring the participants’ eye movements during example study (i.e., EMME or ME) and problem solving. ActionScript 3.0 was used to program the geometry problems and Flash Adobe CC was used for the presentation of the geometry problems.

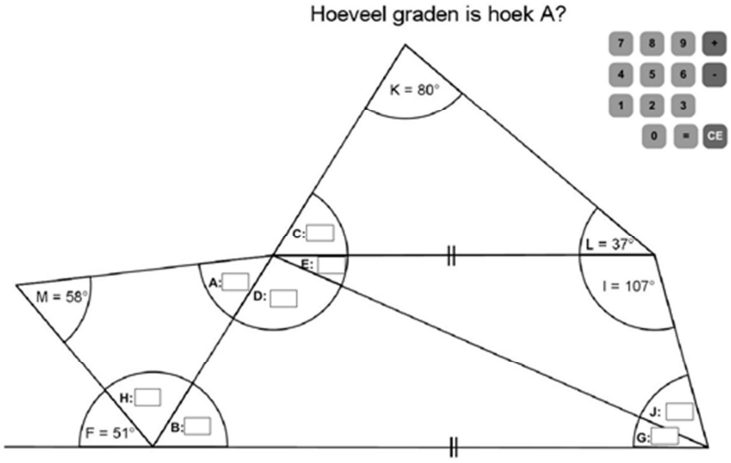


Figure 1. An example of a geometry problem used both in Experiment 1 and Experiment 2. On top is the problem statement, (translated from Dutch) “How many degrees is angle A?”.

Pretest. To check whether there were no prior knowledge differences between conditions a pretest was administered. It consisted of two open questions (e.g., Angle A is equal to:…) and three multiple-choice questions about a geometry figure. The questions tested knowledge of corresponding angles, alternating angles, straight lines, and triangles.

Geometry problems. The program *Geogebra* (www.geogebra.org) was used to create the geometry problems. The problems consisted of line drawings (i.e., black lines on top of a white background) of triangles and parallel lines which combined different geometrical principles regarding parallel lines (i.e., Z-rule or alternating angle rule; F-rule or corresponding angle rule). Above each geometry figure the problem statement was provided (e.g., “How many degrees is angle B?”), which required four angles (i.e., four sub steps) to be solved. All problems contained 13 angles coded A, B, C, etc. The values of some angles were provided in the problem and unknown values of angles were indicated with a question mark. Two geometry problems were created for the modeling examples together with two isomorphic versions of these problems (i.e., exact same layout of the line drawings but with different numbers). Additionally, four transfer problems were created which had different visual layouts and numbers than the isomorphic problems. Due to the different visual

layouts of the transfer problems, participants had to search for the correct starting point and order to solve the problems; for the isomorphic problems, in contrast, participants had observed the correct starting point and order in the modeling examples. To facilitate calculations (and to prevent participants from making errors), a digital calculator was placed right next of the problem statements for all geometry problems. For all problems, the rectangular area of the line drawings containing the problem statement, the calculator and the geometry figure had a width ranging from 1092-1351 pixels and a height ranging from 787-847 pixels (see Figure 1).

Video modeling examples. SMI Experiment Center 3.4.165 was used to present the experiment and to record the eye movements of the model for the modeling examples. SMI BeGaze 3.4.52 was used to visualize the eye movements and to create the video examples. In both conditions a female model explained verbally how to solve the problem, using ambiguous references to the angles (e.g., “Now that you know *this* angle, you can calculate the *other* angle.” (see Appendix for a screen capture of a video modeling example along with the transcript of the verbal explanation). In all modeling examples the model began by searching for the location of the angle mentioned in the problem statement. Once the angle was localized, the model started working backwards until a starting point to solve the problem was found. Then, the model started to determine the angle of the starting point and explained each solution step until the angle of the problem statement was determined. Both conditions were provided with the same verbal explanations along with the answers to each solution step (i.e., angles) that became visible during the problem-solving process. In the EMME condition, participants additionally saw the model’s eye movements superimposed onto the modeling example. The eye movements were visualized as a blue translucent dot with a 30-pixel diameter and were created using the Bee Swarm utility in SMI BeGaze. The length of the modeling example videos was the same across conditions (with one video lasting 122 s and the other video lasting 131 s).

Procedure

All participants were tested in individual sessions of ca. 25 min. As participants entered the lab, they were briefly given a general overview of the experiment. Then participants were presented with the pretest. Once the participants finished the pretest they were seated in front of the eye tracker at approximately 57 cm distance of the monitor

by means of a forehead and chin-rest. After a five-point calibration (including four-point validation) procedure, participants were instructed that they were about to see video examples about how to solve geometry problems and that each video demonstrated the correct solution to the geometry problem. In the EMME condition, participants were also instructed that they would see the model's eye movements, visualized as a moving blue translucent dot. To familiarize participants with such a visualization of a model's eye movements, participants in the EMME condition were shown a short EMME illustration video, in which the model was inspecting an image of a living room. Then, participants of both conditions were instructed that they would be presented with a similar problem after each video example and that they should solve this problem as rapidly and as accurately as possible. After these instructions, participants were presented with the two example-problem pairs. Subsequently, after re-calibration of the eye tracker, participants solved the four transfer problems. The order of the video examples with the matching isomorphic problems were counterbalanced across participants while the transfer problems were presented in a fixed order. During each problem-solving task, participants' eye movements, performance, and response times were recorded.

Data Analysis

Prior knowledge. Participants could score one point for each correctly answered pretest question (i.e., max. score = 5 points).

Eye tracking measures. To investigate whether EMME was beneficial for guiding participants' attention towards the right information in the geometry figure at the right time, we determined the onset time of the verbal referents in the narration of the verbal explanations (e.g., in the sentence "Now that you know *this* angle, you can calculate the *other* angle*."* the verbal referents are the words 'this' and 'other'; each video modeling example contained 15 referents). For each verbal referent an area of interest (AoI) was constructed around the corresponding angle in the geometry figure (e.g., angle D or E). We then determined the proportion of fixations on the AoI corresponding with the verbal referents by dividing the number of fixated AoIs of the verbal referents by the total number of referents (*proportion fixated*). In addition, we determined how much time participants needed to first fixate the AoI of the verbal referent (*time lag*) after the onset of the verbal referent, and how long participants fixated the AoI of the verbal referent (*fixation duration).* Fixations (i.e.,

peak velocity ≤ 40°/s and fixation duration ≥ 100 ms; cf. Jarodzka et al., 2013; Litchfield et al., 2010) were only included in the analyses if the fixation was within the time window of 1500 ms after the onset time of the verbal referent (cf. Dahan & Tanenhaus, 2005). Based on research findings suggesting that programming and initiating eye movements triggered by language input cost approximately 100 ms or longer (Altmann, 2011), fixations occurring within the first 100 ms were excluded from all eye tracking analyses.

Three participants (ME condition) were excluded from all eye movement data analyses because of poor calibration measures (i.e., deviation > 1 deg for x-axis or y-axis; x-axis deviation: $M = 0.46°$; $SD = 0.17°$; y-axis deviation: $M = 0.41°$; $SD = 0.15°$). One other participant (EMME condition) was identified as an outlier due to having very long time lags (*z*-score > 2.5). In addition, three participants (EMME condition $n = 2$, ME condition $n = 1$) were identified as outliers due to having very long fixation durations (z-score > 2.5).

Performance. One point was given for a geometry problem if all four steps had been performed correctly. Performance accuracy was computed by summing the number of points obtained and dividing that sum by the number of tasks, to obtain the proportion correctly solved problems; this was done separately for isomorphic and transfer problems. Two participants were identified as outliers for the analysis of isomorphic test problem accuracy and two participants were identified as outliers for the analysis of transfer test problem accuracy because of poor performance scores (*z*-score < -2.5). Performance speed was computed by averaging the response times on the correctly solved problems, separately for isomorphic and transfer problems. One participant was identified as an outlier for the response times isomorphic analysis and one participant was identified as an outlier for the response times transfer analysis because of very high response times (*z*-score > 2.5).

Statistical Analyses. After the exclusion of all participants who were outliers on one of the performance and eye-tracking measures, the final sample consisted of forty-four participants ($n = 22$ EMME condition; $n = 22$ ME condition) and this sample was used for all the analyses reported. The data were analyzed with non-parametric Mann-Whitney *U* tests due to violation of the normality assumption with the exception of the time to first fixation analysis, which was analyzed with a *t*-test. We used *r* as a measure of effect size, with $r = .10$, $r = .30$, $r = .50$, representing small, medium, and large effects

respectively (Cohen, 1988). We additionally conducted Bayesian analyses[3] with JASP (version 0.8.6.0; jasp-stats.org; JASP team, 2018). One of the advantages of Bayesian analyses is that it expresses how much more likely the alternative or null hypothesis is given the obtained data instead of just rejecting the null hypothesis (see for an overview article: Wagenmakers et al., 2018). How much more likely the obtained data are under one of the hypotheses compared to the other hypothesis is expressed as a Bayes factor (BF). For instance, a BF_{10} = 8.00 indicates that the obtained data are eight times more likely under the alternative hypothesis than the null hypothesis. For each main analysis, we added the BF.

Results

We first checked for differences in prior knowledge. Prior knowledge was quite high in both conditions, and there was a small but significant difference in prior knowledge between the EMME condition (M = 4.00 out of 5; SD = 1.07) and the ME condition (M = 4.59 out of 5; SD = 0.80), U = 159.00, z = -2.20, p = .028, r = .33.

Eye Tracking Measures

See Table 1 for the means and standard deviations for the eye-tracking measures used in the following analyses. To address our first hypothesis, that ambiguous EMME would guide learners' attention to the relevant information in the video modeling examples (H1), a Mann-Whitney U test was conducted on the proportion of fixated verbal referents and fixation duration on referents, and a t-test on the time to first fixation on referents. The analysis on the proportion of fixated verbal referents (H1a) revealed that, in line with our hypothesis, participants in the EMME condition fixated the referents more often than participants in the ME condition, U = 97.50, z = -3.42, p = .001, r = .52, BF_{10} = 78.633. The Bayes factor indicates that the observed data are 78.6 times more likely under the alternative hypothesis (i.e., the conditions differ in the proportions of fixations) than under the null hypothesis (i.e., the conditions do not differ in proportions of fixations). In terms of the time (in ms) required to first fixate the verbal referents after onset (H1b), participants in the EMME condition were –as expected- significantly faster than participants in the ME condition, t (42) = 3.24, p = .002, r = .49, BF_{10} = 15.694. The Bayes factor indicates that the observed data are 15.6 times more likely under the alternative hypothesis (i.e., the conditions differ in the

[3] We would like to thank an anonymous reviewer for this suggestion.

Table 1. Mean (and SD) and Median (and Range) of Performance and Eye Tracking Measures of the Eye Movement Modeling Example (EMME) and Modeling Example (ME) Conditions in Experiment 1.

	EMME *M* (SD) | *Mdn* (range)	ME *M* (SD) | *Mdn* (range)
Eye Tracking Measures		
Proportion of Fixations (*n* = 44)	0.46 (0.13) | 0.53 (0.47)	0.32 (0.12) 0.30 (0.47)|
Time Lag (in ms; *n* = 44)	563.44 (111.21) | 563.55 (440.74)	667.90 (102.18) | 679.49 (391.57)
Fixation Duration (in ms; *n* = 44)	499.03 (187.51) | 465.94 (699.83)	418.95 (159.91) | 361.13 (678.29)
Proportion Correct		
Isomorphic (*n* = 44)	0.93 (0.18) | 1.00 (0.50)	0.91 (0.20) | 1.00 (0.50)
Transfer (*n* = 44)	0.86 (0.18) | 1.00 (0.50)	0.88 (0.17) | 1.00 (0.50)
Response Times (s)		
Isomorphic (*n* = 44)	71.43 (17.32) | 67.16 (65.13)	61.54 (12.96) | 61.00 (62.34)
Transfer (*n* = 44)	110.43 (36.93) |101.42 (127.86)	96.19 (27.32) | 89.14 (113.08)

time to first fixation) than under the null hypothesis. Contrary to our hypothesis (H1c), however, there was no significant difference in fixation duration (in ms) between participants in the EMME condition and participants in the ME condition, $U = 170.00$, $z = -1.69$, $p = .091$, $r = .25$, $BF_{10} = 0.751$. The Bayes factor indicates that the observed data are less than one time more likely under the alternative hypothesis (i.e., the conditions differ in the fixation duration) than under the null hypothesis.the ME condition, t (42) = 3.24, $p = .002$, $r = .49$, $BF_{10} = 15.694$. The Bayes factor indicates that the observed data are 15.6 times more likely under the alternative hypothesis (i.e., the conditions differ in the time to first fixation) than under the null hypothesis. Contrary to our hypothesis (H1c), however, there was no significant difference in fixation duration (in ms) between participants in the EMME condition and participants in the ME condition, $U = 170.00$, $z = -1.69$, $p = .091$, $r = .25$, $BF_{10} = 0.751$. The Bayes factor indicates that the observed data are less than one time more likely under the alternative hypothesis (i.e., the conditions differ in the fixation duration) than under the null hypothesis.

Performance

See Table 1 for the means and standard deviations of the performance measures. To address our second hypothesis that seeing an EMME would result in enhanced performance accuracy (H2a) and speed (H2b) compared to the ME condition, non-parametric Mann-Whitney *U* tests were conducted. These revealed that neither

accuracy on the isomorphic problems, $U = 231.00$, $z = -.41$, $p = .684$, $r = .06$, $BF_{01} = 3.147$, nor accuracy on the transfer problems, $U = 237.50$, $z = -.12$, $p = .904$, $r = .02$, $BF_{01} = 3.300$, differed significantly between conditions. The Bayes factors for the proportion correct for both the isomorphic and transfer problems indicates that the observed data are more than three times more likely under the null hypothesis (i.e., no difference between conditions) than under the alternative hypothesis. Regarding the response times for the isomorphic problems the results indicate that participants were faster in the ME condition than in the EMME condition, $U = 153.00$, $z = -2.09$, $p = .037$, $r = .32$, $BF_{10} = 1.806$. However, for the transfer problems there were no differences in response times between the conditions, $U = 191.00$, $z = -1.20$, $p = .231$, $r = .18$ $BF_{01} = 1.445$. The Bayes factor for the response times for the isomorphic problems indicates that the observed data are 1.8 times more likely under the alternative hypothesis (i.e., the conditions differ in the response times) than under the null hypothesis. However, for the transfer problems the Bayes factor indicates that the observed data are more than 1.4 times more likely under the null hypothesis (i.e., no difference between conditions) than under the alternative hypothesis.

Discussion

In line with our first hypothesis, results of the eye-tracking analyses revealed that seeing the model's eye movements in verbally ambiguous modeling examples (EMME) guided learners' attention more often (H1a) and faster (H1b) towards the verbally referred information than verbally ambiguous modeling examples that did not display the model's eye movements (ME). However, and in contrast to our second hypothesis, this did not result in better (H2a) performance on the geometry test problems in the EMME vs. the ME condition. Regarding response times no support was found for our hypothesis (H2b) as we found that learners of the ME condition were faster in solving the isomorphic problems, however for the transfer problems the solving speed did no longer differ between conditions. Although we had expected EMME to be more effective under conditions of verbal ambiguity, the results are largely in line with Authors (2016b) who used examples with unambiguous explanations. This seems to suggest that the ambiguity of the model's verbal explanation (in terms of being unspecific) might not play a role in the effectiveness of EMME.

It would be premature to draw that conclusion, however, as a possible explanation for the current findings might lie in the high amount of prior knowledge (i.e., average score of 4.1-4.6 out of 5 in both conditions) and the (resulting) overall high performance on the test problems in both conditions. Under conditions of high prior knowledge, the cognitive load imposed by the learning task is lower (Sweller, Ayres, & Kalyuga, 2011), and learners in the condition without attention guidance will probably be able to accommodate the additional load imposed by having to search for the verbal referents, without losing track of the model's demonstration and explanation. This explanation has also been offered for mixed findings in research on (other forms of) visual cueing (Van Gog, 2014) and is in line with findings from a recent meta-analysis showing that the effectiveness of attention guidance by means of visual cues is moderated by the amount of prior knowledge a learner has (Richter et al., 2016): learners with low levels of prior knowledge generally benefitted more from visual cues in multimedia learning materials than learners with higher levels of prior knowledge. However, note that even advanced learners or experts can benefit from EMME provided that there is still room for improvement in their performance. This was shown in a recent study by Gegenfurtner and colleagues (2017) in which experts and novices performed a diagnostic task on CT and PET medical images before and after watching an EMME of an expert verbally explaining how to perform the task. It was found that even though experts had high prior knowledge, they benefited from seeing the EMME, which suggests that the effectiveness is not necessarily limited to lower levels of prior knowledge but might also depend on the extent to which there is room for improvement.

Therefore, Experiment 2 investigated the effects of ambiguous EMME and ME in a lower prior knowledge sample of secondary education students who have not yet been taught the F and Z principles in geometry.

Experiment 2

Next to the ambiguous conditions of Experiment 1, we also included unambiguous EMME and ME in Experiment 2, because these had not yet been tested in a lower prior knowledge sample. Thus, Experiment 2 had a 2 (EMME vs. ME) x 2 (ambiguous vs. unambiguous verbal explanations) between-subjects design and was conducted with secondary education students. Because this

experiment was conducted in a regular classroom setting, it did not involve eye tracking but only assessed test performance. Based on previous research showing beneficial effects of studying EMME (Jarodzka et al., 2012; 2013) and assuming the results of Experiment 1 were indeed due to the high prior knowledge, it was hypothesized that studying EMME would result in better test performance than studying ME (H1), that examples with unambiguous verbal explanations would lead to better performance than examples with ambiguous explanations (H2), and that the beneficial effects of EMME on performance would be larger in the ambiguous verbal explanation condition (interaction effect, H3).

Methods

Participants and Design

Initially 142 first-year students from five classes of a Dutch school for secondary education participated in Experiment 2. However, due to software difficulties the experiment did not run properly in one class (n= 30). In addition, four participants did not provide informed consent to use their data. This resulted in a final sample of 108 students (M_{age} = 12.05, SD = 0.46, 53 male). Participants were randomly assigned to one of the four conditions resulting from a 2 (modeling example: EMME vs. ME) x 2 (verbal explanation: ambiguous vs. unambiguous) between-subjects design: ambiguous EMME (n = 28), unambiguous EMME (n = 26), ambiguous ME (n = 26), and unambiguous ME (n = 28).

Materials

The experiment was created and conducted with the online questionnaire software Qualtrics (www.qualtrics.com). The pretest and the geometry test problems (isomorphic and transfer) were the same as those used in Experiment 1. The modeling examples (EMME and ME) with the ambiguous verbal explanations were exactly the same as in Experiment 1; the modeling examples with unambiguous verbal explanations were created by replacing the audio file with ambiguous verbal referents (e.g., “this angle”) by an audio file (recorded by the same model) with unambiguous verbal referents (e.g., “angle A”).

Procedure

The experiment, which was conducted during math class lasting approximately 50 min, was run in a computer classroom at participants’ school. Participants were assigned to one of the conditions at random in advance of the session. There was a sheet of

paper next to every computer with the name of the participant and the corresponding participant number to ensure that participants would work with the correct version of the computer program. Participants first received general and practical instructions about the experiment and were given the opportunity to ask clarification questions. Participants were then instructed to start the computer program and work through the program at their own pace. The program first asked students for their consent to use their performance data for research purposes. Then the pretest followed. After the pretest, all participants were presented with brief definitions and example images of the different types of angles used in the experiment[4]. Then participants were instructed that they were about to see video examples that would teach them how to solve geometry problems. In the EMME conditions, participants were also instructed that they would see the model's eye movements, visualized as a moving blue translucent dot (and they were shown a screenshot illustrating this; cf. Experiment 1). Then, participants in all conditions were instructed that after each video example they first had to indicate on a 5-point rating scale ranging from 1 (*very unclear*) to 5 (*very clear*) how clear they thought the verbal instructions were in the video example. Participants were additionally instructed that after rating the video example that they would be presented with a similar problem after each video example and that they should solve this problem as rapidly and as accurately as possible. After these instructions, participants were presented with the two example-problem pairs in the format of their assigned condition. Subsequently, participants solved the four transfer problems. The order of the video examples with the matching isomorphic problems were counterbalanced across participants and the transfer problems were presented in random order. During each problem-solving task participants' performance was recorded.

Data Analysis

Prior knowledge. See Experiment 1.

Video Ratings. The ratings of how clear learners perceived the verbal instructions of the video examples to be, were averaged across the two videos for each participant.

Performance. See Experiment 1 for details regarding the calculation of performance accuracy. Twenty-one participants were

[4] This was done to prevent a bottom effect, as participants had not yet been taught the relevant geometry principles. Participants in all conditions received the same information.

unable to finish all the transfer problems before the end of the math lesson and therefore the data of these participants were excluded, leaving 87 participants for the transfer performance analysis.

Results

The prior knowledge, video ratings and performance data are presented in Table 2. The data were analyzed with 2 (modeling example: EMME vs. ME) x 2 (instruction: ambiguous vs. unambiguous verbal referents) ANOVAs, and partial eta squared is reported as a measure of effect size, with $\eta_p^2 = .01$, $\eta_p^2 = .06$, $\eta_p^2 = .14$, representing small, medium, and large effects respectively (Cohen, 1988). In addition, we performed the Bayesian ANOVAs and reported the inclusion Bayes Factor (BF_{inc}). The BF_{inc} estimates the likelihood of the model if it contains the effect. For instance, when reporting a main effect of modeling example a $BF_{inc} = 7.500$, would indicate that the model including the effect of modeling example is 7.5 times more likely than the null model without an effect of modeling example. A check on the pretest scores revealed no main effect of the type of modeling example, F (1, 104) < 1.00, $p = .954$, no main effect of instruction, F (1, 104) < 1.00, $p = .589$, and no interaction, F (1, 104) < 1.00, $p = .469$.

Video Ratings

To test whether participants perceived the verbal referents as less clear in the ambiguous conditions and to test whether the ratings of the verbal referents of the modeling examples are also affected by the presence of the model's eye movements a 2x2 ANOVA[5] with the video ratings as dependent variable was conducted. This analysis revealed a main effect modeling example, F (1, 104) = 9.21, $p = .003$, $\eta_p^2 = .08$, $BF_{inc} = 15.950$, no significant effect of instruction, F (1, 104) = 3.06, $p = .083$, $\eta_p^2 = .03$, $BF_{inc} = 2.453$, and a significant interaction, F (1, 104) = 6.82, $p = .010$, $\eta_p^2 = .06$, $BF_{inc} = 6.927$. The results of the inclusion Bayes Factor suggest that the model including an effect of modeling example is almost 16 times more likely than the null model. Simple effects analysis revealed that the interaction was caused by lower video ratings for the ME ambiguous condition compared with the

[5] Because the ME unambiguous instruction condition did not meet the normality assumption, we additionally conducted non-parametric Kruskal-Wallis tests to examine whether the main effects from the 2x2 ANOVA would hold. This revealed the same effect for modeling example, H (1) = 7.18 p = .007, and instruction, H (1) = 1.67, $p = .197$.

Table 2. Mean (and SD) and Median (and Range) of the Prior Knowledge Test, Video Ratings, and the Proportion Correct of the Solved Geometry Problems for the Eye Movement Modeling Example (EMME) Conditions and the Modeling Example (ME) Conditions in Experiment 2.

	EMME		ME	
	Ambiguous Instructions *M* (SD) | *Mdn* (range)	Unambiguous Instructions *M* (SD) | *Mdn* (range)	Ambiguous Instructions *M* (SD) | *Mdn* (range)	Unambiguous Instructions *M* (SD) | *Mdn* (range)
Prior Knowledge (1-5; *n*=108)	1.93 (1.30) | 2.00 (5.00)	1.88 (1.24) | 2.00 (5.00)	1.77 (1.21) | 2.00 (4.00)	2.07 (1.18) | 2.00 (5.00)
Video Ratings (1-5; *n*=108)	3.79 (0.75) | 4.00 (3.00)	3.65 (0.80) | 3.50 (3.00)	2.92 (0.90) | 2.50 (3.50)	3.59 (0.72) | 3.50 (3.50)
Proportion Correct				
Isomorphic (*n*=108)	0.59 (0.41) | 0.50 (1.00)	0.54 (0.40) | 0.50 (1.00)	0.19 (0.32) | 0.00 (1.00)	0.38 (0.35) | 0.50 (1.00)
Transfer (*n*=87)	0.28 (0.36)) | 0.00 (1.00)	0.30 (0.31) | 0.25 (0.75)	0.11 (0.27) | 0.00 (1.00)	0.21 (0.28) | 0.00 (0.75)

EMME ambiguous condition, F (1, 104) = 15.94, $p < .001$, $r = .38$, BF_{10} = 75.981. The Bayes factor indicates that the observed data are almost 76 times more likely under the alternative hypothesis (i.e., the ambiguous conditions differ) than under the null hypothesis. The video ratings did not differ between the ME unambiguous condition and the EMME unambiguous condition, F (1, 104) = 0.09, $p = .766$, $r = .03$, BF_{10} = 0.286. The Bayes factor indicates that the observed data are only 0.286 times more likely under the alternative hypothesis (i.e., the unambiguous conditions differ) than under the null hypothesis. *4.2.2* **Performance**

To examine the effectiveness of EMME and also to examine whether this was affected by the ambiguity of verbal referents, a 2x2 ANOVA[6] was conducted. For the isomorphic problems this analysis revealed a main effect of modeling example, F (1, 104) = 15.36, $p <$ .001, $\eta_p^2 = .13$, BF_{inc} = 98.329, no main effect of instruction, F (1, 104) < 1.00, $p = .359$, $\eta_p^2 = .01$, BF_{inc} = 0.351 and no interaction, F (1, 104) = 2.67, $p = .106$, $\eta_p^2 = .03$, BF_{inc} = 0.696. This analysis indicates that learners in the EMME conditions performed better on the isomorphic problems compared with learners in the ME conditions, however the ambiguity of the verbal referents did not affect the performance nor the effectiveness of EMME relative to ME. The Bayes inclusion factor suggest that the model including an effect of modeling example is almost 98 times more likely than the null model. Similar results were found for the transfer problems as this analysis revealed a main effect of modeling example, F (1, 83) = 4.13, $p = .045$, $\eta_p^2 = .05$, BF_{inc} = 0.916, no main effect of instruction, F (1, 83) < 1.00, $p = .364$, η_p^2 = .01, BF_{inc} = 0.247 and no interaction, F (1, 83) < 1.00, $p = .543$, $\eta_p^2 <$ .01, BF_{inc} = 0.185 (see Table 2). This indicates that performance was only affected by the type of modeling examples with learners in the EMME conditions outperforming the learners in the ME conditions.

[6] Due to the non-normal distribution of the performance data, we additionally conducted non-parametric Kruskal-Wallis tests to examine whether the main effects from the 2x2 ANOVA would hold. For the isomorphic problems, we found that EMME outperformed the ME condition, H (1) = 12.89, $p < .001$, and there was no difference between the ambiguous instruction condition and the unambiguous condition, H (1) = 0.68, $p = .410$. Similar results were found for the transfer problems, EMME outperformed the ME condition, H (1) = 4.15, p = .042, and there was no difference between ambiguous instruction condition and the unambiguous condition, H (1) = 1.38, $p = .241$. In sum, the results of the non-parametric tests are in line with the results regarding the main effects of the 2x2 ANOVA.

The Bayes inclusion factor suggest that the model including an effect of modeling example is roughly one time more likely than the null model.

Discussion

The aim of Experiment 2 was to examine the effects of EMME and explanation ambiguity on learning to solve a procedural task in secondary education students. In line with our hypothesis (H1), EMME yielded higher learning outcomes than ME. However, and contrary to our hypothesis (H2), ambiguity of the verbal explanation did not affect learning outcomes, either directly or in interaction with the type of modeling example. We will discuss these results in more detail in the next section.

General Discussion

In two experiments, we investigated the role of the ambiguity of the model's verbal explanation and students' prior knowledge in the effectiveness of EMME for learning a procedural geometry task. Experiment 1 examined the effects of EMME on attention allocation and learning when the verbal explanation was ambiguous (i.e., unspecific). This experiment revealed that EMME helps to guide learners' visual attention towards the verbal referents faster (H1b) and more often (H1a) compared to a modeling example that did not show the model's eye movements. However, this attentional guidance by EMME did not result in higher accuracy (H2a) or consistently faster (H2b) geometry problem solving compared to ME, possibly due to the high prior knowledge (i.e., ceiling effect). Consequently, Experiment 2 investigated the effects of EMME and explanation ambiguity on learning to solve a procedural geometry task in secondary education students who had limited prior knowledge.

Experiment 2 revealed that secondary education students in the EMME conditions outperformed their counterparts in the ME conditions (H1) on both isomorphic and transfer problems. This is an interesting and important finding, as it extends the tasks to which EMME are applicable. There was some evidence that EMME can be effective for learning classification tasks (Jarodzka et al., 2012; 2013; Sridharan et al., 2012; Vitak et al., 2012) and strategy learning (Mason et al., 2015; 2016), but to our knowledge, this is the first study to demonstrate that EMME can be effective to enhance learning of a procedural problem-solving task. Earlier studies found no beneficial

effect of EMME for learning a procedural geometry problem-solving task in university students (Van Marlen et al., 2016) and a study using a procedural puzzle problem even found a negative effect on learning when the EMME was accompanied by a verbal explanation (Van Gog et al., 2009). The present study demonstrates that EMME can be effective for learning a procedural problem-solving task, but suggests that this is only the case if prior knowledge is low. Findings by Van Marlen et al. (2016) and from Experiment 1 show that for university students, who had relatively high prior knowledge, EMME successfully guided their attention but this did not affect their learning outcomes. Presumably, students who have more working memory capacity at their disposal, for instance because they have more prior knowledge, are able to accommodate the additional working memory load imposed by having to search for the verbal referents, without losing track of the model's demonstration and explanation. So even if students who do not receive EMME may take a little more time to locate the referent, this does not hamper their learning. For the secondary education students in Experiment 2, who had lower prior knowledge than the university students from Experiment 1, the guidance provided by EMME did foster learning.

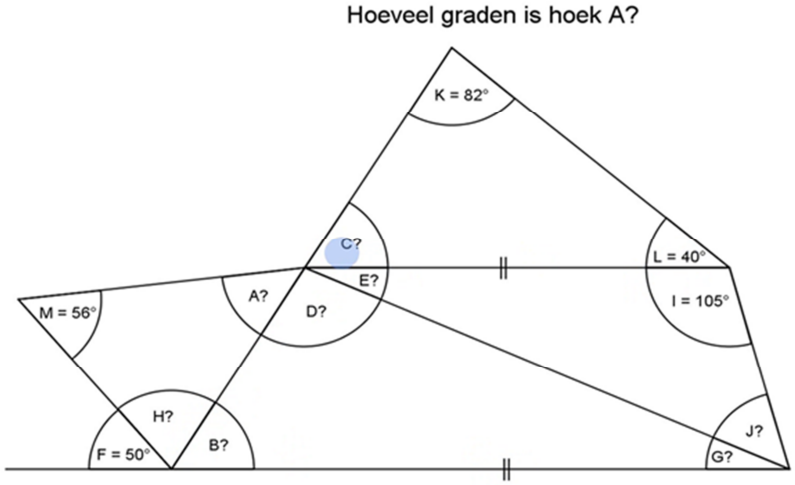
Interestingly, the ambiguity of the verbal explanation did not significantly affect learning outcomes in Experiment 2 (H2). Learners did perceive the verbal explanations in the video examples to be clearer when they were accompanied by the model's eye movements, especially when the verbal explanation was ambiguous. This interaction effect on perceived clarity seems to be mirrored by the test performance data shown in Table 2, with the difference between the mean scores of the EMME and ME conditions being larger when the explanation was ambiguous than when it was unambiguous. However, there was no significant interaction effect on performance. There are several possible explanations for this lack of effect of verbal ambiguity. First, in Experiment 2, it is possible that given the substantial variation in test performance (which is not uncommon in the classroom), the sample size was too low to be able to detect a significant interaction. Another potential explanation might be that the ambiguity manipulation was too subtle. Our ambiguity manipulation consisted of using unspecific rather than specific referents (e.g., "that angle" instead of "angle A"). Although it may not be immediately clear for learners where to look based on such unspecific referents, they may be less detrimental for learning than the use of referents that learners

will not know the meaning of when they cannot see what the model is looking at (e.g., the use of jargon like “dorsolateral fin”; Jarodzka et al., 2013). Moreover, the fact that the model could be observed performing the problem-solving steps by clicking and typing, would have synchronized learners’ attention intermittently throughout the example. So even if they lost track due to the unspecific referents in the no EMME condition, this would be resolved once the model performed a step.

Therefore, it would be interesting in future research to replicate Experiment 2 with a larger sample size and to extend it to different problem-solving tasks (with/without physical actions on the computer screen) and other sources of ambiguity of the model’s verbal explanation. In addition, it would also be interesting to directly manipulate prior knowledge within a single experiment. A limitation of the present study is that we only manipulated prior knowledge between experiments, and therefore cannot directly verify that this is indeed the factor that moderates the effectiveness of EMME. In addition, future research might explore whether other factors like working memory capacity, spatial ability, and reasoning ability also moderate the effectiveness of EMME. In general, further research on EMME (and other forms of cueing) in which prior knowledge is manipulated is important considering that the literature shows on the one hand that people with low prior knowledge benefit most of the visual cues (Richter et al., 2016), while on the other hand, a recent study suggests that for certain tasks, experts may profit more from EMME than novices (Gegenfurtner et al., 2017).

To conclude, the present study shows that –in contrast to prior findings– EMME can be effective in enhancing learning of a procedural problem-solving task, at least when prior knowledge is low. This is also relevant for educational practice, as our study demonstrates that EMME –once created– can easily be implemented in secondary education classrooms.

Appendix. Example of a transcript of the verbal explanation in a modeling example in Experiment 1 and Experiment 2.



Screenshot of a modeling example (EMME condition) in Experiment 1 and 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 this angle. 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 it right now, because the second angle is unknown. The second angle 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, because the right angle is also unknown, it is not possible to calculate this angle no. You cannot calculate the right angle directly, but it can be derived from this angle, because these are equal. This can be seen by this 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 these angles are equal. This angle is unknown for now but can be calculated. This angle equals 180 degrees minus the known angles, equals 58 degrees. Now this angle is known, you know that the other angle, by means of the corresponding angle principle, also equals 58 degrees. With this angle known, you can now calculate the other angle. This angle equals 180 degrees minus the known angles, equals 72 degrees. With that angle 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."*

Chapter 4

Chapter 5

Does the Purported Expertise of a Model Influence the Effectiveness of Eye Movement Modeling Examples?

This chapter is submitted for publication as:
Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018). *Does the Purported Expertise of a Model Influence the Effectiveness of Eye Movement Modeling Examples?* Manuscript submitted for publication.

Acknowledgement of author contributions: TM, MW, HJ, and TG designed the experiments, TM recruited participants and collected the data, TM analyzed the data, MW checked the data package, TM drafted the manuscript, all authors contributed to critical revision of the manuscript, MW, HJ, and TG supervised the experiments.

Chapter 5

Abstract
In contrast to regular examples, Eye movement modeling examples (EMME) not only show students what actions the model (e.g., teacher or expert) is performing on the computer screen content, but also where the model is looking, by visualizing the model's eye movements (e.g., as circles or dots) overlaid on the screen content. EMME thus provide attention guidance to the learner, which has been shown to foster performance and learning. Because students' inclination to follow the model's eye movements might depend on how they perceive the model, the question addressed here is whether the purported expertise of the model affects learning with EMME. Secondary education students studied EMME about geometry problems in two conditions with identical example content but different purported model expertise: In the Expert condition, the model introduced herself as a math teacher (N=54), and in the Non-Expert condition, as a Dutch teacher (N=52). Subsequently, students received two EMME and afterwards solved isomorphic and transfer problems. Results revealed no differences in learning outcomes between conditions, suggesting that learning from EMME is not influenced by the model's purported expertise.

Introduction

Video modeling examples, in which a human model (e.g., teacher or expert) demonstrates and explains how to perform a learning task, have proven to be an effective tool to foster learning (Van Gog & Rummel, 2010; Renkl, 2014;). One common type of video modeling examples are screen recordings in which the model is performing a computer-based task. The videos show the student the content the model has on the computer screen and what actions they are performing on the screen content. Although these examples can be very helpful for learning, they carry the risk that students miss out on important information when the information in the video examples is transient (i.e., information is only temporarily available), which might hamper learning (Ayres & Paas, 2007). One way to overcome this problem is by providing attentional guidance, for instance, by displaying where the model is looking by means of a cursor (e.g., circle or dot) that represents the gaze position, overlaid on the screen recording. Such examples, which have been referred to as eye movement modeling examples (EMME), have been shown to enhance learning of perceptual classification tasks (Jarodzka et al., 2012; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012), text and picture integration (Mason, Pluchino, & Tornatora, 2015; 2016; Mason, Scheiter, & Tornatora, 2017; Scheiter, Schubert, & Schüler, 2018), navigating digital hyperlinked texts (Salmerón & Llorens, 2018) and problem-solving tasks (Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2018).

The reason why EMME are effective for learning compared with regular video modeling examples (i.e., video examples without the model's eye movements superimposed) is that seeing where the model was looking during the task demonstration helps to synchronize the visual attention of the learner with that of the model, creating joint attention (Brennan, Chen, Dickinson, Neider & Zelinsky, 2008; Frischen, Bayliss, & Tipper, 2007). Joint attention between the model and the learner helps the learner to attend the relevant information (i.e., the information the model is currently considering) at the right time, which helps learners to follow the model's demonstration and explanation of the task. Hence, the effectiveness of EMME depends on the learners' following of the visualization of the model's eye movements. Recent research suggests, however, that this gaze following may be affected by the social status of the model (Gobel,

Tufft, & Richardson, 2018). Therefore, we investigate whether the model's purported expertise plays a role in learning with EMME. Considering that the availability of eye trackers is increasing, the use of eye trackers to create EMME is also likely to increase . Thus, knowing what factors affect learning from EMME is a very significant and timely issue for the field.

Effects of (purported) model expertise on performance and learning

There is some evidence from the literature on example-based learning that model characteristics, such as the model's (perceived) expertise, can influence the effectiveness of (regular) modeling examples. According to the Model-Observer Similarity (MOS) hypothesis, examples are more effective when students perceive themselves to be more similar to the model (Schunk, 1987). Indeed, there is some evidence that novice students learn more from a model with low expertise than from a model with high expertise (e.g., Braaksma, Rijlaarsdam, & van den Bergh, 2002; Sonnenschein & Whitehurst, 1980). However, in these studies, the content of the examples also differed as a function of model expertise. In a recent study, the content of video modeling examples in physics was kept equal while the model's age (peer vs. adult) and purported expertise (low vs. high) were manipulated (Hoogerheide, Van Wermeskerken, Loyens, & Van Gog, 2016). Results revealed that the explanations in the video examples (which were identical) were rated as being higher in quality when the model was an adult or an alleged expert, although only model age (presumably also associated with expertise or task appropriateness) positively affected learning outcomes.

The fact that the model's social status affected perceived quality of the model's explanations, is likely associated with attention paid to the model (which might explain the effects on learning). That individuals who are perceived as having more expertise, attract more attention, was shown in a study by Cheng, Tracy, Foulsham, Kingstone, and Henrich (2013) in which participants were instructed to do a group assignment and were rated by fellow group members and independent raters afterwards on how much prestige (i.e., the sharing of expertise to gain respect) the participants had expressed during the group assignment. In Experiment 2, observers were presented with the video recordings of the groups of Experiment 1 (featuring individuals scoring both high and low on prestige), while their eye movements were recorded. Results showed that individuals scoring

higher on prestige in Experiment 1were attended more often and longer by the observers than individuals scoring lower on prestige.

While they provide some evidence that the model's perceived or purported expertise might affect attention and learning from video examples, the above studies pertain to situations in which the model is visible in the video, which is not the case in EMME (as they are screen-recording examples). Yet a recent study provides some specific evidence that the purported expertise of the model might also influence how learners interpret and attend to the visualized eye movements (which could affect the effectiveness of EMME). Gobel et al. (2018) conducted two experiments in which participants had to respond as fast as possible to the appearance of a target (a blue square) in one of four quadrants on the screen. However, just before the appearance of the target shape, participants were presented with a red dot in one of the four quadrants. In Experiment 1 participants were led to believe either that this red dot was a computer-generated cue or that it represented the gaze location of a partner present on the other side of the room. Results showed an increased inhibition of return effect (i.e., longer response time to detect the target when the target was located in the same quadrant as where the red dot had previously been) when participants thought it was a gaze cue; that is, it took participants longer to respond to targets in locations that were supposedly already looked at by a partner. This increased inhibition of return effect suggests that participants relied on the confederate as having searched that part of the stimulus for the target, as a result of which the participant took more time to attend to that target area again.

In Experiment 2, Gobel et al. (2018) examined whether this effect was further modulated by the partner's social rank. More specifically, participants were introduced to their partner (a confederate) and were led to believe her social rank was either low or high (by means of a cover story) before conducting the spatial cueing task. Participants were instructed that the red dot represented the gaze location of their partner and on half of the trials they thought they were both engaged in the same task, on the other half that their partner was engaged in another task. Results showed an inhibition of return effect, with participants taking longer to respond to targets that were just looked at by the confederate when his/her purported social rank was high but not when his/her purported social rank was low, and only when participants thought their partner was engaged in the same

task. In sum, these findings show that participants may react differently to gaze cursors depending on the (purported) social status of the person whose eye movements they are observing.

The Present Study

The aim of the current study was to examine whether the purported expertise of the model who provides the verbal explanation and guidance in the EMME would affect secondary education students' learning outcomes. We used EMME in which a female model demonstrated and explained how to solve geometry problems (cf. Van Marlen et al., 2018). The verbal explanation and visual guidance in both conditions were kept equal, but the model's purported expertise was varied by an introductory cover story. In the Expert condition the model introduced herself as a math teacher who has to make videos in the context of a professional development course on multimedia learning, and is confident that she can clearly explain how to solve the geometry problem. In the Non-Expert condition, the model introduced herself as a Dutch teacher who has to make videos in the context of a professional development course on multimedia learning, and hopes she can clearly explain how to solve the problem. Based on the findings reviewed above, we hypothesized that students might pay closer attention to a model with high purported expertise, which would lead to higher learning outcomes in the *Expert condition* than in the *Non-Expert* condition. In line with the results of Hoogerheide et al. (2016) we also expect that students in the Expert condition would rate the model's explanations as being of higher quality than students in the Non-Expert condition. If we would indeed find that perceived explanation quality and learning outcomes would be affected by purported model expertise, an interesting question would be whether all students would be equally affected, or whether this would depend on their sensitivity to authority. We therefore explored this by administering two subscales of the Schommer Epistemological Questionnaire (EQ; Schommer, 1998).

Method

Participants

Participants were 107 Dutch first-year pre-university students (i.e., the highest level of secondary education in the Netherlands) from four classes of one school. One student did not consent to the use of his/her data and was therefore excluded, resulting in a final sample of

106 students (M_{age} = 11.95, *SD* = 0.42, 64 male). Students were randomly assigned to either the Expert (*n*= 54) or the Non-Expert (*n*= 52) condition.

Materials

The online questionnaire software Qualtrics (www.qualtrics.com) was used for creating and conducting the experiment. The geometry examples and problems used in the current study were developed and used in previous studies in which EMME were shown to be effective for guiding attention and to be more effective for secondary education students' learning outcomes than examples without the model's eye movements (Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016; Van Marlen et al., 2018).

Pretest. A pretest was administered to be able to check and control whether there were prior knowledge differences between conditions. The pretest measured the knowledge about corresponding angles, alternating angles, straight lines and triangles and consisted of three multiple choice questions and two open questions (e.g., Angle A is equal to:…) about two geometry figures.

Geometry problems. The learning tasks consisted of geometry problems created with the program Geogebra (www.geogebra.org). Above each line drawing of a geometrical shape the problem statement was given (e.g., "How many degrees is angle A?"). The line drawings depicted triangles, parallel lines, and straight lines, which combined several geometrical principles regarding alternating angles, corresponding angles, triangles, and straight lines. Each geometry problem required four angles to be solved with the last angle being the answer of the problem statement. All angles in the line drawings were coded with letters (e.g., angle A, B, C, etc.) and for some of the angles the value was given whereas for angles with an unknown value a question mark was given.

Eye movement modeling examples. The two EMME videos were created by recording the model's eye movements with a SMI 250Hz remote eye tracker, using SMI Experiment Center software version 3.4.165. Subsequently, SMI BeGaze software version 3.4.52 was used to visualize and integrate the model's eye movements onto the screen-recording of the task. The eye movements of the model were displayed as blue translucent dots with a diameter of 30 pixels. The length of the two EMME videos was 122 s and 132s, respectively. In both videos, a female model verbally explained how to solve the geometry problem. The model started by searching for the angle

mentioned in the problem statement. Once this angle was found the model started working backwards until a starting point was found to solve the problem. The model would then explain and solve the problem step by step until at the end the unknown angle requested in the problem statement was solved. See Appendix A for a screenshot of the geometry problem along with a (translated) transcript of the verbal explanation.

Model introduction. The EMME was preceded by an audio recording in which the female model (same as in the EMME) introduced herself. The audio introduction was similar between conditions with the exception of the key words or sentences that were manipulated. The full introduction was as follows: "*My name is Laura. I am 30 years old and I've been working as a [math/Dutch] teacher for 10 years. As part of a professional development course about the usage of multimedia in education, I've been given the assignment to make instructional videos for math class. So, I'm about to explain and show to you how to solve a geometry problem by means of two video examples. The blue dot in the videos shows you where I am looking while solving the geometry problem. I [know for sure that as a math teacher I can very / hope that as a Dutch teacher I can] clearly explain to you how to solve these geometry problems.*" The length of the audio introduction was 35s for the Expert condition and 34s for the Non-Expert condition.

Test problems. In total six test problems were created: two isomorphic problems and four transfer problems. The two isomorphic problems had a similar visual layout as the problems demonstrated in the video examples, but included different values thus requiring students to apply the same procedure as in the video example by themselves. The transfer problems had different visual layouts and different values.

Questionnaire. To explore students' sensitivity to authority regarding knowledge and learning, we administered a short questionnaire which consisted of the subscales *depend on authority* and *criticize authority* of the Schommer Epistemological Questionnaire (Schommer, 1998). The subscale *depend on authority* consisted of four statements (e.g., "When you first encounter a difficult concept in a textbook, it's best to work it out on your own.") and the subscale *criticize authority* consisted of six statements (e.g., "Often, even advice from experts should be questioned."). Students had to indicate on a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) to what extent they agreed with the statements. See Appendix B for the full list of

statements (which were translated from English into Dutch by the authors).

Quality ratings verbal explanation. Students were asked to rate the quality of the model's verbal explanation of each video on a 5-point Likert scale ranging from 1 (*very* bad) to 5 (*very* good).

Procedure

The experiment was run during a math class and lasted approximately 50 minutes. The students were randomly assigned to one of two conditions in advance of the lesson. On every table a headphone and a sheet of paper stating the student name, participant number, version number and website URL was placed. Students were instructed to find their seat and to take out their laptop. Then, the experimenter gave general instructions about the experiment and answered practical questions of the students. Subsequently, students were instructed to type in the URL and to start to run through the program at their own pace and by themselves (i.e., students were seated separately to prevent collaboration). The program started by asking informed consent, some demographics (gender and age) and presenting the pretest. Following the pretest, students received brief definitions and examples of the different types of angles that were going to be explained in the video examples. In addition, the students were instructed that they were about to see video examples in which they also saw the eye movements of that person visualized as a blue dot. To clarify this visualization an example screenshot was presented. Subsequently, the students were instructed that the person who created the videos would first introduce herself. At this point in the experiment the purported expertise of the model was manipulated as the student either heard the Expert or the Non-Expert audio introduction. After the introduction, the students were presented with the EMME (which were the same in both conditions). After each EMME students rated the quality of the verbal explanation and were then presented with a geometry problem that was isomorphic to the problem demonstrated in the video example. After these two example-problem pairs participants received the four transfer problems. The order of the example-problem pairs and the transfer problems was random. Students' performance was logged by the Qualtrics website on all geometry problems. One week later during the math lesson the students filled out the Schommer Epistemological Questionnaire with

the ten statements of the subscales *depend on authority* and *criticize authority*.

Data Analysis

Due to technical difficulties two participants were unable to hear the audio during the experiment and were therefore excluded from all analyses. For each correctly answered pretest question one point was assigned (range: 0-5). For both the isomorphic and transfer problems one point was given if all four substeps were answered correctly. The proportion correct was calculated for the isomorphic (range: 0-2) and transfer (range: 0-4) problems separately by summing all points and dividing the sum by the number of problems. Two participants were excluded from all analyses as they did not finish the isomorphic problems. In addition, nine students did not calculate the values of the angles but simply stated angle names and were therefore excluded from all analyses. This left 46 participants in the Expert condition and 48 in the Non-Expert condition for the analysis of performance on the isomorphic problems. For the analysis of the performance on the transfer problems an additional 22 participants had to be excluded as they did not manage to finish the transfer problems before the class period finished, leaving 34 participants in the Expert condition and 38 in the Non-Expert condition.

The quality ratings of the verbal explanation of the model in the videos were averaged across both videos for each participant (range: 1-5). The ratings on the items of each subscale of the Schommer Epistemological Questionnaire (i.e., *depend on authority* and *criticize authority*) were averaged for each participant after repooling reversely formulated items so that a high score would indicate a high dependence on authority/high reluctance to criticize authority (range: 1-5). Only the questionnaire data of the 94 participants who were included in the analysis of the proportion correct on the isomorphic problems were included. The questionnaire data of five participants, of which one participant was already excluded from all analyses, was missing because they were absent during the second lesson in which the questionnaire was administered. In addition, one participant left one or more questions unanswered on the subscale *depend on authority* and six participants on the subscale *criticize authority* and were therefore excluded from the analyses. This left 89 participants for the subscale *depend on authority* (Expert condition $n = 43$; Non-Expert $n = 46$) and 84 participants for the

Table 1. Mean (SD) and Median (minimum and maximum) of Performance and Schommer Epistemological Questionnaire Measures of the Expert and Non-Expert Conditions.

	Expert *M* (SD) | *Mdn* (Min.–Max.)	Non-Expert *M* (SD) | *Mdn* (Min.–Max.)
Proportion Correct		
Pretest (*n* = 94)	0.54 (0.26) | 0.60 (0.00–1.00)	0.60 (0.26) | 0.60 (0.00–1.00)
Isomorphic (*n* = 94)	0.66 (0.37) | 0.50 (0.00–1.00)	0.71 (0.35) | 1.00 (0.00–1.00)
Transfer (*n* = 72)	0.33 (0.35) | 0.25 (0.00–1.00)	0.46 (0.33) | 0.50 (0.00–1.00)
Video Quality Rating (*n* = 94)	4.13 (0.59) | 4.00 (3.00–5.00)	4.26 (0.44) | 4.00 (3.50–5.00)
Epistemological Questionnaire		
Depend on Authority (*n* = 89)	3.29 (0.51) | 3.25 (2.25–4.25)	3.11 (0.39) | 3.06 (2.25–4.00)|
Criticize Authority (*n* = 84)	2.53 (0.51) | 2.50 (1.67–3.67)	2.72 (0.41) | 2.67 (1.67–3.50)

subscale *criticize authority* (Expert condition *n* = 41; Non-Expert *n* = 43).

Results

The pretest scores, video quality ratings, questionnaire ratings and problem-solving performance data are presented in Table 1. All data were analyzed with non-parametric Mann-Whitney *U* tests due to the violation of the normality assumption. The measure *r* was used for effect size, with *r* = .10, *r* = .30, *r* = .50, representing small, medium, and large effects respectively (Cohen, 1988). We first checked for prior knowledge differences between the students in both conditions; results indicated no significant differences in prior knowledge, *U* = 961.00, *z* = -1.11, *p* = .267, *r* = -.11.

Learning Outcomes

Performance on the isomorphic problems did not differ between conditions, *U* = 1029.00, *z* = -0.63, *p* = .531, *r* = -.06, nor did performance on the transfer problems, *U* = 496.50, *z* = -1.73, *p* = .084, *r* = -.20.

Quality Ratings Verbal Explanation

To test whether the purported expertise manipulation affected how students perceived the quality of the verbal explanation, we examined whether the average quality rating of the videos differed between the conditions. The results suggest that the quality of the verbal explanation was perceived as high (Expert: M = 4.13, SD = 0.59; Non-Expert: M = 4.26, SD = 0.44), however the ratings did not differ between conditions, U = 970.50, z = -1.11, p = .267, r = -.11.

Schommer Epistemological Questionnaire

The groups did not differ on the subscales *depend on authority* (U = 799.00, z = -1.59, p = .112, r = -.17) and *criticize authority* (U = 695.00, z = -1.68, p = .093, r = .18). We explored whether the problem-solving performance within each condition was related to the scores on these subscales. Due to non-normality Spearman rank correlation analyses were conducted. There were no significant correlations between performance on the isomorphic problems and the two subscales in either condition (Expert condition: *depend on authority*, r = -.24, p = .122; *criticize authority,* r = .21, p = .183; Non-Expert condition: *depend on authority*, r = -.29, p = .055; *criticize authority,* r = .00, p = .991). Similar results were found regarding the transfer problems (Expert condition: *depend on authority*, r = -.16, p = .385; *criticize authority,* r = -.30, p = .114; Non-Expert condition: *depend on authority*, r = -.08, p = .650; *criticize authority,* r = .08, p = .670).

Discussion

The aim of the current study was to examine whether the purported expertise of the model would affect learning from EMME. Based on previous research regarding the effects of social status on attention allocation and learning (Hoogerheide et al., 2016; Cheng et al., 2013; Gobel et al., 2018), we hypothesized that students might devote more attention and thus show better problem-solving performance when observing EMME of a model who is purportedly an expert than a non-expert. Contrary to our hypothesis, there were no differences in learning outcomes between the Expert and Non-Expert EMME conditions.

The fact that the ratings of the quality of the model's explanation were very high and did not differ among conditions, suggests that either our manipulation of purported model expertise was too weak, or that the actual quality of the explanation (in terms of content or in terms of the expertise conveyed in the model's voice) is

the critical factor for learning. That both of these issues might have played a role in our study is suggested by comparing our findings with those of Hoogerheide et al. (2016), who used a similar purported expertise manipulation. They did find an effect of the model's purported expertise on the perception of the quality of the model's explanation, however, they also failed to find an effect on students' learning outcomes. In that sense, our results suggest that the model's purported expertise does not affect learning from EMME (cf. findings on learning from other types of video modeling examples: Van Marlen et al., 2016). The fact that the model's purported expertise does not seem to affect learning might be reassuring for use of EMME in computer-based learning environments, in the sense that this might mean that as long as the quality of the video example is high, it does not matter much who the model is. As for the discrepancy in findings regarding quality ratings, one difference between our study and that of Hoogerheide et al. (2016), which may explain why they did find an effect on quality ratings, is that in their examples the models were visible in the video (and in the introductory cover story), whereas in our study only a voice-over was heard. It is possible that only hearing the model's introduction was not pervasive enough to have an effect lasting long enough to influence the students while they were watching the video examples.

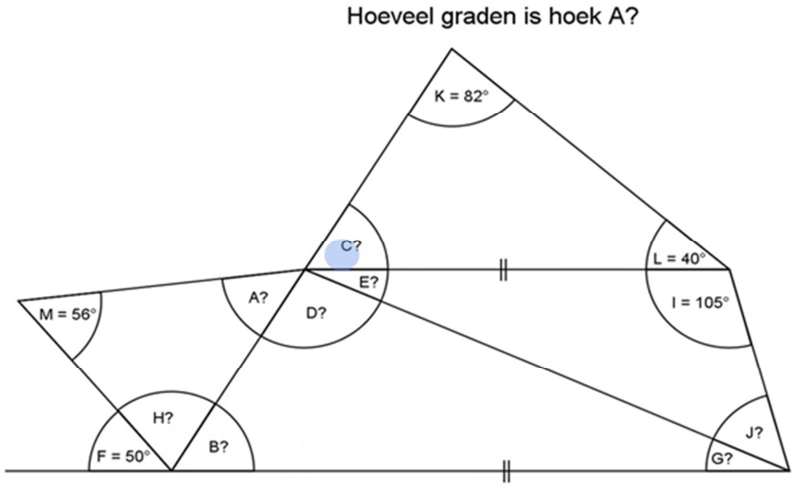
As neither study found an effect on learning outcomes, however (even though students in the Hoogerheide et al. 2016, study perceived the quality of the purported expert model's explanations to be higher), and both studies kept the actual explanations the same across conditions, it seems that either the explanation content or other social cues conveyed by the model's voice may be most important for learning. That is, given the high quality ratings, it is possible that students recognized that the content of the explanations is beneficial to them, regardless of the cover story. It is also possible that another social cue, namely the acoustic properties of the model's voice, conveyed to the students that the model had expertise or was quite confident in teaching the topic. Some support for this second explanation comes from a study by Ko, Sadler, and Galinksy (2015) in which they found that people were judged to be more likely to have a high social rank when their voices were higher in pitch, were louder, and had more variability in loudness. To further investigate these explanations, future research could, for instance, keep the content the same but manipulate the voice-over to make the model in the Non-

Expert condition sound more insecure and hesitant (this would also provide a more pervasive cue regarding the model's expertise throughout the example).

Next to the fact that our manipulation of purported model expertise may have been too weak, another potential limitation is that we did not include eye-tracking to study whether purported expertise affected students' attention. Based on the literature about social status and interpretation of visualizations of other people's gaze (Gobel et al., 2018), we expected that students viewing the Expert EMME would follow the attention allocation of the model (visualized as blue dots) more closely than students viewing the non-expert EMME. It is possible that this was the case, even if that did not result in differences in learning outcomes (cf. studies on cueing that show differential effectiveness of different types of cueing on attention allocation but not on learning outcomes; Van Gog, 2014; De Koning & Jarodzka, 2017). Using eye-tracking in future research would not only shed light on this matter but could also provide an indirect manipulation check to test whether the students indeed perceived the model as having more/less expertise.

To conclude, in the current study we found no support that – keeping all else equal- the purported expertise of the model would influence learning from EMME. When this result is not due to a manipulation failure it would seem to indicate that it does not matter whether the model is purportedly an expert as long as the content of the EMME is clear and perceived to be of high quality. This would be good news as eye trackers are becoming cheaper and therefore can be expected to become more common in schools and even in households (e.g., in gaming equipment). Thus, it is only a matter of time before 'teacher-made' or 'home-made' EMME videos will start making an appearance online, leading to wider variation in models' social status.

Appendix A. Screenshot of an EMME used in the current study 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 this angle. 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 it right now, because the second angle is unknown. The second angle 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, because the right angle is also unknown, it is not possible to calculate this angle no. You cannot calculate the right angle directly, but it can be derived from this angle, because these are equal. This can be seen by this 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 these angles are equal. This angle is unknown for now but can be calculated. This angle equals 180 degrees minus the known angles, equals 58 degrees. Now this angle is known, you know that the other angle, by means of the corresponding angle principle, also equals 58 degrees. With this angle known, you can now calculate the other angle. This angle equals 180 degrees minus the known angles, equals 72 degrees. With that angle 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."*

Appendix B. The items of the subscales *depend on authority* (DEPEND) and *criticize authority* (CRIT) of Schommer Epistemological Questionnaire with the original English statements and the Dutch translations.

DEPEND 1. Whenever I encounter a difficult problem in life, I consult with my parents/Wanneer ik een moeilijk probleem tegenkom in het leven, vraag ik mijn ouders om raad.

DEPEND 2. When you first encounter a difficult concept in a textbook, it's best to work it out on your own/Wanneer je een moeilijk begrip tegenkomt in een tekstboek, kun je het beste proberen om er zelf uit te komen.

DEPEND 3. How much a person gets out of school mostly depends on the quality of the teacher/Wat je leert op school hangt vooral af van de kwaliteit van de leraar.

DEPEND 4. Sometimes you just have to accept answers from a teacher even though you don't understand them/Soms moet je gewoon aannemen wat je leraar zegt, ook al begrijp je het niet.

CRIT 1. Often, even advice from experts should be questioned/Vaak moet je zelfs het advies van experts in twijfel trekken.

CRIT 2. You should evaluate the accuracy of information in a textbook, if you are familiar with the topic/Als je bekend bent met het onderwerp, zou je moeten nagaan of de informatie in een tekstboek klopt.

CRIT 3. You can believe almost everything you read/Je kan bijna alles geloven wat je leest.

CRIT 4. People who challenge authority are over-confident/Mensen die opkomen tegen het gezag zijn overmoedig.

CRIT 5. I often wonder how much my teachers really know/Ik vraag me vaak af hoeveel mijn leraren echt weten.

CRIT 6. For success in school, it's best not to ask too many questions/Om succesvol te zijn op school is het beter om niet teveel vragen te stellen.

Chapter 6

Looking through Sherlock's eyes: Effects of eye movement modeling examples with and without verbal explanations on deductive reasoning

This chapter is submitted for publication as:
Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., Raijmakers, M., & Van Gog, T. (2018). *Looking through Sherlock's eyes: Effects of eye movement modeling examples with and without verbal explanations on deductive reasoning* Manuscript submitted for publication.

Acknowledgement of author contributions: TM, MW, HJ, MR and TG designed the experiments, TM recruited participants and collected the data, TM analyzed the data, MW checked the data package, TM drafted the manuscript, all authors contributed to critical revision of the manuscript, MW, HJ, and TG supervised the experiments.

Abstract
Eye movement modeling examples (EMME) are demonstrations in which learners' not only see a model's (e.g., a teacher's) task performance on a computer screen (as in regular video examples) but also the model's eye movements (represented as moving colored dots overlaid on the screen). Thereby EMME help guide learners' attention towards the relevant information and can model cognitive strategies which are otherwise unobservable for learners. This study investigated whether EMME can help to learn deductive reasoning strategies in Mastermind. Secondary education students (N=137) were randomly assigned to study video examples under one of four conditions in a 2 (EMME: yes/no) x 2 (verbal explanations: yes/no) between-subjects design. Results revealed only a beneficial effect of the presence of verbal explanations on performance on the practice problems, but no pretest-to-posttest learning gains. Our results suggest that EMME do not promote learning of deductive reasoning strategies.

Introduction

Studying video modeling examples is an effective way to learn new skills (Van Gog & Rummel, 2010). One specific type of video modeling examples, referred to as 'eye movement modeling examples' (EMME; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009), consists of a screen recording that shows the learner a demonstration by a model (e.g., a teacher or an expert) of how to perform a task, while simultaneously visualizing showing where the model was looking during this task demonstration (e.g., as a colored dot or circle). EMME may or may not include a verbal explanation by the model (as a voice-over). EMME can serve two different functions. First, they align the attention of the learner with the attention of the model, which may help the learner to make more sense of the model's demonstration (and explanation, when present) than regular modeling examples (ME; i.e., the same examples but without the visualization of eye movements). Second, by making the model's eye movements visible, EMME allow for modeling perceptual and cognitive strategies that would otherwise remain unobservable for learners. This second function has received less attention, but prior research suggests that EMME can be effective for modeling perceptual/cognitive strategies, such as learning to classify locomotion patterns (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013) and epileptic seizure symptoms (Jarodzka et al., 2012), or learning study strategies such as text-picture integration (Mason, Pluchino, & Tornatora, 2015; 2016; Mason, Scheiter, & Tornatora, 2017; Scheiter, Schubert, & Schüler, 2018), or multiple document integration (Salmerón & Llorens, 2018). The aim of the current study was to investigate whether EMME can foster the learning of cognitive problem solving strategies and whether this depends on the absence or presence of the model's verbal explanation.

Eye Movement Modeling Examples

As stated above, EMME are video examples in which you see a screen recording of a model solving a task while also seeing where the model was looking at during task performance. EMME can serve two functions. The first function of EMME is that they align the learner's attention with that of the model, which then can help the learner to select and integrate the relevant information of the task demonstration. Compared to regular modeling examples (ME; i.e., video examples of screen recordings without the model's eye

movements superimposed) in EMME the learner's attention is directed and synchronized with that of the model, thus creating a state of *joint attention* (i.e., *joint attention* is characterized as the phenomenon of automatically attending an object someone else is attending; Brennan, Chen, Dickinson, Neider & Zelinsky, 2008; Frischen, Bayliss, & Tipper, 2007). The creation of joint attention between the model and the student is important as information within video examples are often transient (i.e., information is only temporarily available) meaning that students can miss out on important information when they do not attend to the right information at the right time, which can negatively affect students' learning (Ayres & Paas, 2007). By offering attentional guidance by means of EMME, the risk of not attending important information can be reduced. In this way, the learner's processing of visual information (i.e., the on screen learning material with visible interactions –e.g., clicks, drags, typing- of the model with the material) and visual-verbal information (i.e., on screen learning material with the model's verbal explanation) can be facilitated.

That the attentional guidance provided by EMME affects the learner's visual attention, aligning it more with the model's visual attention, and also enhances learning, is supported by research. Several studies have compared the effects of EMME with ME on attention allocation (by measuring the learner's eye movements during example study) and learning outcomes (i.e., posttest performance; Jarodzka et al., 2012; 2013; Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016; 2018). For instance, the studies by Jarodzka et al. (2012; 2013) demonstrated that EMME compared to ME enhanced learning to perform classification tasks. Also they found that the learner's eye movements while watching the video examples in the EMME conditions were more similar to those of the model than in the ME condition, as evidenced by a higher scanpath similarity (Jarodzka et al., 2013) or smaller Euclidean distances between the model's gaze position and that of the learner (Jarodzka et al., 2012). In addition, Van Marlen et al. (2018) have recently demonstrated that college students watching EMME fixated on the verbally referred visual task elements more often and faster than students watching ME.

The second function of EMME is that they make it possible to visualize perceptual and cognitive strategies that would otherwise remain unobservable for learners. This can be done either in the presence or absence of the model's verbal explanation. EMME

without the model's verbal explanation have been shown to enhance study strategies for digital hyperlinked texts (Salmerón & Llorens, 2018) and illustrated texts (in seventh grade students: Mason et al., 2015; 2016; 2017; in college students: Scheiter et al., 2018). More specifically, the studies by Mason et al. (2015, 2016, 2017) examined whether observing EMME prior to studying an illustrated text, would enhance text picture integration during study. In the EMME, the model demonstrated how to integrate information from the text and picture by making transitions between certain terms in the text and the corresponding part of the picture. Compared to students who did not observe EMME, students in the EMME condition showed better text-picture integration while studying the (new) illustrated text, and also performed better on a text comprehension test.

That EMME without verbal explanations can also model perceptual strategies was demonstrated in studies showing that EMME enhance visual search performance when people had to search for errors in software code (Stein & Brennan, 2004), errors on printed circuit boards (Nalanagula, Greenstein, & Gramopadhye, 2006) or lung-nodules on X-ray scans (Litchfield, Ball, Donovan, Manning, & Crawford, 2010).

Other research has shown that EMME *with* verbal explanations can be used to model inspection strategies aimed at enhancing to learn to classify in classification tasks (Jarodzka et al, 2012; 2013; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012) and can model procedural problem-solving strategies to enhance problem solving (Experiment 2, Van Marlen et al., 2018). For instance, in the study of Van Marlen et al. (Experiment 2, 2018) EMME demonstrated procedural problem-solving strategies about solving geometry problems. In this study students watched EMME in which the students saw the model making transitions between elements of the problem (i.e., the different angles that had to be solved) while verbally explaining the underlying principles to solve the angles.

In contrast, there is also research in which EMME demonstrated procedural problem-solving strategies but were not effective to enhance learning even though the model's verbal explanations were included (Van Gog et al., 2009). For instance, in the study by Van Gog et al. (2009) students watched EMME or ME about how to solve a procedural puzzle problem with or without the model's verbal explanations. The procedural puzzle problem was a mathematical problem about getting puzzle pieces (e.g., frogs) from

one side towards the other side. The model in the video examples verbally explained from the start till the end of the problem how to solve this problem. Results revealed no positive effect of EMME on learning. However, on the transfer task it was even found that students who watched EMME including the model's verbalization had lower performance than students who watched the EMME without the model's verbalization. An alternative explanation given by the authors for these findings was that perhaps the verbal explanations were already sufficient for the students to guide their attention making EMME redundant. This suggests that the model's verbalizations might play an important role in learning from EMME.

In sum, some studies (Jarodzka, et al. 2012; 2013; Marlen et al., 2018; Vitak et al., 2012) found positive effects of EMME conveying perceptual/cognitive strategies on learning if the EMME included verbal explanations, whereas other studies found a negative effect (Van Gog, 2009) or no effect (Van Marlen et al., 2016) of EMME with verbal explanations on learning. Drawing conclusions about the role of the model's verbalizations on learning from EMME remains difficult because, with the exception of Van Gog et al. (2009), none of the discussed studies have manipulated the presence or absence of verbal explanations. Therefore, in the current study we aimed to examine whether the presence or absence of verbal explanations affects learning deductive reasoning strategies from EMME.

The Present Study

The aim of the present study was to examine whether EMME would be effective for fostering learning of cognitive strategies (more specifically: deductive reasoning) and whether this would be affected by the presence or absence of the model's verbal explanations. Secondary education students were presented with video modeling examples in which the model demonstrated how to break the code in a Deductive Mastermind task, either with or without verbal explanation. The Deductive Mastermind task (cf. Gierasimczuk, Van der Maas, & Raijmakers, 2013) is an adapted version of the classic board game Mastermind. A code has to be deduced from a set of code breaking attempts. In the Deductive Mastermind task, the code-breaker is provided with an image depicting several code-breaking attempts and corresponding feedback, which the code-breaker must use to deduce the correct code by systematically comparing the entered codes with its corresponding feedback. In the EMME, the cognitive strategies involved in systematically comparing the entered

codes and feedback become visible through the depiction of the model's eye movements.

We expected that the students in the EMME conditions would benefit from seeing the model's problem-solving strategies and would therefore show greater learning gains from pretest to posttest and transfer problems than students in the ME conditions. In addition, we expected students in the verbal explanation conditions to show higher learning gains than students in the no verbal explanation conditions. Even though Van Gog et al. (2009) found that the presence of a verbal explanation presumably made the attention guidance provided by the EMME redundant, we expected that the presence of verbal explanations would further enhance learning from EMME in the present study, as the attention guidance provided by the EMME would complement the strategy information (i.e., the model demonstrated multiple deductive reasoning strategies which had to be combined to solve the problems) conveyed in the verbal explanations.

Method

Participants

Hundred-forty Dutch secondary education students in their first year of pre-university education (the track that prepares students for enrollment at a university, and has a six-year duration) were recruited out of five classrooms. Three students did not provide informed consent to use their data and were therefore excluded from the study. The final sample consisted of 137 students (M_{age} = 12.66, *SD* = 0.51, 64 male). Participants were randomly assigned to one of four conditions resulting from a 2 (modeling example: EMME vs. ME) x 2 (verbal explanation: present vs. absent) between-subjects design: EMME verbal explanation present (*n* = 34), EMME verbal explanation absent (*n* = 35), ME verbal explanation present (*n* = 33), ME verbal explanation absent (*n* = 35).

Materials

Deductive Mastermind task. The deductive reasoning task used in the current study is an adaptation of the board game Mastermind. This Deductive Mastermind task is part of the popular online learning platform called 'Math Garden' (in Dutch 'de Rekentuin' www.mathgarden.com; Gierasimczuk et al., 2013). In the digital Deductive Mastermind game the learner plays the role of a code-

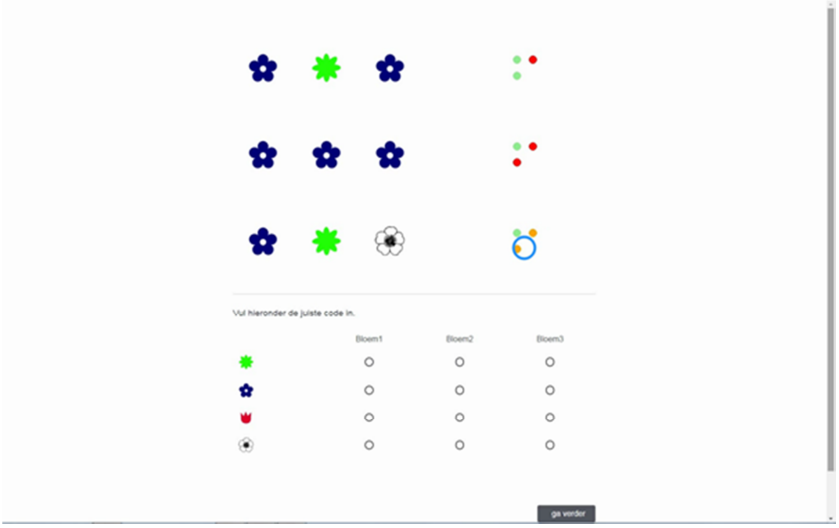


Figure 1. A screenshot of an eye movement modeling example depicting the Mastermind task. The blue circle represents the gaze location of the model, in this case inspecting the feedback of the third code-breaking attempt. On top the three rows with the flower codes along with the corresponding feedback on the right are displayed. Underneath the codes the answer selection pane with the different possible flower types are displayed.

breaker. The task for the learner (code-breaker) is to unravel the correct code. To do so, the learner is shown an image depicting previously entered codes with the corresponding feedback (see Figure 1). The learner must use the feedback of all previous code breaking attempts, to enter the correct code, and code-breaker is given only one chance to break the code. Hence the term 'Deductive Mastermind', as the learner has to deduce the correct code by processing all the feedback of earlier code breaking attempts. In the current study, the colored pins of the deductive Mastermind task were represented as colored flowers. In addition, the feedback pins were represented as colored dots: a green dot for correct flower and location, an orange dot for correct color only (i.e., wrong location), and a red dot for stating that a flower does not occur in the code.

Mastermind test problems. The Mastermind problems used in the present experiment differed in terms of the number of flowers in the code (two, three, or four flowers) and the number of code-breaking attempts being shown (i.e., this ranged between two to five attempts). The pretest, consisted of six problems in which the code consisted of two flowers and six problems in which the code consisted of three flowers resulting in a total of twelve problems. Based on difficulty

ratings obtained from the database of ‘Math Garden’ (Gierasimczuk et al., 2013) we selected problems of which half of the two flower codes and half of the three flower codes were rated as easy and the remaining half as difficult.

For the two video examples we used Mastermind problems that consisted of three flower codes, each showing three code breaking attempts. After each video example students were presented with the opportunity to practice applying the modeled strategy themselves on an isomorphic practice problem (created by replacing the flowers with different flowers so that the task looked different, but it was structurally identical).

Twelve posttest problems were created by replacing the flowers of the pretest problems with different types of flowers, so that the posttest problems looked different from the pretest problems but were otherwise identical. See Figure 1 for an example.

Finally, six transfer problems were created, which were more complex than the pretest/posttest problems and consisted of four flower codes. Within the four-flower code category half of the transfer problems were rated as easy and the other half as difficult based on the ‘Math Garden’ database.

Eye movement modeling examples. A SMI 250Hz remote eye tracker (SensoMotoric Instruments, GmbH) was used to create the two EMME videos. SMI Experiment Center 3.7.60 software to display the tasks and iViewX 2.8 software was used for recording the model’s eye movements while the model completed the tasks. After recording, SMI BeGaze 3.7 software was used to visualize the model’s eye movements in the videos of the screen-recording of the task. The eye movements were represented as a moving blue colored circle with a diameter of 30 pixels and a line width of 3 pixels.

In the EMME videos, the male model started by inspecting the flower codes and the corresponding feedback from top to bottom. Once all the codes and feedback were inspected, the model made transitions between two different lines of code and the corresponding feedback, hereby implying that the comparison of these two lines of code and feedback would provide information regarding the correct code. In the conditions in which verbal explanations were also present, the model was then explaining strategies regarding what could be deduced from these lines of code and feedback. In the video examples three strategies were demonstrated. The first strategy is the least difference strategy. In this strategy, two rows of code are

compared in which only one flower in the code differs from the other code. By looking at the corresponding feedback, you can deduce whether this single difference in the code indicates whether a particular flower belongs in the code. The second strategy being demonstrated in the video examples is the integration of knowledge. By this we mean that the knowledge of a part of the code is further used to deduce later part(s) of the code. For example, knowing that the leftmost flower of the code must be a green flower instead of a blue flower might enable you to deduce that the blue flower must be placed on a different location in the code. The third strategy concerns the usefulness of feedback. If the feedback of code breaking attempt consist of only red pins, then you can deduce that no flower in that code is present in the final code. The other way around, if the feedback pins consist of green and orange pins, then you can deduce that at least you know which color flowers must be present in the final code. One video example demonstrated all strategies and the other video example demonstrated the least difference strategy and the integration of knowledge strategy. The least difference strategy was needed in nine pretest/posttest problems and in five transfer problems, the integration of knowledge strategy was needed in seven pretest/posttest problems and in three transfer problems, and the usefulness of feedback strategy was needed in seven pretest/posttest problems and in five transfer problems.

In the EMME, the model made comparisons (i.e., comparisons between lines of code and corresponding feedback) until the full code was deduced. Every solved part of the code, or sub step of the problem, was entered immediately by clicking on the correct answer option underneath the problem and was visible in the screen recording of the task. The model behaved didactically, so the eye movements between the code and feedback were very deliberate. The length of the two EMME videos was 135 s and 139 s, respectively. For the regular modeling example conditions (ME conditions) the screen recordings were exported without the eye movements superimposed. The screen recordings were exported with or without the verbal explanations. Therefore, the videos across conditions were equal regarding the screen recording and length and only differed regarding the presence/absence of the model's eye movements and presence/absence of the verbal explanation. See Appendix for the translated transcript of the verbal explanation about how to solve the Mastermind problem depicted in Figure 1.

Procedure

The experiment was conducted during a math lesson that lasted approximately 50 minutes. The tables in the classroom were separated to ensure the students would not collaborate or look at each other's laptop screens. On each table a sheet of paper was placed stating the student name, participant number, version number, headphone (if a student was assigned to a condition including verbal explanations) and a website URL. As the students entered the classroom they were instructed to find their table and to take out their laptop. Once every student was seated, the experimenter gave general instructions about the experiment and answered practical questions. After these instructions the students were asked to type in the URL, which opened the online questionnaire used for the experiment hosted by Qualtrics (www.qualtrics.com) and to type in the given participant number and experiment version. Subsequently, the students were instructed to start with the online program at their own pace. The program then started by asking the informed consent followed by demographic questions (age and gender). Students then received instructions about the Mastermind task stating that the students had to break the code by using the feedback provided for each code-breaking attempt. The students were explained the meaning of the different types of feedback (i.e., what the different colored pins meant). However, how the students should use the feedback to unravel the code was not explained. After the Mastermind instructions, students made the twelve pretest problems. Then the students were presented with the video examples. They received the instruction that they were about to see how someone else solved a Mastermind problem, and only students in the EMME condition were additionally instructed that the blue circle showed them where the person had looked while solving the problem. Students in the conditions with verbal explanations in the examples were asked to check whether the headphone was connected properly and the volume was on. Then the students watched the video example, followed by the corresponding isomorphic problem, and the second video example followed by the corresponding isomorphic problem. Subsequently, students received the twelve posttest problems followed by the six transfer problems. The order of the problems in the pretest, the video examples, the problems of the posttest, and the transfer problems was random. The answer options selected by the students were registered by the Qualtrics software.

Data Analysis

One student was unable to finish the pretest due to technical problems, and was therefore excluded from all analyses. In addition, eight students were unable to finish the isomorphic practice problems in time, 25 additional students were unable to finish the posttest in time, three students were considered outliers on the posttest (i.e., absolute *z*-score larger than 2.5), and an additional 14 students did not finish the transfer problems in time. Table 1 shows the total number of participants included in the different types of analyses.

One point was given for each correctly solved Mastermind problem. In total students could earn 12 points for the pretest, 12 points for the posttest, 2 points for the isomorphic problems and 6 points for the transfer problems. For the pretest, isomorphic problems and transfer problems the proportion of correctly solved problems was calculated by dividing the number of points by the maximum obtainable points. To measure the student's progression from pretest to posttest, we calculated a gain score. For each participant a gain score was calculated as the number of points earned in the posttest minus the number of points earned in the pretest (posttest-pretest).

Results

The performance data of the pretest problems, isomorphic problems, transfer problems, and the gain scores are presented in Table 1. The data were analyzed with 2 (modeling example: EMME vs. ME) x 2 (verbal explanation: present vs. absent) ANOVAs [7] and partial eta squared is reported as a measure of effect size, with $\eta_p^2 = .01$, $\eta_p^2 = .06$, $\eta_p^2 = .14$, representing small, medium, and large effects respectively (Cohen, 1988). We first examined whether the prior knowledge measured as the performance on the pretest was equal across conditions. Results of this 2 x 2 ANOVA with the

[7] Due to the non-normal distribution of the performance data, we additionally conducted non-parametric Kruskal-Wallis tests to examine whether the main effects from the 2x2 ANOVA would hold. For the isomorphic problems, we found that EMME outperformed the ME condition, $H(1) = 12.89$, $p < .001$, and there was no difference between the ambiguous instruction condition and the unambiguous condition, $H(1) = 0.68$, $p = .410$. Similar results were found for the transfer problems, EMME outperformed the ME condition, $H(1) = 4.15$, $p = .042$, and there was no difference between ambiguous instruction condition and the unambiguous condition, $H(1) = 1.38$, $p = .241$. In sum, the results of the non-parametric tests are in line with the results regarding the main effects of the 2x2 ANOVA.

Table 1. Mean (and SD) and Median (and min–max) for the Pretest, the Gain score (posttest-pretest), and the Transfer problems for the four conditions: Modeling Example (EMME vs. ME) x Verbal Explanations (Verbal Explanations vs. No-Verbal Explanations).

	EMME		ME	
	Verbal Explanations	No-Verbal Explanations	Verbal Explanations	No-Verbal Explanations
	M (SD) | *Mdn* (min–max)	*M* (SD) | *Mdn* (min–max)	*M* (SD) | *Mdn* (min–max)	*M* (SD) | *Mdn* (min–max)
Proportion Correct				
Pretest (*n*=136)	0.43 (0.29) | 0.33 (0.00–0.83)	0.40 (0.32) | 0.33 (0.00-1.00)	0.39 (0.30) | 0.42 (0.00-0.92)	0.34 (0.26) | 0.25 (0.00-0.83)
Isomorphic (*n*=128)	0.65 (0.42) | 1.00 (0.00–1.00)	0.34 (0.37) | 0.50 (0.00–1.00)	0.55 (0.42) | 0.50 (0.00–1.00)	0.31 (0.38) | 0.00 (0.00–1.00)
Posttest (*n*=100)	0.51 (0.26) | 0.58 (0.00–0.92)	0.34 (0.28) | 0.25 (0.00–0.92)	0.47 (0.28) | 0.46 (0.00–0.83)	0.38 (0.29) | 0.38 (0.00–0.83)
Transfer (*n*=87)	0.23 (0.22) | 0.17 (0.00–0.67)	0.15 (0.17) | 0.17 (0.00–0.50)	0.14 (0.19) | 0.00 (0.00–0.50)	0.20 (0.21) | 0.17 (0.00–0.67)
Gain score (*n*=100)	0.63 (1.95) | 0.00 (-2.00–4.00)	0.46 (2.49) | 0.00 (-4.00–6.00)	0.46 (1.77) | 0.00 (-3.00–4.00)	0.50 (2.30) | 0.50 (-5.00–6.00)

the proportion correct of the pretest as dependent variable indicated no main effect of modeling example, F (1, 132) < 1.00, p = .341, no main effect of verbal explanation, F (1, 132) < 1.00, p = .380, and no interaction, F (1, 132) < 1.00, p = .829.

Performance Isomorphic Practice Problems

To test our hypothesis that students would learn more from an EMME than a regular modeling example and whether learning was affected by the presence/absence of verbal explanations, a 2 x 2 ANOVA with the proportion correctly solved isomorphic problems as the dependent variable was conducted. The results revealed no main effect of modeling example, F (1, 124) < 1.00, p = .375, η_p^2 < .01, a main effect of verbal explanation, F (1, 124) = 15.04, p < .001, η_p^2 = .11, indicating that students presented with video examples that also included verbal explanations outperformed students who did not receive verbal explanations. There was no significant interaction, F (1, 124) < 1.00, p = .607, η_p^2 < .01.

Pretest to Posttest Gain Score

To test whether students with EMME showed larger learning benefits than students with the regular video examples and to examine whether this is influenced by the presence/absence of verbal explanations, a 2 x 2 ANOVA with the gain score as the dependent variable was conducted. The results revealed no main effect of modeling example, F (1, 96) < 1.00, p = .882, η_p^2 < .01, no main effect of verbal explanation, F (1, 96) < 1.00, p = .888, η_p^2 < .01, and no significant interaction, F (1, 96) < 1.00, p = .813, η_p^2 < .01.

Performance Transfer Problems

A 2 x 2 ANOVA with the proportion correctly solved transfer problems as dependent measure was conducted to examine whether students in the EMME conditions outperformed students in the ME conditions on the transfer problems and whether this was influenced by the verbal explanations. The results revealed no main effect of modeling example, F (1, 82) < 1.00, p = .674, η_p^2 < .01, no main effect of verbal explanation, F (1, 82) < 1.00, p = .783, η_p^2 < .01, and no significant interaction, F (1, 82) = 2.56, p = .114, η_p^2 = .03.

Discussion

The aim of the present study was to examine whether EMME would be effective for fostering learning of cognitive strategies and whether this would be affected by the presence or absence of the model's verbal explanations. Thus, the present study set out to

conceptually replicate Van Gog et al. (2009), with a deductive reasoning task. We hypothesized that the students in the EMME conditions would show greater learning gains from pretest to posttest than students in the regular modeling example conditions (i.e., video modeling example without the model's eye movements superimposed). In addition, we expected that EMME with verbal explanations would be more effective than all other conditions (i.e., interaction effect) as the attention guidance provided by the EMME would complement the strategy information conveyed in the verbal explanations.

In contrast to our hypothesis, we found no effect of attention guidance: there were no differences in learning outcomes between the EMME and ME conditions. The only exception is the finding that students performed better on the isomorphic practice problems if the modeling example included verbal explanations. However, this benefit of having heard the model's verbal explanations during the video examples did not translate into a higher learning gain score or better performance on the transfer problems. Also, in contrast to our hypothesis we found no interaction between the type of modeling example (EMME vs. ME) and the presence/absence of verbal explanations. Thus, although we did not find a negative effect of EMME with verbal explanations on learning as found by Van Gog et al. (2009), we also did not find the positive effects of EMME with verbal explanations often found with classification tasks (Jarodzka et al., 2012; 2013; Vitak et al. 2012).

A possible explanation for the lack of differences in learning outcomes between the conditions might be due to the difficulty of the task. Support for the assumption that students experienced the task to be difficult stems from the fact that across conditions the gain score was below one. This means that from pretest to posttest, students hardly improved in solving the mastermind problems even though on average, students in the verbal explanation conditions solved more than half of the isomorphic problems. This suggests that students were able to apply the strategies that were just demonstrated in the example, on an isomorphic practice problem, but could not apply what they had learned to solve the posttest problems. This is further underlined by the low performance on the transfer problems. Thus, taken as a whole, our data suggest that students were unable to abstract the cognitive strategies being demonstrated by the model, irrespective of whether they were accompanied by verbal

explanations. Possibly, students would need more example-problem pairs in the learning phase to practice how to apply the strategies. The use of two example-problem pairs in the learning phase is similar to a previous EMME study that found EMME to enhance learning of geometry problem solving (Van Marlen et al., 2018). However, the current study differs with respect to the task being demonstrated. Perhaps the underlying cognitive strategies involved in the current task are too difficult to grasp with only two example-problem pairs, especially since the students did not receive feedback. An interesting research avenue could be to see whether students would learn more if the learning phase consisted of video examples with the corresponding isomorphic practice problems followed by feedback on their performance, for instance by means of a correct demonstration (i.e., video example) of the isomorphic practice problem.

The finding that students were unable to abstract and/or apply the modeled cognitive strategy for solving the posttest and transfer problems, is at odds with research regarding enhancing text and picture integration with EMME (Mason et al., 2015; 2016; Mason, Scheiter, & Tornatora, 2017). However, in these studies the modeled strategy was arguably less complex: the model in the EMME demonstrated how to read and process texts by making transitions between key concepts in the text and the corresponding elements in the picture (and this resulted in higher performance on text comprehension tests compared to a control condition). In the present study, the model in the EMME conditions emphasized how to compare and deduce the correct code by making transitions back and forth between specific parts of the code and the corresponding feedback. However, in the current study the function of EMME was not only to indicate which information sources needed to be integrated but also to convey the underlying cognitive strategies involved to solve the deductive reasoning problems. It is possible that the combination of both attending the video example while also trying to understand the underlying cognitive strategies was too demanding for the students.

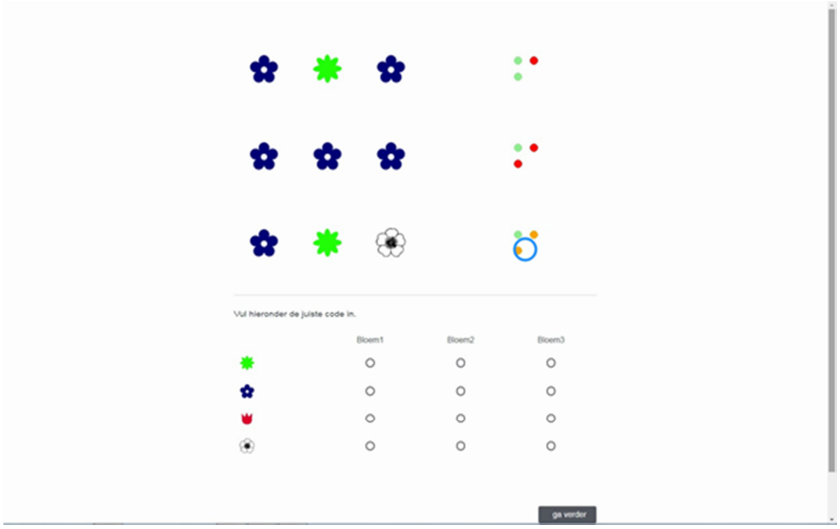
A study by Scheiter et al. (2018) also illustrates that students may not be able to learn from EMME when these are too demanding. Their findings suggest that in order to fully benefit from the attentional guidance in EMME the student needs to have a minimal amount of cognitive prerequisites (i.e., more broad knowledge regarding scientific thinking enabling the student to understand instructions, thus not to be confused with specific prior knowledge). In their study,

students either watched EMME demonstrating how to process multimedia learning materials regarding mitosis or students were given an equivalent amount of time to read the text themselves in the control condition. Results showed that on learning outcomes that required more in-depth processing of the learning materials actually the stronger students (i.e., with higher cognitive prerequisites) benefitted more from having watched EMME compared to weaker students. Similar results were also found in an EMME study regarding medical image diagnosis in which radiologist vs. medical residents seemed to benefit more from having watched EMME (Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017). Thus, perhaps the students in the present study did not have the necessary cognitive prerequisites to use the guidance in the EMME to their full potential. For future research it would therefore be interesting to test an older sample, and/or to take measures of cognitive abilities into account when investigating EMME.

Besides including broader measures of cognitive ability, the addition of eye tracking would also be informative for future research. Because the present study was conducted in the classroom, we were not able to measure the students' eye movements while they were watching the video examples. Although many studies regarding EMME found that the student's visual attention allocation was affected by EMME in terms of fixating relevant information more often, faster and longer (Jarodzka et al., 2013; Mason et al. 2015; Scheiter et al., 2018; Van Marlen et al., 2016; 2018) in the current study we cannot know for sure whether the EMME affected the visual attention allocation of the students. It is possible that they attempted to engage in solving the problem shown in the example themselves, without following the model's gaze, in which case the attentional guidance provided by EMME would not be very useful.

To conclude, EMME may not be effective for learning to solve deductive reasoning problems, regardless of whether or not the gaze guidance is combined with verbal explanations. However, before being able to draw a definitive conclusion on the usefulness of EMME for acquisition of deductive reasoning strategies, further research is needed that includes students with higher cognitive prerequisites, and measures their eye movements during example study.

Appendix A. A screenshot of the EMME with the translated (from Dutch) transcript of the verbal explanations.



"*Here we see the Mastermind task and as you can see the code consists of three flowers. You see that three attempts have already been made and to the right of every attempt you also see the feedback. Let's start to crack this code. First thing that stands out is the third attempt. Here you see that the feedback consists of one green dot and two orange dots. The green dot means that one flower is in the correct location and the orange dots mean that there are two flowers that are in the code but are not yet placed in the correct location. Thus, this third attempt indicates which flowers must be present in the code. The second thing that stands out when looking at the first and second attempt is that only one flower changed, namely the middle green flower in the first attempt changes into a blue flower in the second attempt. Subsequently, when we look at the feedback, then you see that first there were two green dots and at the second attempt there is only one green dot. Thus, from two correct flowers we went to one correct flower. So, we can deduce that the middle flower must be a green flower. Now that we know this flower, we can figure out the rest with the third attempt. That is, there is one flower in the correct location and two are not yet placed in the correct location. But now we know that the middle flower is green, so this is the correct flower. This means that the blue flower and the white flower are not yet correctly placed. Thus, these have to be switched. So the first flower of the code should not be the blue flower but should be the white flower and the third flower of the code should not be the white flower but should be the blue flower. So finally, the code should be white flower, green flower, blue flower.*"

Chapter 7
Summary and Discussion

In recent years, studies have been accumulating evidence that *eye movement modeling examples* (EMME) can be used to guide the learner's attention and enhance learning and performance of various tasks. EMME are video examples in which learners not only see the model's (e.g., a teacher's) task performance on a computer screen but also a visualization of the model's eye movements (i.e., colored circle or dot) overlaid on the screen recording showing the learner where the model was looking during task performance (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). Often EMME also contain a verbal explanation by the model of what s/he is doing, although this does not necessarily have to be the case (e.g., EMME did not contain a verbal explanation in the studies by Mason, Pluchino, & Tornatora, 2015; 2016; Mason, Scheiter, & Tornatora, 2017; Salmerón & Llorens, 2018; Scheiter, Schubert, & Schüler, 2018).

EMME have been shown to improve performance on visual search tasks (Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Nalanagula, Greenstein, & Gramopadhye, 2006; Stein & Brennan, 2004). Moreover, EMME also improved learning (i.e., later performance on similar tasks in the absence of the guidance) of strategies for integrating textual and pictorial information (Mason et al., 2015; 2016; 2017; Salmerón & Llorens, 2018; Scheiter et al., 2018), and learning of classification tasks (Gegenfurtner, Lehtinen, Jarodza, & Säljö, 2017; Jarodzka et al., 2012; 2013; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012). However, a study in which EMME were used to learn how to solve a procedural puzzle problem found a negative effect of EMME when combined with the model's verbal explanation (Van Gog et al., 2009).

Considering the fact that EMME have proven effective for several different kinds of tasks, and that the lack of evidence on effectiveness of EMME for learning procedural problem-solving tasks was limited to a single study, it would be premature to conclude that EMME are not suitable to be used to enhance learning in procedural tasks. Given that students have to acquire problem-solving skills in many subjects (e.g., math, science, and chemistry) at school, it is important to examine whether and under which conditions EMME can also be used to foster learning of procedural tasks. Therefore, the central question addressed in this dissertation is whether EMME guide the learner's attention and enhance learning of procedural problem-solving tasks. Five empirical studies, with a total of eight experiments,

were conducted to answer the main research question: *Do EMME enhance learning of procedural problem-solving tasks?* Together, these studies also make a first step towards shedding light on several possible boundary conditions that could affect learning procedural problem-solving tasks from EMME, by using tasks of different complexity levels, tasks in which the visualizations of the model's eye movements have a different function, tasks in which the model's verbal explanations are clear versus ambiguous, and by studying effects of EMME with learners of different ability levels.

In this Chapter, a summary of the main findings of the studies is provided, followed by a discussion of those findings. Subsequently, implications for practice and research are discussed.

Summary of Main Findings

In the study reported in **Chapter 2**, it was investigated whether EMME affected attention and enhanced learning how to solve a procedural geometry problem-solving task. In Experiment 1, university students watched video examples in which a model demonstrated how to solve single-step geometry problems, either without the model's eye movements, with the model's eye movements visualized, or with a random pattern of eye movements (this last condition was added to examine whether we could conceptually replicate the findings of Litchfield et al., 2010, showing that only task relevant eye movements improved concurrent task performance, to rule out that the effect occurred because the mere presence of visualized eye movements itself would make students more attentive). Results revealed no differences in learning outcomes between the three conditions. However, learning outcomes were relatively high, leaving little, if any, room for improvement. Therefore, in Experiment 2 more complex four-step geometry problems were used and given the complexity of the problems, the video examples included the model's verbal explanations. Students were either shown EMME or regular video examples (ME; i.e., without the model's eye movements superimposed). In addition, the student's eye movements were recorded while watching the video examples, to investigate how EMME affected attention. Learning outcomes were assessed with isomorphic problem-solving tasks and transfer problems. In line with the hypothesis, results showed that students in the EMME condition located the verbally referred information within the videos faster and fixated it longer compared to students in the ME condition. Thus,

EMME were able to guide the visual attention of the students. However, despite the fact that EMME were successful at guiding attention, they did not result in higher learning outcomes compared to the regular video modeling example condition.

One possible explanation for this finding might be that the verbal explanations were already clear and easy to understand, therefore reducing the need for attentional guidance. In situations in which the verbal explanations of the model are ambiguous for the learner (e.g., because the learner lacks prior knowledge to know what aspects of the visually displayed task the model's terminology is referring to, or because there are multiple parts of the display the model could be referring to), attentional guidance towards what the model is referring to might be of help, and this may have been the case in studies that found beneficial effects of EMME on learning (e.g., Jarodzka et al., 2012, 2013).

Therefore, a fundamental proof of concept study on effects of verbal ambiguity and visual complexity on visual search (i.e., selection of information) was conducted, reported in **Chapter 3**. In two experiments, this study examined whether visual search would be affected by verbal ambiguity of the descriptions (which referred to the target that had to be searched) and visual complexity of the images. A 2 (verbal ambiguity: high vs. low) x 2 (visual complexity: high vs. low) within-subjects design was used in which both the verbal ambiguity of the description and visual complexity of the search image was manipulated. It was expected that both high verbal ambiguity and high visual complexity would result in slower and less accurate visual search with a possible additive effect of both factors. In case verbal ambiguity would indeed negatively affect visual search, this would provide indirect evidence that in situations in which verbal explanations are ambiguous for the learners, EMME might be helpful to disambiguate the model's explanation. In line with the hypothesis results showed that higher verbal ambiguity led to a lower proportion of correctly identified targets and to slower response times. However, overall no reliable effect was found of visual complexity and also no interaction of verbal ambiguity and visual complexity on visual search performance. Thus, the results of this study indicated that verbal ambiguity affects visual search performance and thus EMME might be especially helpful when the model's explanation is ambiguous for learners.

The study reported in **Chapter 4** examined whether the ambiguity of the model's verbalizations would affect learning from EMME. Two experiments were conducted using the same geometry task as in Experiment 2 of Chapter 2, but with a more ambiguous verbal explanation. In Experiment 1, university students either watched regular video examples (i.e., without the model's eye movements visualized) or EMME. In both conditions the model's verbalizations contained ambiguous language due to the use of deictic terms and sentences like "now you know *this* angle, you can calculate the *other* angle". Students' eye movements were recorded while they watched the video examples, and learning outcomes were assessed by means of isomorphic and transfer problems. Again, results showed that EMME were successful at guiding students' visual attention as indicated by the fact that students in the EMME condition were fixating the verbally referred information faster and more often compared to students in the regular modeling example condition. Even though EMME guided the student's visual attention, this again did not result in higher learning outcomes for the EMME condition compared to regular modeling example condition. One explanation for this finding might be that students already performed relatively high on the pretest, which consequently left little room for improvement. Indeed research on cueing has shown that attentional guidance is most effective for learners with only limited prior knowledge (Richter et al., 2016). Therefore, participants in Experiment 2 were secondary education students. Experiment 2 had a 2 (modeling example: EMME vs. regular video example) x 2 (verbal explanation: ambiguous vs. unambiguous) between-subjects design. Because this study was conducted in regular classrooms, eye tracking could not be applied to measure students' attention, and only their learning outcomes were measured. The results revealed that EMME were effective for these secondary education students: regardless of the ambiguity of the model´s verbal explanations, students in the EMME conditions outperformed students in the regular video example conditions on the isomorphic problems and the transfer problems.

The study in Chapter 5 investigated the effect of the model's social status. Fundamental research on attention has shown that individuals with higher social status are looked at more by observers (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013) and that social status seems to affect how visualizations of eye movements are interpreted (Gobel, Tufft, & Richardson, 2018). Furthermore, social

status (or the model's competence) is also regarded as an important factor in research on learning from modeling examples (Braaksma, Rijlaarsdam, & van den Bergh, 2002; Sonnenschein & Whitehurst, 1980). Thus, the study reported in **Chapter 5** examined whether the purported expertise of the model would affect learning from EMME. Secondary education students watched EMME with unambiguous verbal explanations in which the model was explaining and demonstrating how to solve geometry problems (same as used in Chapters 2 and 4). The purported expertise of the model in the EMME was manipulated by means of a cover story preceding the EMME. In one condition, the model was introduced as a math teacher confident in her ability to explain the geometry problem (Expert condition) in the other condition as a Dutch teacher insecure about her ability to explain the geometry problem (No-Expert condition). Subsequently, students in both conditions watched the exact same EMME (i.e., only the purported expertise was different). Contrary to the hypothesis, the results revealed no differences in learning outcomes between the Expert vs. No-Expert conditions, suggesting that purported expertise of the model did not influence learning from EMME.

The study described in **Chapter 6** used a problem-solving task that capitalized more on the perceptual and cognitive strategy-conveying function of EMME and addressed the effects of EMME in the presence or absence of the model's (unambiguous) verbal explanations. An experiment was conducted that was a conceptual replication of the study by Van Gog et al. (2009) described above. In a 2 (modeling example: EMME vs regular modeling example) x 2 (verbal explanation: present vs. absent) between-subject design, secondary education students watched video modeling examples demonstrating how to solve deductive reasoning problems, and their learning outcomes were assessed. Results revealed no effect of EMME, no effect of verbal explanation, and no interaction between EMME and verbal explanation on learning to solve deductive reasoning problems.

Discussion of Main Findings

The studies presented in this dissertation demonstrated for the first time that EMME are successful at guiding attention of learners while studying examples on procedural problem-solving tasks (Chapter 2, Experiment 2; Chapter 4, Experiment 2) and, importantly, that EMME can also enhance learning to solve procedural problems, at least for adolescents who had low prior knowledge (Chapter 4,

Experiment 2). For university students with more prior knowledge, EMME successfully guided attention but did not improve learning (cf. Chapter 2 and Chapter 4, Exp.1). These effects on attention and learning did not seem to be influenced by verbal ambiguity of the model's explanation (Chapter 4) or by the model's purported expertise (Chapter 5). Moreover, EMME in which the model's eye movements served to model the problem-solving strategies that would have otherwise remained invisible, were not effective for learning (Chapter 6). Below, I will discuss the findings regarding each of the factors that were investigated in the studies in this dissertation in light of research on example-based learning and multimedia learning.

Verbal ambiguity. Because the studies that showed positive effects of EMME on learning (Jarodzka et al., 2012; 2013) might have been more ambiguous in the verbal explanations than the procedural problem-solving study in which no effects were found (Van Gog et al., 2009), the ambiguity of the model's verbal explanation looked like a potential boundary condition for the effectiveness of EMME for learning. Arguably, when it is clear to students what the model is looking at from the verbal explanation, guiding their attention will not be necessary for learning, whereas if it is unclear to them what the model is referring to, they may not look at the right information at the right time, which might hamper their learning (i.e., the integration of visual task information with the model's explanation).

Indeed, the initial proof of concept study showed that verbal ambiguity affected visual search performance (Chapter 3). In demonstrating that, this study made an interesting contribution to the more fundamental research literature about the interplay between language processing and visual search using complex visual stimuli. Moreover, it seemed to underline the idea that EMME might be most effective under high verbal ambiguity conditions. However, no effect of verbal ambiguity was found when used in EMME (Chapter 4, Experiment 2). One possible explanation might lie in how the verbal ambiguity was manipulated. For instance, in Chapter 4 the verbal explanation of the model were made more ambiguous by using deictic references (e.g., "now you know *this* angle, you can calculate the *other* angle"). However, in other studies of which it was reasoned that they may have shown positive effects of EMME because verbal explanations may have been ambiguous, the ambiguity would have arisen partly from technical or jargon language (Jarodzka et al., 2012; 2013), which may be ambiguous especially to students with low prior

knowledge. Such an explanation would be in line with the results of Experiment 2 in Chapter 4, in which EMME were found to be effective for learning when the student's prior knowledge was low.

Prior knowledge. Regarding prior knowledge the studies presented in this dissertation have shown that when the student's prior knowledge was low there was a positive effect of EMME on learning (Chapter 4 Experiment 2) but not when the prior knowledge was high (Chapter 2). These results suggests that prior knowledge might be a boundary condition as the effectiveness of EMME for learning appears to be moderated by the amount of prior knowledge of the students. Next to the fact that for students with low prior knowledge, the model's verbal explanation might be ambiguous, for instance due to technical jargon language, prior knowledge can also affect learning by influencing the amount of cognitive load a learner experiences (Mayer, 2014; Sweller, Ayres, & Kalyuga, 2011).

According to the *Cognitive Theory of Multimedia Learning* (Mayer, 2014) there are three processes involved in learning from multimedia materials: *selecting* relevant verbal and pictorial information, *organizing* these sources of information into coherent mental representations, and lastly *integrating* the mental representations with each other and to the prior knowledge. Because the information in dynamic visualizations like EMME is often transient, these learning processes have to be executed under time pressure, which places heavy demands on working memory (i.e., high cognitive load). For instance, if the right information is not selected at the right time, it may no longer be available for selection, or organization and integration (i.e., learning) will be hampered. Thus, by guiding attention, EMME (and other types of cueing; Van Gog, 2014) may be necessary for learning when students have low prior knowledge. Because the cognitive load imposed by the learning material is lower for learners with higher prior knowledge (Sweller et al., 2011), they may have sufficient cognitive capacity for processing the material even without attentional guidance. Support that visual guidance is especially effective for learners with low prior knowledge comes from a meta-analysis regarding the use of visual cues in animations which found that prior knowledge moderates the effects of cueing on learning (Richter, Scheiter, & Eitel, 2016). This would explain why a positive effect of EMME was found in Chapter 4 (Experiment 2) in which secondary education studies with limited prior knowledge participated,

whereas in other studies (Chapter 2 and Chapter 4 Experiment 1) university students with higher prior knowledge participated.

However, this explanation that EMME is mainly effective for learners with low prior knowledge seems to be at odds with the results of the study reported in Chapter 6, which found no positive effect of EMME even though secondary education students with low prior knowledge participated. This suggests that the relationship between prior knowledge and EMME is more complicated. It should be noted, though, that in Chapter 6 the main function of EMME was to depict the underlying cognitive strategy involved in a deductive reasoning task which would otherwise remain unobservable for the learner, whereas the main function of EMME in Chapters 2, 4, and 5, was to establish joint attention between the learner and the model. It is possible that when the function of EMME is aimed more towards depicting cognitive strategies, instead of merely establishing joint attention, that learners need to have a certain minimal amount of prior knowledge in order for them to correctly interpret and understand the strategy depicted in the EMME. The findings of a recent study of Scheiter, Schubert, and Schüler (2018) seem to support such a view. In this study, EMME were used to learn students how to read and integrate textual and pictorial information in multimedia learning materials. Compared to a control condition (i.e., without the EMME intervention) different effects of learning with EMME were found depending on the prior knowledge of the student. On the one hand, it was found that students with less prior knowledge had poorer recall performance in the EMME condition compared to the control condition. On the other hand, in a forced-choice verification task, students with higher prior knowledge showed better performance in the EMME condition than the control condition. The authors concluded that for learners to benefit from watching EMME it might require a minimal amount cognitive prerequisites (i.e., broad knowledge regarding scientific thinking enabling the student to understand instructions). This might explain why the study in Chapter 6 found no positive effect of EMME on learning how to perform a deductive reasoning task. Considering deductive reasoning is a complex skill it is possible that the secondary education students did not yet have the minimal required cognitive prerequisites to fully understand the demonstrated strategies in the video examples both in the EMME and ME conditions.

Model characteristics. The study reported in Chapter 5 explored whether model characteristics would affect learning from

EMME. This research question was inspired by research findings regarding regular modeling examples in which it was found that model characteristics like (perceived) competence (Braaksma et al., 2002; Sonnenschein & Whitehurst, 1980) or age (Hoogerheide, Van Wermeskerken, Loyens, & Van Gog, 2016), affected learning. For instance, Hoogerheide et al. (2016) found that secondary education students learned more from (female) adult models than peer models even though the actual content of the video examples was otherwise identical. A possible explanation for this result might be that the model's (perceived) social status affects the amount of attention that is devoted to the model's performance by the student. However, in those studies the model was visible for the learners (Braaksma et al., 2002; Hoogerheide et al., 2016). Whether these model characteristics would also affect learning from EMME, in which only the model's actions and eye movements are visible for learners, was an open question. Research from outside the field of education, suggests that this might be the case; for instance, it was found that the social rank of a person (who is visible) can affect visual attention of others to that person (Cheng et al., 2013), and that the purported status as high ranked person affected observers' attention (even when this person was not visible) to visualizations of that person's eye movements as a moving dot (Gobel et al., 2018). In contrast to these findings, the study presented in Chapter 5 revealed no difference in learning outcomes between the Expert and the No-Expert EMME conditions. This was surprising given that the representation of the gaze location of the socially high ranked person in the study by Gobel et al. (2018) was very similar to how gaze is displayed in EMME; however, the learning task used in Chapter 5 was very different from the task in Gobel et al. and it cannot be ruled out that learners' attention was affected even if their learning was not, as learners' eye movements were not recorded in this classroom study.

On the positive side, these findings may indicate that as long as the content of the modeling examples itself is of high quality the *purported* expertise does not affect learning from EMME. It is likely, however, that the results would have been different if the models differed in *actual* expertise and were behaving naturally rather than didactically. In most EMME studies, the model behaves in a didactical manner (i.e., demonstrating the task, making deliberate eye movements). It is an open question whether novices would even be able to interpret EMME if the model would behave naturally and

perform the task as s/he always does. Novices might in that case have trouble interpreting expert models' eye movements as eye-tracking research on expertise has shown that experts compared to novices tend to fixate relevant information faster and longer, and are better able to ignore task irrelevant information (Charness, Reingold, Pomplun, & Stampe, 2001; Haider & Frensch, 1999). On the one hand, this behaviour might differ so much from that of a novice, that it be more difficult for a novice to interpret than, for instance, the behaviour of a less-expert model (who is only somewhat more advanced than the student). On the other hand, the results of the study by Litchfield et al. (2010) suggest that novices can actually improve current task performance in a classification task when observing 'natural' EMME as long as the eye movements are task related. Whether such results would also be obtained for different types of tasks in which the aim is to improve learning and transfer (rather than immediate performance) remains to be seen.

Implications for Practice

Developing EMME requires the use of eye-tracking technology. At present, eye-tracking technology is not readily available at schools. However, this should not pose a problem for the implementation of EMME considering that EMME can be recorded elsewhere. Once a teacher has recorded the EMME they can be placed in online learning environments or video-streaming channels, where learners can be view them whenever and wherever they want, including at home when they are working on homework assignments, for example. In this way, even in the absence of the teacher, s/he can guide students' attention towards the relevant information in an instructional video. Moreover, due to technological advancement eye-tracking technology is rapidly becoming more mobile and affordable. This trend makes it possible for eye-tracking technology to be within reach for schools in the near future. The fact that eye-tracking technology is becoming more accessible is already reflected by its increased use on video game streaming websites like TwitchTV (www.twitch.tv). On such streaming platforms video gamers are screen-recording their video game with their eye movements being recorded and visualized. In essence, these video game streams are live EMME.

For teachers and educational designers who wish to develop and implement EMME, the results of the studies presented in this

dissertation offer some practical implications. The results of Chapter 4 (Experiment 2) indicate that EMME can be used for learning procedural problem-solving strategies in STEM-related subjects like science or math. For instance, a math teacher could demonstrate in short videos how to solve math problems and record his/her eye movements while explaining. In addition, our results (Chapter 4 Experiment 2) indicate that the ambiguity of the verbal explanations in terms of deictic references did not differentially affect learning from EMME or regular video examples, so it is probably safe for teachers to use a similar approach to explaining as they would do in a live situation (where people often rely on gestures to disambiguate speech; McNeill, 1985). Finally, the results of Chapter 5 suggest that (as long as the quality of the EMME is high), the purported expertise of the model does not seem to matter, so when using a good script, a diversity of models could be used, although future research should address whether other model characteristics would play a role in learning from EMME.

Future Research

One direction for future research is related to the type of control conditions that were used in the studies in this dissertation and in other studies investigating EMME (e.g., Jarodzka et al., 2012; 2013; Mason et al., 2015; 2016; Van Gog et al., 2009). The effectiveness of EMME and how EMME affects learning has almost exclusively been compared to a control condition in which the model's eye movements are not shown (i.e., using a regular modeling example). One could argue that a better control condition would be to use other visual cues (i.e., arrows or colored highlights). By doing so, the difference in visual information between conditions is controlled for. By comparing EMME to video examples with other visual cues, it can be investigated whether the effectiveness of EMME is purely driven by the visual guidance component or that the visualized eye movements have additional value beyond mere cueing (e.g., visualizing a cognitive strategy involved in a task; raising the feeling of a 'social' learning situation, cf. Gobel et al., 2018, who found that a red dot affects performance differently when people feel it indicates another person's eye gaze than just a red dot).

Whether EMME can act as more than a visual cue might also depend on the task being demonstrated and the function of EMME. For instance, in studies in which EMME were mainly used to establish

joint attention and facilitate the integration of visual and verbal information, one could probably replace the model's visualized eye movements with well-timed arrows or visual highlighting. However, when the main role of EMME is to visualize and to model the underlying processes regarding perceptual strategies (cf. visual search tasks; Litchfield et al., 2010; Nalanagula et al., 2006; Stein & Brennan, 2004) or cognitive strategies (cf. deductive reasoning tasks; Chapter 6), eye movements might be more effective cues than static cues like arrows or highlighting. Although one could use other dynamic, continuous visual cues (e.g., Boucheix & Lowe, 2010), such other forms of cueing will always have the drawback compared to EMME that they require post-processing. That is, the expert, teacher, or instructional designer has to indicate / decide where cues have to be placed. And due to task automatization (Chi, 2006), experts might not even be consciously aware of what they were looking at while performing the task, which may make it difficult to determine how other visual cues should be implemented. Thus, in future research it would be interesting to investigate whether and under which conditions (see e.g., the prior knowledge aspect discussed above) EMME can be applied to model perceptual and cognitive strategies, and are more effective at that than other forms of cueing.

Another interesting research avenue is related to how the model behaved in the EMME. In most EMME studies (e.g., Jarodzka et al., 2012; 2013; Mason et al., 2015; 2016; 2017) and also in the studies in this dissertation, the model in EMME behaved in a didactical manner by making very deliberate eye movements towards the different elements in the task demonstration. The question is whether the positive effects of EMME on learning would also be present if the model behaved naturally during the task demonstration. An advantage of recording the model's natural behavior is that it would be less demanding for the teachers and could expand the potential applications of EMME with real-time live streaming (cf. the gaming example given above). Think, for instance, of real-time live streaming of programmers' eye movements during debugging, or surgeons' eye movements during [simulated] operations, to learners who do not even need to be present in the same room. With EMME, learners can literally learn to look at a task through the teacher's eyes.

Chapter 7

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://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://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://doi.org/10.1002/acp.1351

Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animation. *Learning and Instruction*, *20*, 123–135. http://dx.doi.org/10.1016/j.learninstruc.2009.02.015.

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://doi.org/10.1037/0022-0663.94.2.405

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., Zelinsky, G J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477. doi:10.1016/j.cognition.2007.05.012

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*, 1283-1297. doi:10.3758/APP.72.5.1283

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://doi.org/10.3758/BF03206384

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, *104*, 103–125. http://dx.doi.org/10.1037/a0030398.

Chi, M. T. H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167–184). Cambridge, UK: Cambridge University Press.

Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., & Shelepin, Y. (2012). Complexity of images: Experimental and computational estimates compared. *Perception*, *41*, 631–647. http://dx.doi.org/10.1068/p6987.

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

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

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://doi.org/10.3758/BF03193787
Davis, E. T., Shikano, T., Peterson, S. A., & Michel, R. K. (2003). Divided attention and visual search for simple versus complex features. *Vision Research*, *43*, 2213-2232. doi:10.1016/S0042-6989(03)00339-0
De Koning, B. B., & Jarodzka, H. (2017). Attention guidance strategies for supporting learning from dynamic visualizations. In R. Lowe & R. Ploetzner (Eds.) *Learning from Dynamic Visualization* (pp. 255–278). Springer International Publishing.
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.
De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, *27*, 41–45. doi:10.1016/j.chb.2010.05.010
Donderi, D.C. (2006a). An information theory analysis of visual complexity and dissimilarity. *Perception*, *35*, 823–835. http://dx.doi.org/10.1068/p5249.
Donderi, D.C. (2006b). Visual complexity: A review. *Psychological Bulletin*, *132*, 73–97. doi: 10.1037/0033-2909.132.1.73
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://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*. Advance online publication. http://doi.org/10.1037/edu0000065.
Forsythe, A. (2009). Visual complexity: Is that all there is? *Engineering Psychology and Cognitive Ergonomics*, *LNCS*, *5639*, 158–166.
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*, 694–724. dx.doi.org/10.1037/0033- 2909.133.4.694
Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers and Education*, *113*, 212–225. http://dx.doi.org/10.1016/j.compedu.2017.06.001
Gerlach, C., & Marques, J. F. (2014). Visual complexity exerts opposing effects on object categorization and identification. *Visual Cognition*, *22*, 751–769. http://doi.org/10.1080/13506285.2014.915908.
Gierasimczuk, N., van der Maas, H. L. J., & Raijmakers, M. E. J. (2013). An analytic tableaux model for deductive mastermind empirically tested with a massively used online learning system. *Journal of Logic, Language and Information*, *22*, 297–314. DOI: 10.1007/s10849-013-9177-5
Gobel, M. S., Tufft, M. R. A., & Richardson, D. C. (2018). Social beliefs and visual attention: How the social relevance of a cue influences spatial orienting. *Cognitive Science*, *42*, 161–185. DOI: 10.1111/cogs.12529
Godwin, H. J., Walenchok, S. C., Houpt, J. W., Hout, M. C., & Goldinger, S. D. (2015). Faster than the speed of rejection: Object identification processes during visual search for multiple targets. *Journal of Experimental Psychology: Human Perception & Performance*, *41*, 1007-1020. doi: 10.1037/xhp0000036.
Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*, 462–466.

http://doi.org/10.1111/1467-9280.02454
Groen, M., & Noyes, J. (2010). Solving problems: how can guidance concerning task-relevancy be provided? *Computers in Human Behavior*, *26*, 1318–1326. http://doi.org/10.1016/j.chb.2010.04.004
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://doi.org/10.1016/j.learninstruc.2013.05.001
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.
Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 37–53. DOI: 10.1037/a0017683
Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search:Evidence from search efficiency and eye movements. *Journal of Vision*, *9*, 1–8. http://journalofvision.org/9/1/32/,doi:10.1167/9.1.32.
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://doi.org/10.1016/j.chb.2014.09.013
Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, *44*, 22–30. https://doi.org/10.1016/j.learninstruc.2016.02.004
Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*, 128-149. Doi: 10.3758/s13414-014-0764-6.
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://doi.org/10.1016/j.actpsy.2010.11.003
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://doi.org/10.1016/j.actpsy.2010.11.003
Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, *32*, 47–53. http://dx.doi.org/10.1016/j.chb.2013.11.013
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. http://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://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://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 edn; pp.

247–262). New York: Cambridge University Press.
Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2015). The sound of power: conveying and detecting hierarchical rank through voice. *Psychological Science*, *26*, 3–14. https://doi.org/10.1177/0956797614553009
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., 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://doi.org/10.1037/a0020082
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://doi.org/10.1080/17470218.2011.558628
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://doi.org/10.1111/j.1551-6709.2010.01135.x
Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, *14*, 225–244. doi:10.1007/BF03172967
Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*, 1-13. doi: 10.1167/9.11.8
Marin, M. M., & Leder, H. (2016). Effects of presentation duration on measures of complexity in affective environmental scenes and representational paintings. *Acta Psychologica*, *163*, 38–58. http://dx.doi.org/10.1016/j.actpsy.2015.10.002
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. http://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. doi.org/10.1111/bjet.12271
Mason, L., Scheiter, K.,&Tornatora, C. (2017). Using eye movements to model the sequence of textpicture 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: Cambridge University Press.
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
McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*, 350-371.
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://doi.org/10.1016/j.ergon.2005.11.008
Neider, M. B., & Zelinksky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, *11*, 1-16. doi:10.1167/11.14.7
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://doi.org/10.1002/acp.3097
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1–37. http://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://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://doi.org/10.1111/j.1467-9280.2007.01914.x
Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. http://dx.doi.org/10.1016/j.edurev.2015.12.003
Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*. Advance online publication. https://doi.org/10.1177/0735633117751605
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 Technology, 88*, 80–94. https://doi.org/10.1111/bjep.12175.
Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, *62*, 1904-1914. doi: 10.1080/17470210902853530
Schommer, M. (1998). The influence of age and schooling on epistemological beliefs. *The British Journal of Educational Psychology*, *68*, 551–562. https://doi.org/10.1111/j.2044-8279.1998.tb01311.x
Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*, 149–174.
Sonnenschein, S., & Whitehurst, G. J. (1980). The development of communication: when a bad model makes a good teacher. *Journal of Experimental Child Psychology*, *3*, 371–390. http://dx.doi.org/10.1016/0022-0965(80)90101-0.
Sridharan, S., Bailey, R., McNamara, A., & Grimm, C. (2012). *Proceedings of the 2012 Symposium on eye tracking research & Applications ETRA '12*, 75–82. Santa Barbara, CA. ACM. http://dx.doi.org/10.1145/1743666.1743737.
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://doi.acm.org/10.1145/1027933.1027936
Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. New York: Springer.
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://doi.org/10.3758/BF03196818
Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). New York, NY: Cambridge University Press.

http://dx.doi.org/10.1017/CBO9781139547369
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://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://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://doi.org/10.1007/s10648-010-9134-7
Van Gog, T., Rummel, N., & Renkl, A. (in press). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *Cambridge handbook and cognition and education*. New York, NY: Cambridge University Press.
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://doi.org/10.1016/j.compedu.2013.12.004
Van Marlen, T., Van Wermeskerken, M., Boven, L., Ketelaar, J., van Trijp, K., Vernooij, S., & Van Gog, T. (2016). Are eye movement modeling examples most useful when verbal instructions are ambiguous? *Paper presented at the EARLI Special Interest Group 6 and 7 Dijon, France.*
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 Marlen, T., Van Wermeskerken., & Van Gog, T. (in press). Effects of visual complexity and ambiguity of verbal instructions on target identification. *Journal of Cognitve Psychology*.
Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., De Bock, J. J. R. P., & Van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, *32*, 10–21. http://dx.doi.org/10.1016/j.learninstruc.2014.01.004
Vitak, S. A., Ingram, J.E., Duchowski, A.T., Ellis, S., & Gramopadhye, A.K. (2012). Gaze-augmented think-aloud as an aid to learning. *In Proceedings of ACM CHI 2012 Conference on Human Factors in Computing Systems*, 2991–3000, Austin, Texas, 2012. ACM Press.
Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research, 34,* 1187–1195.
Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*, 2985-3004. doi: 10.1016/S0042-6989(02)00388-7
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://doi.org/10.1007/s11251-016-9367-z

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*, 2095-2103. doi: 10.1016/j.visres.2009.05.017

Yoon, S. H., Lim, J. H., & Ji, Y. G. (2015). Perceived visual complexity and visual search performance of automotive instrument cluster: A quantitative measurement study. *Journal of Human–Computer Interaction*, *31*, 890–900. http://dx.doi.org/10.1080/10447318.2015.1069661.

Nederlandse samenvatting

Summary in Dutch

Eye movement modeling examples (EMME) zijn videovoorbeelden waarin lerenden niet alleen via een schermopname zien hoe een model (bijvoorbeeld een leraar) voordoet hoe je een taak uitvoert, maar ook zien waar hij of zij naar keek tijdens het uitvoeren van de taak. Dit wordt gedaan door de oogbewegingen van het model te visualiseren door middel van (bijvoorbeeld) een gekleurde stip of cirkel (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). Vaak bevatten EMME ook een verbale uitleg van het model over wat hij/zij aan het doen is, hoewel dit niet altijd het geval is (zo werd er geen uitleg gegeven in de EMME in de studies van Mason, Pluchino, & Tornatora, 2015; 2016; Mason, Scheiter, & Tornatora, 2017; Salmerón & Llorens, 2018; Scheiter, Schubert, & Schüler, 2018).

Eerder onderzoek heeft aangetoond dat EMME de prestaties op visuele zoektaken kunnen verbeteren (Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Nalanagula, Greenstein, & Gramopadhye, 2006; Stein & Brennan, 2004). Bovendien is er evidentie dat EMME ook een positief effect kunnen hebben op het *leren* (d.w.z. op latere prestaties op soortgelijke taken zonder de ondersteuning middels EMME). Zo verbeterden EMME (in vergelijking met videovoorbeelden zonder de oogbewegingen van het model), het leren van strategieën voor het integreren van teksten en bijbehorende illustraties (Mason et al., 2015; 2016; 2017; Salmerón & Llorens, 2018; Scheiter et al., 2018) en leren van classificatietaken (Gegenfurtner, Lehtinen, Jarodza, & Säljö, 2017; Jarodzka et al., 2012; 2013; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012). Echter, een studie waarin EMME werden gebruikt om te laten zien hoe een procedureel puzzelprobleem opgelost kon worden, vond een negatief effect van EMME op leren wanneer deze werden gecombineerd met de verbale uitleg van het model (Van Gog et al., 2009).

Gezien het feit dat EMME effectief zijn gebleken voor verschillende soorten taken en dat het bewijs over de ineffectiviteit van EMME voor het leren van procedurele probleemoplostaken zich beperkte tot één enkele studie, zou het voorbarig zijn om te concluderen dat EMME niet geschikt zijn om het leren van procedurele probleemoplostaken te verbeteren. Omdat zulke taken een belangrijke component zijn van veel vakken op alle onderwijsniveaus (bijv. Rekenen/Wiskunde, Natuurkunde, Scheikunde, Economie, Statistiek), is het van belang om te

onderzoeken of –en onder welke voorwaarden- EMME ook gebruikt kunnen worden om het leren van procedurele probleemoplostaken te bevorderen. De vraag die in dit proefschrift centraal staat, is dan ook of EMME ook bij procedurele probleemoplostaken de aandacht van de lerende sturen en het leren verbeteren. Vijf empirische studies, met in totaal acht experimenten, werden uitgevoerd om de hoofdonderzoeksvraag te beantwoorden: *Verbeteren EMME het leren van procedurele probleemoplostaken?* Deze studies zetten tevens een eerste stap richting het in kaart brengen van verschillende mogelijke randvoorwaarden die van invloed kunnen zijn op het leren van procedurele probleemoplostaken middels EMME. Dit is gedaan door gebruik te maken van taken van verschillende complexiteitsniveaus, taken waarbij de visualisaties van de oogbewegingen van het model een verschillende functie hebben, taken waarbij de verbale uitleg van het model duidelijk versus ambigu was en door het onderzoek uit te voeren onder lerenden op verschillende onderwijsniveaus die verschilden in voorkennis.

Samenvatting van de belangrijkste bevindingen

In het onderzoek gerapporteerd in **Hoofdstuk 2** werd onderzocht of EMME waarin getoond werd hoe procedurele geometrie problemen opgelost moesten worden, invloed hadden op de aandacht en de leerresultaten. In Experiment 1 keken studenten naar videovoorbeelden waarin een model demonstreerde hoe geometrieproblemen bestaande uit één stap konden worden opgelost. Hierbij werden in één groep niet en in een andere groep wel de oogbewegingen van het model getoond in de videovoorbeelden (middels een lichtblauwe stip), en in een derde groep werd er een willekeurig patroon van oogbewegingen getoond. Deze laatste conditie was toegevoegd om te onderzoeken of we conceptueel de bevindingen van Litchfield et al. (2010) konden repliceren. Litchfield et al. (2010) toonden aan dat alleen taakrelevante oogbewegingen de prestaties op een zoektaak verbeterden. De conditie met willekeurige patronen was toegevoegd om uit te sluiten dat een eventueel positief effect van EMME louter optrad wegens de aanwezigheid van de visualisaties van oogbewegingen, die ertoe zouden kunnen leiden dat studenten beter opletten. De resultaten lieten geen verschillen zien in leerresultaten tussen de drie condities. De leerresultaten waren echter relatief hoog, waardoor er weinig of geen ruimte voor verbetering was.

Om die reden werden in Experiment 2 complexere geometrieproblemen bestaande uit vier stappen gebruikt en vanwege

de hogere complexiteit van de problemen bevatten de videovoorbeelden ook de verbale uitleg van het model. Studenten kregen EMME of reguliere videovoorbeelden te zien (zgn. *Modeling Examples*, *ME*; d.w.z. zonder de oogbewegingen van het model). Daarnaast werden de oogbewegingen van de studenten opgenomen tijdens het bekijken van de videovoorbeelden om te onderzoeken hoe EMME de aandacht beïnvloedden. Leerresultaten werden vastgesteld aan de hand van de prestatie op isomorfe problemen (d.w.z. problemen die op precies dezelfde manier opgelost konden worden als de problemen die uitgelegd werden in de videovoorbeelden) en transfer problemen (d.w.z. problemen die net even anders waren dan de problemen die uitgelegd werden in de videovoorbeelden). In overeenstemming met de hypothese bleek uit de resultaten dat studenten in de EMME conditie sneller en langer keken naar de informatie waaraan in de uitleg gerefereerd werd, in vergelijking met studenten in de ME conditie. Dus EMME waren succesvol in het sturen van de aandacht van de studenten. Echter resulteerde dit niet in hogere leerresultaten in vergelijking met de ME conditie.

Een mogelijke verklaring voor deze bevinding kan zijn dat de verbale uitleg al duidelijk en gemakkelijk te begrijpen was, waardoor het sturen van aandacht minder noodzakelijk was voor het leren. In situaties waarin de verbale uitleg van het model ambigu is voor de lerende (bijv. omdat de lerende geen voorkennis heeft en dus niet weet naar welk aspect van de taak een gebruikte term verwijst, of omdat een term naar meerdere onderdelen van de taak zou kunnen verwijzen), kan het sturen van de aandacht naar waar het model naar verwijst cruciaal zijn voor het leren. Dit was waarschijnlijk ook het geval in studies die gunstige effecten van EMME op leren hebben gevonden (bijv. Jarodzka et al., 2012; 2013).

Derhalve werd eerst een fundamentele proof-of-concept studie uitgevoerd naar de effecten van verbale ambiguïteit en visuele complexiteit op een visuele zoektaak (d.w.z. selectie van informatie), gerapporteerd in **Hoofdstuk 3**. In twee experimenten werd onderzocht of de prestatie op een visuele zoektaak zou worden beïnvloed door verbale ambiguïteit van de beschrijving (die verwees naar het item dat moest worden gezocht) en visuele complexiteit van de afbeeldingen. De studie had een 2 (verbale ambiguïteit: hoog versus laag) x 2 (visuele complexiteit: hoog versus laag) within-subjects design waarin zowel de verbale ambiguïteit van de beschrijving als de visuele complexiteit van de afbeelding werden gemanipuleerd. Er werd

verwacht dat zowel hoge verbale ambiguïteit als hoge visuele complexiteit zouden resulteren in een langzamere en minder nauwkeurige prestatie op de zoektaak, met een mogelijk additief effect van beide factoren. Indien verbale ambiguïteit inderdaad de prestatie op de zoektaak negatief zou beïnvloeden, zou dit indirect bewijs opleveren dat in situaties waarin verbale uitleg ambigu is voor de lerende, EMME nuttig kunnen zijn omdat het zien waar het model naar kijkt, de ambiguïteit van de uitleg kan opheffen. In overeenstemming met de hypothese, lieten de resultaten zien dat hogere verbale ambiguïteit leidde tot een lagere prestatie op de zoektaak (d.w.z. een lagere proportie correct geïdentificeerde targets) en tot langzamere reactietijden. Over het algemeen vonden we echter geen betrouwbaar effect van visuele complexiteit en ook geen interactie van verbale ambiguïteit en visuele complexiteit op de prestatie op de zoektaak. De resultaten van dit onderzoek gaven dus aan dat verbale ambiguïteit van de taakbeschrijving van invloed is op de prestaties op een zoektaak, wat suggereert dat EMME mogelijk vooral nuttig zijn wanneer de uitleg van het model ambigu is voor de lerende.

Het onderzoek beschreven in **Hoofdstuk 4** onderzocht of de ambiguïteit van de verbalisaties van het model van invloed zou zijn op het leren van EMME. Twee experimenten werden uitgevoerd met dezelfde geometrietaken als in Experiment 2 van Hoofdstuk 2, maar met een meer ambigue verbale uitleg. In Experiment 1 keken studenten naar ME of naar EMME. In beide condities was de uitleg van het model ambigu wegens het gebruik van deiktische termen en zinnen zoals "nu je *deze* hoek weet, kun je de *andere* hoek berekenen" (in plaats van het niet-ambigue “nu je hoek A weet…”). Net als in hoofdstuk 2, werden de oogbewegingen van studenten opgenomen terwijl ze de videovoorbeelden bekeken en werden de leerresultaten gemeten middels isomorfe en transfer problemen. Wederom lieten de resultaten zien dat EMME succesvol waren in het sturen van de aandacht van studenten, wat bleek uit het feit dat studenten in de EMME conditie sneller en vaker keken naar de informatie waaraan in de uitleg gerefereerd werd, dan studenten in de ME conditie. Echter, resulteerde dit opnieuw niet in hogere leerresultaten in de EMME conditie in vergelijking met de ME conditie. Een verklaring voor deze bevinding kon zijn dat studenten al relatief hoog presteerden op de voorkennistest, waardoor er weinig ruimte was voor verbetering. Eerder onderzoek naar cueing had ook

aangetoond dat aandachtsturing het meest effectief is voor lerenden met beperkte voorkennis (Richter et al., 2016). Om die reden werden in Experiment 2 middelbare scholieren getest, die relatief weinig voorkennis hadden. Experiment 2 had een 2 (voorbeeld: EMME versus ME) x 2 (verbale uitleg: ambigu versus duidelijk) between-subjects design. Omdat deze studie werd uitgevoerd in reguliere klaslokalen kon eye tracking niet worden toegepast om de aandacht van de scholieren te meten en werden alleen de leerresultaten gemeten. De resultaten lieten zien dat EMME effectief waren voor deze middelbare scholieren: ongeacht de ambiguïteit van de verbale uitleg van het model presteerden studenten in de EMME condities beter dan studenten in de ME condities op de isomorfe problemen en de transfertaken.

De studie in Hoofdstuk 5 onderzocht het effect van de sociale status van het model. Fundamenteel onderzoek naar aandacht heeft laten zien dat mensen meer kijken naar mensen met hogere sociale status (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013) en dat sociale status ook de interpretatie van visualisaties van oogbewegingen lijkt te beïnvloeden (Gobel, Tufft, & Richardson, 2018). Bovendien wordt sociale status (of de competentie van het model) eveneens als belangrijke factor gezien in onderzoek naar leren van voorbeelden (Braaksma, Rijlaarsdam, & van den Bergh, 2002; Sonnenschein & Whitehurst, 1980). Daarom onderzocht de studie gerapporteerd in **Hoofdstuk 5** of het leren van EMME beïnvloed werd door de (zogenaamde) sociale status van het model. Middelbare scholieren keken naar EMME over geometrie problemen (vgl. Hoofdstuk 2 en 4) met duidelijke verbale uitleg. De (zogenaamde) expertise van het model werd gemanipuleerd door middel van een korte introductie voorafgaand aan de EMME. In één conditie werd het model geïntroduceerd als een wiskundelerares die zelfverzekerd is over haar vermogen om het geometrieprobleem uit te leggen (Expert conditie), in de andere conditie als een lerares Nederlands die onzeker is over haar vermogen om het geometrieprobleem uit te leggen (No-Expert conditie). Vervolgens keken scholieren in beide condities naar precies dezelfde EMME (d.w.z. de expertise verschilde niet daadwerkelijk). In tegenstelling tot de hypothese, lieten de resultaten geen verschillen in leeruitkomsten zien tussen scholieren in de Expert versus No-Expert condities, wat suggereert dat de zogenaamde expertise van het model het leren van EMME niet beïnvloedt.

Naast het sturen van aandacht, maken EMME ook perceptuele en cognitieve strategieën van het model zichtbaar die normaal gesproken onzichtbaar zijn. Het onderzoek beschreven in **Hoofdstuk 6** gebruikte een andere probleemoplostaak die sterker leunde op deze strategie-overdragende functie van EMME (Mastermind problemen) en onderzocht de effecten van EMME in de aanwezigheid of afwezigheid van de (duidelijke) verbale uitleg van het model (dit experiment was een conceptuele replicatie van het onderzoek van Van Gog et al., 2009). In een 2 (voorbeeld: EMME versus ME) x 2 (verbale uitleg: aanwezig versus afwezig) between-subjects design keken middelbare scholieren naar videovoorbeelden waarin werd gedemonstreerd hoe ze de deductieve Mastermind problemen konden oplossen. Na afloop werden hun leerresultaten bepaald door het meten van hun prestatie op isomorfe en transfertaken. De resultaten lieten echter geen effect zien van EMME, geen effect van verbale uitleg, en geen interactie tussen EMME en verbale uitleg op het leren oplossen van deductieve redeneerproblemen. Een mogelijke verklaring hiervoor was dat de voorkennis van de scholieren wellicht te laag was om de getoonde strategieën goed te kunnen begrijpen.

Concluderend, zijn de belangrijkste resultaten van dit proefschrift dat het tonen van oogbewegingen van het model in videovoorbeelden (EMME) succesvol de aandacht van lerenden kan sturen en ook het leren van procedurele probleemoplostaken kan bevorderen vergeleken met reguliere videovoorbeelden –mits lerenden weinig of geen voorkennis hebben. De ambiguïteit van de uitleg lijkt geen invloed te hebben op het leren van EMME. Echter, werd er in dit proefschrift slechts één specifieke manipulatie van ambiguïteit (deiktische referenties) onderzocht, dus toekomstig onderzoek zou moeten uitwijzen of het bij andere vormen van ambiguïteit van de uitleg (bijvoorbeeld door het gebruik van jargon) wel zo is dat EMME nog beter zijn voor het leren dan reguliere videovoorbeelden. Het feit dat het leren van EMME niet werd beïnvloed door de sociale status (zogenaamde expertise) van het model, is relevant voor het gebruik van EMME in online leeromgevingen. Dit suggereert namelijk dat –zolang de inhoudelijke kwaliteit van EMME hoog is- het niet uitmaakt voor de leeruitkomsten of de lerende denkt dat het model expert is of niet. De resultaten van het laatste empirische hoofdstuk laten zien dat het goed zou zijn om

nog meer onderzoek te doen naar de strategie-overdragende functie van EMME en de mogelijke rol van voorkennis hierin.

Op dit moment is eye-tracking technologie nog niet beschikbaar op scholen, maar gezien het feit dat eye-tracking technologie steeds goedkoper en toegankelijker wordt, zal dit in de nabije toekomst mogelijk wel het geval zijn. Bovendien kunnen EMME al wel ingezet worden in het onderwijs. Want eenmaal opgenomen kunnen EMME worden geplaatst in online leeromgevingen of op videosites zoals YouTube, zodat lerenden de EMME zo vaak ze willen en wanneer ze maar willen kunnen bekijken. Met EMME kunnen lerenden dus zelfs in de afwezigheid van een docent, letterlijk leren om door de ogen van een docent naar een taak te kijken.

Dankwoord

Acknowledgements in Dutch

Na bijna 5 jaar is mijn proefschrift eindelijk af. Hierbij hoort natuurlijk ook een dankwoord. In de loop der jaren zijn er een hoop mensen geweest in mijn leven die op directe dan wel indirecte wijze een bijdrage hebben geleverd aan het tot stand komen van dit proefschrift. Ik ga proberen iedereen te noemen die volgens de hierboven genoemde beschrijving in aanmerking komt voor dit dankwoord. Sommige mensen zullen vaker worden genoemd, omdat deze in verschillende contexten een bijdrage hebben geleverd.

Als eerste wil ik graag Tamara bedanken. Eind 2013 heb je me de kans gegeven om op het Vidi project aan de slag te gaan. Deze beslissing heeft voor mij veel betekend. Tamara, ik wil je niet alleen bedanken voor de feedback waar ik veel van heb geleerd en die hebben bijgedragen aan dit proefschrift, ook wil ik je bedanken voor de gezelligheid. Ik ken geen andere vrouw die zo ontzettend druk is, maar tegelijkertijd bijna altijd tijd vrijmaakt om aanwezig te zijn bij sociale gelegenheden. Je persoonlijke betrokkenheid ondanks je drukke agenda vind ik erg knap.

Margot, als dagelijkse begeleider was jouw rol in het tot stand komen van dit proefschrift cruciaal. Jouw laagdrempelige bereikbaarheid en steun hebben me vaak goed geholpen. Niet alleen heb je me vaak goed geholpen met de meer praktische lab achtige zaken maar ook was je altijd supersnel met je feedback en kon ik weer een stapje vooruit. In alle eerlijkheid moest ik in het begin van onze samenwerking wel een beetje wennen aan je. Zo had ik het ineens regelmatig over (kleine) kinderen! Dit had echter later wel als voordeel dat ik al wat tips en tricks had gehoord voordat ik zelf begon met het hebben van een kind. Ongetwijfeld heb je ook aan mij, een game nerd, moeten wennen. Gelukkig waren we snel aan elkaar gewend en groeide onze band mede dankzij de vaak leerzame en leuke congressen. Vooral ons 1^{e} congres samen in Freiburg heeft echt het ijs doen breken. Zo zal ik nooit vergeten wat een hoop lol we die avond hebben gehad bij het conference dinner. Verder heb ik altijd fijn met je kunnen samenwerken niet alleen binnen ons project maar ook bij het verzorgen van het onderwijs op de afdeling. Bedankt hiervoor! Ik zal deze fijne samenwerking met jou gaan missen nu je recent bent vertrokken uit Utrecht. Ik wens je nogmaals heel veel succes in je nieuwe functie!

Halszka, je was aanvankelijk bij het project betrokken als external advisor maar later werd ook jij net als Margot co-promotor. Ik wil je graag bedanken voor zowel je feedback op mijn stukken als op je enthousiasme die je altijd bij ons overleg had. In de loop der jaren heb ik gemerkt dat deze enthousiaste houding niet alleen aanwezig was bij onze project overleggen maar ook bij andere gelegenheden. Zo ook bij een borrel bij het EARLI SIG 27 conferentie in Finland. Je zat die avond rechts van mij en sindsdien heb ik nog steeds last van piepjes in mijn rechteroor. Ook daar wil ik je graag voor bedanken!

Mijn interesse in het uitvoeren van wetenschappelijk onderzoek was al ruim voor mijn promotietraject aangewakkerd tijdens mijn bachelor- en masterscriptie. René en Diane bedankt voor het betrekken van mij als student in jullie labmeetings en daarbuiten. Ik heb via jullie niet alleen vooraf een kijkje gekregen in de wereld van de wetenschap maar heb ook veel van jullie geleerd. Dit geldt natuurlijk ook voor alle andere Memorylab members!

Geen enkel project kan worden uitgevoerd zonder soms wat (technische) ondersteuning en de studies in mijn proefschrift zijn daar dan ook zeker geen uitzondering op. Christiaan, Marcel en Gerrit-Jan, bedankt voor alle technische hulp in het lab in Rotterdam. Jos, bedankt voor alle hulp bij het opzetten van het eye-track lab in Utrecht. Charly, bedankt voor het programmeren van mijn Flash experiment en de gezellige game avonden. Jacqueline bedankt voor je voice acting werk voor mijn Guess Who studie. Daarnaast kan geen enkel experiment worden uitgevoerd zonder proefpersonen. Dus bij deze wil ik graag alle proefpersonen bedanken voor hun deelname aan mijn experimenten. In het bijzonder wil ik alle stafleden, docenten en scholieren van het Emmaus college en het Krimpenerwaard college bedanken voor hun deelname aan mijn onderzoeken.

Daarnaast heb ik tijdens het uitvoeren van mijn studies het geluk gehad dat ik Susan als assistent had. Susan, zonder jouw hulp zouden de studies uitgevoerd op de middelbare scholen veel chaotischer zijn verlopen. Vooral die keer dat ineens de computers in het computerlokaal niet werkte heeft jouw rust ervoor gezorgd dat uiteindelijke niet alle metingen verloren zijn gegaan.

Doordat ik tijdens mijn promotietraject van de EUR naar UU ben gegaan, heb ik het plezier gehad om op verschillende afdelingen

te werken. In Rotterdam wil ik graag iedereen van het HLP team bedanken. Daarnaast wil ik alle andere collega's bedanken voor de interessante en gezellige gespreken tijdens de aio-etentjes, bowlingavonden en andere gelegenheden. Ook wil ik graag de volgende mensen van de pubgroup bedanken: Vincent, Steven, Gertjan, Charly, Marloes, Wim, Martine, Kim, Tamara, Marit, Margot, Jacqueline, Lysette, Mario, Gerdien en Margina. Daarnaast wil ik graag Jesper, Marien en Vincent bedanken voor mijn dagelijkse portie humor. Het was altijd lachen op jullie kamer! Maar ook buiten het werk zorgden jullie voor een hoop lol met de (foute) Fifa avonden. Zonder jullie had ik mezelf nooit Fifa zien spelen, bedankt!

Verder verdienen sommige collega's iets meer spotlight dan enkel een naam in een opsomming. Gertjan, wat hebben wij samen met Margot op ons 1e congres/summer school in Freiburg niet alleen een hoop informatie tot ons genomen over eye tracking maar ook een hoop met elkaar gelachen. Ons schnitzelbroodje als lunch leverde jaren later nog doorgestuurde foto's van Margot op met daarop (je raad het al…een schnitzelbroodje). Het hotelkamer delen met elkaar verliep zodanig prettig dat je het een jaar later voor het EARLI congres op Cyprus samen met Steven erbij opnieuw aandurfde om met mij een hotelkamer te delen. Wellicht was (achteraf gezien) 3 man in zo'n (achterlijk klein) hotelkamer toch wat ambitieus... We hebben hierna dan ook nooit meer een hotelkamer gedeeld. Desondanks wil ik je toch bedanken niet alleen voor het met elkaar nog oefenen en nalopen van onze presentaties maar ook voor al je (ja het waren er echt veel) flauwe grapjes. De wandeling door de grimmige straatjes van Limassol en aan het einde van het congres het doorhalen van de nacht zal ik nooit vergeten!

Steven, jij durfde het na Cyprus nog wel aan om met mij op congressen een hotelkamer te delen. Zo waren wij weer roommates in Washington voor de AERA. Ik zal het moment dat wij (samen met Marit) net na binnenkomst erachter kwamen dat de bar niet een gewone bar was maar een gaybar nooit meer vergeten. Ook je verjaardagsfeest een dag later in dezelfde bar was erg geslaagd. Echter denk ik nog steeds dat het een tactische fout van je was om een local daar te trakteren op een drankje. Met name Marloes (wiens krullerige haar door hem uitvoerig werd bestudeerd) en Marit (die later op de avond agressief werd uitgemaakt voor een puntjepuntje-blocker) zullen het daar vast met me over eens zijn. Ondanks de kater

hebben we (Marit, Steven en ik) de dag na het verjaardagsfeest nog een fantastische wandeling gemaakt door Washington. Ik wil jullie alle twee bedanken voor deze geweldige week.

Vincent, jij bent het laatste slachtoffer geweest m.b.t. het delen van hotelkamers. Ik wil je zowel voor de EARLI SIG 6/7 in Dijon als de EARLI in Finland bedanken voor de support en gezelligheid. Ook wil ik je bedanken voor het introduceren van mij aan Schillie in Dijon. Ik wacht overigens nog steeds op jouw excuses voor de valse beschuldiging dat wij (Steven en ik) jouw knuffelbeertje kwijt zouden hebben gemaakt in Finland. Gelukkig zag je het verder nog steeds zitten om op het werk kamergenoten te zijn☺.

In 2015 kwam ik te werken in Utrecht op de afdeling Educatie. Ik moest in het begin wel even wennen aan de nieuwe werkomgeving maar dankzij het warme welkom heb ik snel mijn plekje gevonden. Ik wil hiervoor dan ook iedereen van Educatie en O&T bedanken. Ook wil ik al mijn E3.36 kamergenoten bedanken voor het sparren en de gezelligheid: Luce, Marloes, Esther, Bas, Monika, Michelle, Reneé, Eva en Steven. Eva, nogmaals bedankt voor de fijne samenwerking voor onze taak als paranimf voor Steven. Zonder jouw input als medeparanimf hadden we nooit zo'n goed onderzoek kunnen opzetten voor Steven. Michelle, nog bedankt voor de fijne rondleiding in Oulu. Daarnaast wil ik ook graag alle leden van de Eye Track Seminars bedanken voor de super interessante presentaties en discussies van de afgelopen jaren.

Verder wil ik graag ook mijn vrienden bedanken voor de af en toe fijne afleiding van het promotietraject. Jurn, Gavin, Ewoud, Martijn en Olivier: ookal spreken we niet meer zo frequent af als vroeger, het is wel elke keer weer als vanouds gezellig. Bedankt! Boyd, bedankt voor de leuke (board) game avonden!

Martijn, als paranimf wil ik je graag apart bedanken voor je vriendschap en het delen van de mooie momenten van het leven. Ik zal nooit vergeten hoe jij als dienstdoende arts getuigen bent geweest van de bevalling van Jennifer en daarbij Chantal en mij hebben gesteund. Ik was daarom ook vereert dat ik een paar jaar later getuigen mocht zijn op jouw prachtige bruiloft. Dat we na zoveel jaar nog steeds goede vrienden zijn hebben we recent wel bewezen.

Twaalf jaar na onze 1e poging hebben we dan eindelijk samen de 24-uurs race van Le Mans succesvol uitgereden. We finally did it!

Steven, je naam is al meerdere keren gevallen maar als mijn persoonlijke paranimf wil ik je graag apart bedanken. Je bent in een rap tempo van een collega uitgegroeid tot een echte maat. Bedankt voor alle support als collega en als vriend! Sinds jouw promotie hebben we eigenlijk nauwelijks meer nasi-donderdagen gehad maar zijn we over gegaan op broodjes kip. Heb jij al een idee wat we de komende tijd op donderdag gaan eten? Glad to be your buddy, pal!

Naast mijn vrienden wil ik ook mijn familie bedanken. Rianne en Rick, bedankt voor de gezelligheid. Ik ben geneigd te schrijven "dat we nog velen après-ski themafeesten in jullie stamkroeg in Zeist mogen ervaren" maar ik denk dat we het er allebei wel mee eens zullen zijn dat we dat maar beter niet meer kunnen doen☺. Verder wil ik ook mijn vader Ab en Jolanda bedanken. Ab, bedankt dat je me altijd onvoorwaardelijk hebt gesteund in alle keuzes die ik heb gemaakt. Zonder die vrijheid was ik nooit geworden wie ik nu ben. Anja en Gerrit, vanaf de allereerste dag dat ik jullie heb leren kennen heeft de deur altijd voor mij open gestaan, bedankt! Anja, bedankt dat je twee dagen in de week op Jennifer wilt passen. Zo weet ik zeker dat Jennifer altijd in goede handen is!

De laatste en belangrijkste personen die ik wil bedanken zijn Chantal en Jennifer. Chantal, zonder jouw steun en hulp was ik nooit zover gekomen! Je lieve zorgzame karakter zorgde ervoor dat je mij sinds ik je heb leren kennen in 2003 altijd al wilde helpen waar dat kon. Inmiddels jaren later heb je me dan ook ontelbaar vaak geholpen. Wegens gebrek aan ruimte beperk ik me nu met wat specifieke voorbeelden tot de periode binnen m'n promotietraject. Zo heb je me de afgelopen jaren vaak geholpen met het afzetten en ophalen van mij. Zo heb je mij ontzettend vaak afgezet en opgehaald bij Schiphol, een paar keer bij Locus nadat de pubgroup iets te laat geworden was (sorry!), meerdere keren bij o.a. station Gouda als er weer gedoe was op het spoor, en zelfs een keer bij Krimpenerwaard college om ervoor te zorgen dat ik echt op tijd zou zijn voor het uitvoeren van mijn experiment. Maar belangrijker nog is de onvoorwaardelijke liefde, steun, ontspanning en energie die je me de afgelopen jaren hebt gegeven die me in staat stelde om ook bij tegenslagen weer verder te gaan. Chantal je bent mijn soulmate, mijn

rode knop die ik altijd kan indrukken als ik hulp nodig heb. Bedankt dat je altijd voor me klaar staat! Als laatste wil ik je bedanken voor het in ons leven brengen van onze prachtige dochter Jennifer. Jennifer, ik wil jou ook bedanken. Allereerst wil ik je bedanken voor het feit dat je gelukkig al snel hele nachten rustig doorsliep. Ookal je mij niet zo actief hebt geholpen zoals je moeder, is het effect van hele nachten doorslapen minstens zo groot! Daarnaast wil ik je ook bedanken voor de relativerende werking die je hebt op mij, zodat ik met beide benen op de grond blijf staan.

Publications

In press

Van Marlen, T., Van Wermeskerken., & Van Gog, T. (in press). Effects of visual complexity and ambiguity of verbal instructions on target identification. *Journal of Cognitive Psychology*. doi.org/10.1080/20445911.2018.1552700

2018

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. doi.org/10.1016/j.learninstruc.2018.07.005

2016

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. doi.org/10.1016/j.chb.2016.08.041

Submitted

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., Raijmakers, M., & Van Gog, T. (2018). *Looking through Sherlock's eyes: Effects of eye movement modeling examples with and without verbal explanations on deductive reasoning* Manuscript submitted for publication.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018). *Does the Purported Expertise of a Model Influence the Effectiveness of Eye Movement Modeling Examples?* Manuscript submitted for publication.

Presentations

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018, August). *Does the Purported Expertise of a Model Influence the Effectiveness of Eye Movement Modeling Examples?* Paper presented at the EARLI Special Interest Group 6 and 7, Bonn, Germany.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018, June). *Does the Perceived Expertise of a Model Influence the*

Effectiveness of Eye Movement Modeling Examples? Paper presented at EARLI Special Interest Group 27, Warsaw, Poland.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2017, August). *Effectiveness of Eye Movement Modeling Examples: The Role of Instructional Ambiguity*. Symposium paper presented at EARLI 2017, the 17th biennial Conference of the European Association for Research on Learning and Instruction, Tampere, Finland.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2017, August). *Look Where Eye Looked: Eye Movement Modeling Examples Enhance Learning to Solve Geometry Problems.* Symposium paper presented at ECEM, Wuppertal, Germany.

Van Wermeskerken, M., **Van Marlen, T**., & Van Gog, T. (2016, November). *Follow my Gaze: Effects of instructor gaze cues in video examples on attention and learning outcomes*. Paper presented at EARLI Special Interest Group 27, Oulu, Finland.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016, November). *The Role of Ambiguous Instructions in the Effectiveness of Eye Movement Modeling Examples*. Poster presented at EARLI Special Interest Group 27, Oulu, Finland.

Van Marlen, T., Van Wermeskerken, M., Boven, L., Ketelaar, J., van Trijp, K., Vernooij, S., & Van Gog, T (2016, August). *Are eye movement modeling examples most useful when verbal instructions are ambiguous?*. Paper presented at the EARLI Special Interest Group 6 and 7, Dijon, France.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016, April). *Can eye movement modeling examples foster learning of procedural problem-solving tasks?*. Poster presented at AERA, Washington, USA.

Van Marlen, T., Van Wermeskerken, & Van Gog, T.(2015, August). *Give me a cue: Effects of visual complexity and description ambiguity on attention.* Paper presented at the JURE 2015, the biannual pre-conference of the Junior Researchers of EARLI, Limassol, Cyprus.

Van Marlen, T., Van Wermeskerken, & Van Gog, T.(2015, August). *Give me a cue: Effects of visual complexity and description ambiguity on attention.* Poster presented at ECEM, Vienna, Austria.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2015, August). *Can eye movement modeling examples foster learning?*. Paper presented at the EARLI conference, Limassol, Cyprus.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2014, August). *Can eye movement modeling examples foster geometry learning?*. Poster presented at ESSEM, Freiburg, Germany.

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2014, August). *Can eye movement modeling examples foster geometry learning?*. Paper presented at the EARLI Special Interest Group 6 and 7, Rotterdam, The Netherlands.

Curriculum Vitae

Tim van Marlen was born on January 6 1988 in Rotterdam, The Netherlands. After he graduated from high school in 2007 (Montessori Lyceum Rotterdam), he studied Psychology at the Erasmus University Rotterdam. He obtained his Bachelor's degree in Biological and Cognitive Psychology in 2010, followed by a Master's degree in Biological and Cognitive Psychology (2012, cum laude). During his Master Tim was also enrolled in the Advance Research Program, which was aimed at further fostering research skills. Subsequently, he worked as a research assistant at the Erasmus University Rotterdam in which he assisted conducting experiments about human memory. He started his PhD at the Institute of Psychology at the Erasmus University Rotterdam in 2014 and continued working on his PhD at the Department of Education at Utrecht University in 2015. The PhD project was funded by a Vidi grant from the Netherlands Organization for Scientific Research (NWO) and focused on the effects of learning procedural problem-solving from Eye Movement Modeling Examples (EMME). During his PhD, he presented his work at various international conferences. Tim is currently working as a Postdoctoral Researcher at the Department of Education at Utrecht University.