EYE(S) SEE WHAT YOU DO:

The Role of Social Mechanisms in the Effectiveness of Eye Movement Modeling Examples as an Instructional Tool for Multimedia Learning

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Dipl.-Psych. Marie-Christin Krebs
aus Überlingen

Tübingen
2019

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	01.07.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Katharina Scheiter
2. Berichterstatter:	Prof. Dr. Stephan Schwan

Acknowledgements

I want to thank everyone who supported me along the way of my dissertation. Thank you for taking the time to discuss research questions and to give feedback and advice.

First of all, I would like to thank the Leibniz-Institut für Wissensmedien for offering me an excellent research environment. Especially, I want to thank my supervisors Prof. Dr. Katharina Scheiter and Dr. Anne Schüler for their continuous support during my Ph.D. project. Their feedback and expert advice were invaluable in all phases of the dissertation and have contributed significantly to the improvement of the present work.

I also want to thank Prof. Dr. Stephan Schwan who agreed to be the second reviewer of this dissertation. Further thanks goes to Dr. Annika Scholl and Dr. Yvonne Kammerer for giving valuable feedback on my exposé and progress reports. Moreover, I want to thank Dr. Margot van Wermeskerken for extended discussions on the research topic.

I also would like to thank my former colleagues from Multiple Representations Lab for lively discussions, their feedback on (conference) presentations and for the good time we have had in the last years. A special thanks goes to Katrin Schleinschok and Dr. Steffen Schmidgall for the great time spent together in the office, for their support and their friendship.

Another special thanks goes to Dr. Hannah Greving and Dr. Lara Ditrich for many intensive discussions, support in difficult times and continuous encouragement along the way. Thank you for your friendship, for your support and for reminding me that we also have a life outside the office! I also want to thank Romy Brömme for her encouragement and support over the years. Thank you!

Moreover, I want to thank the research assistants for their support during data collection and the participants for participating in my experiments. None of the experiments would have taken place without you.

Last but not the least, I would like to thank my family. Especially I want to thank my parents, my brother and my grandparents for their constant encouragement. Without you, this dissertation would never have been possible.

Declaration on Contributions to Monography

Although the dissertation is written as a monography, it includes content of three manuscripts that are already published or are ready to submit. Passages of the manuscripts are found in the dissertation's theoretical chapters (Chapter 1 to 4) and General Discussion (Chapter 8). The empirical findings of Manuscript 1 are presented in Chapter 5. Chapter 6 includes the findings of Manuscript 2 and Chapter 7 includes the findings of Manuscript 3. Prof. Dr. Katharina Scheiter and Dr. Anne Schüler are co-authors of these manuscripts. Their proportional contributions to the manuscripts are presented in the subsequent tables.

Manuscript 1

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Marie-Christin Krebs	first author	70 %	100 %	100 %	70 %
Anne Schüler	second author	15 %	0 %	0 %	20 %
Katharina Scheiter	third author	15 %	0 %	0 %	10 %
Title of paper:		Eye Movement Modeling Examples – Do social cues really matter?			
Status in publication process:		ready to submit			

Manuscript 2

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Marie-Christin Krebs	first author	30 %	100 %	100 %	60 %
Anne Schüler	second author	35 %	0 %	0 %	20 %
Katharina Scheiter	third author	35 %	0 %	0 %	20 %
Title of paper:		Just follow my eyes: The influence of model-observer similarity on Eye Movement Modeling Examples.			
Status in publication process:		published: Learning and Instruction, 61, 126–137. https://doi.org/10.1016/j.learninstruc.2018.10.005			

Manuscript 3

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Marie-Christin Krebs	first author	60 %	100 %	100 %	60 %
Anne Schüler	second author	15 %	0 %	0 %	20 %
Katharina Scheiter	third author	25 %	0 %	0 %	20 %

Title of paper:	Do prior knowledge, model-observer similarity and social comparison influence the effectiveness of Eye Movement Modeling Examples for supporting multimedia learning?
Status in publication process:	ready to submit

Table of Contents

1 Introduction

Illustrated (scientific) exists since pre-Roman antiquity and over the centuries it has been used to provide readers with information that would have been too complex to be expressed only in written form. Often the term ‘multimedia’ is used for illustrated text (Mayer, 2014a), which describes materials that present information to learners in verbal (written or spoken text) and pictorial form (e.g., static or dynamic pictures). Nowadays, the use of illustrated text is omnipresent in the educational context. It is part of learning materials in textbooks, e-books, animations, and learning software. In the school context, examples for multimedia materials can be found in traditional scientific textbooks. Here, the learning content is often presented as an expository text that is accompanied by a single static picture or a sequence of static pictures. However, multimedia material is not only important in a school context. Further examples of multimedia materials can also be found in illustrated journal articles, video tutorials or instruction manuals in other areas such as training and higher education, as well as in everyday life.

Over many years a large field of research evolved to investigate the effectiveness of learning with multimedia. Theories of multimedia learning such as the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009) or the Integrated Model of Text and Picture Processing (ITPC; Schnotz, 2005) postulate that learning with text-picture combinations results in a richer mental representation of the learning content. In line with the assumptions of these theories, previous research offers empirical evidence that learning with multimedia is more effective than learning with text alone (for reviews see Anglin, Vaez, & Cunningham, 2004; Levie & Lentz, 1982; Mayer, 2009). The empirical finding that learning with text-picture combinations results in better learning outcomes than learning with text alone is called the *multimedia effect* (Mayer, 2009; for review see Butcher, 2014). According to the CTML as well as the ITPC, successful learning with multimedia requires learners to actively perform certain cognitive processing strategies as prerequisites for mental model construction (Mayer, 2014a). Learners have to select and organize relevant pieces of information from the text as well as from the picture, form a mental representation of the text information and the picture information, respectively, and integrate these mental representations with their prior knowledge into a coherent mental model. Thus, learning with multimedia can be cognitively demanding and challenging for learners. This is underlined by empirical findings showing that without instructional guidance not all learners benefit from multimedia (e.g., Hannus & Hyönä, 1999; Scheiter & Eitel, 2015; Schmidt-Weigand, Kohnert, & Glowalla, 2010).

As a consequence, multimedia research is not only interested in the underlying cognitive processes of multimedia learning but also in developing instructional support measures for multimedia learning. Renkl and Scheiter (2015) distinguish between material-oriented interventions and learner-oriented interventions. While the former focus on adjusting the learning materials to learners' needs (e.g., reducing complexity; cueing/signaling; physically integrating text and pictures), the latter focus on the learners and their individual prerequisites (e.g., pre-training, prompting). One of these learner-oriented interventions are Eye Movement Modeling Examples (EMME). EMME are videos with gaze replays of another person, usually an expert model. These gaze replays are recorded while the model is performing a specific (learning) task. In the gaze replays the model's fixations are overlaid onto the stimulus. These fixations are visualized in the gaze replays as a visual cue that changes its size dynamically depending on the fixation duration. The gaze replays are then shown to other learners or problem solvers as instructional materials. It is assumed that by observing when, where and for how long the model looked at certain parts of the task material, observers (usually novice learners) can infer underlying processing strategies which will help them to perform a similar or identical task (cf., Veenman, 2011). EMME have been shown to be effective for various tasks and domains (e.g., Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Salmerón & Llorens, 2018; Mason, Pluchino, & Tornatora, 2015, 2016).

As a means to enhance multimedia learning, EMME provide learners with suitable processing strategies, which they can then apply in a self-regulated manner during the learning phase. Empirical findings for EMME in the context of multimedia learning suggest that EMME are not only beneficial for learning specific tasks, but also for learning or triggering general processing strategies (Mason, Pluchino, et al. 2015; Mason et al., 2016; Mason, Scheiter, & Tornatora, 2017; Scheiter, Schubert, & Schüler, 2017). Although these findings show that in general, EMME are effective for multimedia learning, it remained unclear which mechanisms underlie the effectiveness of EMME for multimedia learning. As a consequence, the present dissertation seeks to shed light on these mechanisms to close this research gap.

Previous research has used several theories to explain the effectiveness of EMME. To investigate the influence of specific mechanisms and / or individual factors that can be derived from these mechanisms and to draw conclusions about them with regard to the effectiveness of EMME, it is necessary to first separate the various explanations from the different theories. For this purpose, I differentiate between three perspectives regarding possible underlying mechanisms for the effectiveness of EMME for multimedia learning: The *perceptual learning perspective*, the *(meta-) cognitive perspective*, and the *social perspective*. Each of these perspectives suggests possible mechanisms that may underlie the effectiveness of EMME. For instance, from a perceptual learning perspective EMME foster learning because they guide learners' visual attention to relevant task information and highlight the order in which the information should be processed (Goldstone, Marghetis, Weitnauer, Ottmar, & Landy, 2017). From a (meta-) cognitive perspective, EMME not only guide visual attention, but also foster higher-level cognitive processes. Hence, EMME are assumed to support learners in acquiring new cognitive processing strategies (Mason, Pluchino, et al. 2015, Mason et al., 2016) and / or in better regulating already existing ones (Mason et al., 2017; Scheiter et al., 2017). From a social perspective, EMME can be regarded as a special case of video-based modeling (Jarodzka et al., 2013) and foster learning because they create a social (learning) situation by using human eye movements as attention-guiding cue. It is assumed that if learners allege that they observe a problem-solving or performance procedure that is performed by human model, this alleged anthropomorphism, in turn, foster deeper learning (social agency theory; Mayer, 2014b; Mayer, Sobko, & Mautone, 2003; Moreno, Mayer, Spires, & Lester, 2001). Although each of the three perspectives offers a specific focus to explain the effectiveness of EMME, it is important to note that the three perspectives are complementary rather than mutually exclusive.

With regard to the different perspectives described above, previous research has relied particularly on factors from a perceptual learning and / or a (meta-) cognitive perspective to explain positive effects of EMME (e.g., Jarodzka et al., 2013; Mason, Pluchino, et al. 2015, Mason et al. 2016, 2017; Scheiter et al., 2017). That is especially surprising as most of this previous research claims EMME to be theoretically grounded in the social cognitive learning theory (Bandura, 1986). To my knowledge, there has been only one study by Litchfield et al. (2010) so far that considered possible consequences of models' and observer's expertise on the effectiveness of EMME for supporting radiographers' diagnostic performance. However, up to now there has been a lack of research focusing on the social aspects of EMME and possible consequences for the effectiveness of EMME for multimedia learning. Hence, the present dissertation focuses on the question of whether social mechanisms and factors contribute to explaining the effectiveness of EMME for multimedia learning above and beyond perceptual and / or (meta-) cognitive mechanisms and factors.

Two different approaches were used for this purpose. In the first approach I attempted to isolate the effect of social mechanisms. For this purpose, EMME were compared to an attention guidance tool without social cues. It was assumed that if social mechanisms contribute to explaining the effectiveness of EMME for multimedia learning, EMME with a social gaze cue (i.e., human eye movements) would be more effective than a comparable attention guidance tool with a non-social attention guidance cue. Based on this hypothesis, it was investigated whether learners with EMME support show more effective multimedia processing strategies (reflected in their gaze behavior as well as better learning outcomes) than learners with a non-social attention guidance tool. In the second approach it was assumed that if social mechanisms underlie the effectiveness of EMME, social factors should moderate the effectiveness of EMME. In contrast, if the effectiveness of EMME is solely based on perceptual and / or (meta-) cognitive mechanisms, social factors should not influence their effectiveness. The second approach involved to experimentally manipulate the salience of social factors and / or to assess social factors as potential moderators.

In the following I will outline the theoretical background of the research presented in this thesis, starting with multimedia learning and its requirements and challenges for learners. Further, I will present eye tracking as a research tool in the context of multimedia learning and focus on empirical results that show the connection between cognitive processing and human gaze behavior regarding learning performance in multimedia learning. Then, I will introduce EMME as a learner-oriented intervention for multimedia learning that support learners in processing multimedia material more adequately. First, empirical evidence for the effectiveness of EMME for task performance and learning in general as well as multimedia learning in particular is provided. This is followed by the theoretical foundations of EMME. Based on the theoretical foundations, I will present and discuss possible mechanisms underlying the effectiveness of EMME for multimedia learning from a perceptual learning, a (meta-) cognitive and a social perspective on EMME. Further, I will also discuss possible factors that may influence the effectiveness of EMME that can be derived from these perspectives. Against this theoretical background, I will give an overview of the research questions underlying the research presented in this thesis and briefly outline the three studies that were conducted. In the second part of the dissertation, I will present and discuss empirical findings of the three studies in more detail. Finally, the results of the three studies are summarized and their implications are discussed in the light of the research questions presented. Subsequently, practical implications are derived from the empirical findings and strength and limitations of the present dissertation will be outlined.

2 Learning with Multimedia

2.1 Cognitive Theories of Multimedia Learning

The term ‘multimedia’ describes material that presents information in verbal and pictorial form (Mayer, 2014a). According to Mayer, “multimedia learning occurs when people build mental representations from words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video)” (Mayer, 2014a, p. 3). Cognitive theories of multimedia learning such as Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009) or the Integrated Model of Text and Picture Processing (ITPC; Schnotz, 2005, 2014) postulate that learning with multimedia results in better understanding of the learning content.

The CTML makes three basic assumptions regarding multimedia learning (Mayer, 2014a). The first assumption is the *dual-channel assumption* which states that the human information processing system consists of two information processing channels: an auditory / verbal channel and a visual / pictorial channel (Mayer, 2014a). This assumption of two different channels is based on two different approaches to explain how multimedia material is processed within the human cognitive system. The first approach is the sensory-modality approach which is – according to Mayer (2009) – most in line with Baddeley’s (1999) conceptualization of working memory. This approach focuses on the modality of the presented information. It distinguishes between information that is initially processed through the eyes (e.g., pictures or printed words) or initially processed through the ears (e.g., spoken words). The second approach, the representation-mode approach, focuses on whether a verbal (e.g. spoken or printed words) or nonverbal (such as pictures, video, animation, or background sounds) stimulus is presented. According to Mayer (2009), this approach is most in line with Paivio’s (1986) distinction between a verbal and a nonverbal information processing system. Referring to both approaches, the CTML assumes that the human cognitive system consists of an auditory-verbal channel and a visual-pictorial channel. Initially, incoming information is processed according to its modality within these channels (i.e., spoken text and sounds in the auditory part of the auditory-verbal channel; pictures, animations, videos and written text in the visual part of the visual-pictorial channel). Subsequently, information is processed according to its representation mode, meaning that all verbal information is processed in the verbal part of the auditory-verbal channel and all pictorial information is processed in the pictorial part of the visual-pictorial channel.

The second assumption is the *limited-capacity assumption* which presumes that humans can only process a limited amount of information in each channel at a time (Mayer, 2014a).

The third assumption - the *active processing assumption* - describes the assumption that in order to build a coherent mental representation of the multimedia material, learners have to be actively engaged in information processing and carry out a coordinated set of five cognitive processes during the learning phase (Mayer, 2014a).

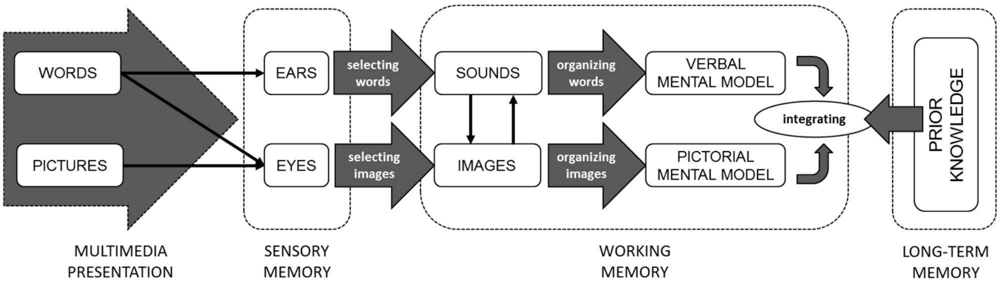


Figure 1: Cognitive Theory of Multimedia Learning (adapted from Mayer, 2014a, p. 52).

These processes are crucial for constructing coherent mental models and therefore for learning from multimedia materials (see Figure 1). The processes are: (1) selecting relevant words from the text; (2) selecting relevant picture elements from the picture; (3) organizing the selected verbal information within working memory into a coherent verbal representation of the text information; (4) organizing the selected pictorial information into a coherent pictorial representation of the pictorial information; (5) integrating both mental representations with each other as well as with relevant prior knowledge activated from long-term memory into a coherent mental model. Because the integration process involves the coordination of both the pictorial and verbal working memory, it is "an extremely demanding process that requires the efficient use of cognitive capacity" (Mayer, 2014a, p. 57). At the end of the process, the constructed mental model should represent the key elements as well as their relations to each other. According to Mayer (2014a), active learning occurs when learners apply the above described cognitive processes to the learning material.

Similar to the CTML, also the Integrated Model of Text and Picture Comprehension (ITPC; Schnotz, 2005; 2014) postulates that multimedia learning requires a multistage cognitive process and that during this process, learners form a coherent mental model of the presented material by connecting verbal and pictorial information. As stated by Schnotz (2014), meaningful learning from multimedia material requires learners to select and organize information, and actively integrate information from different sources. Similar to the assumptions made by the CTML, the ITPC assumes that the human cognitive architecture consists of three components: a modality-specific sensory memory, a working memory, and a long-term memory (see Figure 2). Also comparable to the CTML, the ITPC has the dual channel assumption of initial processing and transmitting verbal and pictorial information in an auditory and a visual register.

However, both theories differ in their assumptions regarding the interaction of verbal and pictorial information. Whereas the CTML suggests that the verbal and visual information do not interact in working memory before the final step of integration (see Figure 1), the ITPC proposes that the verbal and visual information interact as soon as the information is transferred to the respective subsystems in working memory (see Figure 2).

According to Schnotz (2002), texts and pictures are based on different sign systems. Hence, for processing multimedia information, the ITPC model differentiates between a descriptive branch and a depictive branch. The descriptive branch includes the external verbal information, the internal mental representation of this verbal information, and a propositional representation of the semantic content. The depictive branch includes the external pictorial information, the internal visual perception of the picture, and the mental model of the content that is depicted in the picture.

According to the ITPC, learners form a coherent mental model of text and picture (i.e., coherence formation) by a continuous interplay between the propositional representations of the semantic content and the mental model of the content that is depicted in the picture. The coherence has to be established between verbal and pictorial representation (cf. inter-representational coherence formation; Seufert, 2003) on the one hand, and with regard to the information within each representation (cf. intra-representational coherence formation; Seufert, 2003) on the other. Moreover, in both cases coherence has to be established at the surface level as well as the level of semantic deep structures (Schnotz, 2014). Whereas surface structure mapping describes the process of connecting elements of verbal (e.g., words) and pictorial representations (e.g., shapes), semantic deep structure mapping describes the process of establishing connections between conceptual structures (e.g., simple/complex relations between elements; cf. Schnotz, Ludewig, Ullrich, Horz, McElvany, & Baumert, 2014). Based on the ITPC, Schnotz (2014) states that especially learners with lower prior knowledge benefit from a picture, because it is more demanding for them to construct a mental model based only on verbal information when their existing internal source of information is poor.

Although some aspects and assumptions between the CTML and the ITCP differ considerably, the two theories also share some characteristics. Both assume that the successful cognitive processing of multimedia material requires a multistage cognitive process.

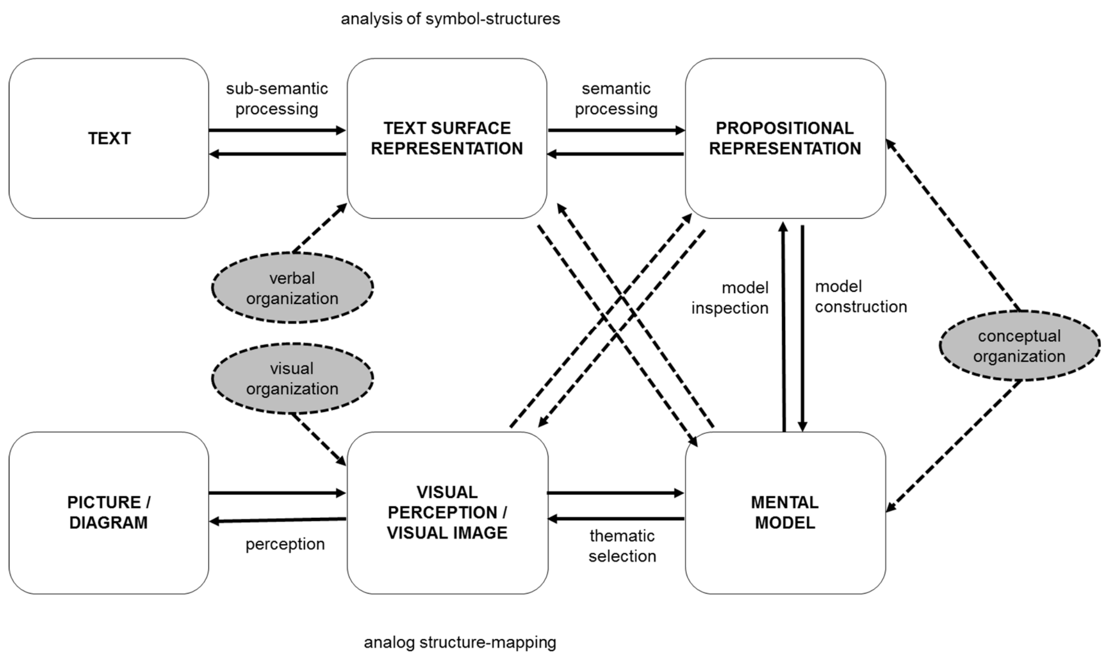


Figure 2: Theoretical framework for the integrative model of text and picture comprehension (adapted from Schnotz & Bannert, 2003, p. 145)

First, learners have to derive relevant information from the text and the picture (i.e., *selection processes*) and organize the information in working memory (i.e., *organizing processes*). Then, they have to integrate the information from the different sources with each other and with their prior knowledge (i.e., *integration processes* or *coherence formation processes*). Both theories suggest that the advantage of learning with multimedia material in contrast to learning with only one representation is that through the integration process / coherence formation process, learners gain a richer mental representation of the learning content. For instance, an advantage of multimedia material is that when learners try to retrieve the processed information from memory later, multimedia information can be retrieved by either the verbal, or the pictorial, or the integrated code. In case of only one representation (e.g., text only), there is also only one code that can be accessed in memory (Paivio, 1986). Another advantage is that the integration of prior knowledge results in a more integrated and elaborate representation of the content (Mayer, 2014a). This allows for the information represented to be retrieved also across longer time spans.

Over many years a large field of research evolved to investigate the effectiveness of learning with multimedia. This research yielded a lot of empirical evidence that learning with multimedia is more effective than learning with text alone (for reviews see Anglin, Vaez, & Cunningham, 2004; Levie & Lentz, 1982; Mayer, 2009). This finding is called the *multimedia effect* (Mayer, 2009).

However, there is also empirical evidence that not all learners benefit from multimedia (e.g., Hannus & Hyönä, 1999; Scheiter & Eitel, 2015; Schmidt-Weigand et al., 2010). Based on the theories described above, one possible explanation for why some learners do not benefit from multimedia materials is that they might fail to apply the necessary cognitive processes. One possibility to investigate this hypothesis is the use of eye tracking, which allows online-assessment of learners' eye movements, thereby providing information about a learner's cognitive processing during the learning process. In the following section, I describe the idea behind eye tracking as research tool and discuss findings from studies using eye tracking as online-based measurement in multimedia learning.

2.2 Eye Tracking as Online-based Measurement for Multimedia Learning

2.2.1 An introduction to eye tracking

In recent years, eye tracking has become a new measurement method in educational research (for reviews, see Alemdag & Cagiltay, 2018; Lai et al., 2013). Initially used primarily in reading research (for a review see Rayner, 1998), eye tracking is nowadays also used in other (educational) research areas such as problem solving (e.g., Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009; Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016), visual search and visual learning tasks (e.g., Dzeng, Lin, & Fang, 2016; Jarodzka et al., 2012; Litchfield et al., 2010), visual expertise (e.g., Gegenfurtner & van Merriënboer, 2017), and also in multimedia learning (e.g., Boucheix & Lowe, 2010; Boucheix, Lowe, Putri, & Groff, 2013; Desjarlais, 2017; Eitel, 2016; Eitel, Scheiter, & Schüler, 2013; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; Hyönä, 2010; Jamet, 2014; Mason, Pluchino et al., 2015, Mason et al., 2016, 2017; Schüler, 2017).

Tracking people's eye movements using technical devices has a longstanding tradition since the late 1800s (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, & Van de Weijer, 2011). Due to constant technical developments in both hardware and software, modern eye tracking devices are relatively user-friendly, non-invasive, and usually provide good data quality. In the eye tracking studies reported in this thesis, a video-based remote eye tracking system was used to measure participants' eye movements. Video-based remote eye tracking systems assess the eye movements using the pupil- and corneal-reflection method (for an overview of eye tracking methods see Holmquist et al., 2011). It is based on measuring the corneal reflection of a (infra-red) light source relative to the location of the pupil center (Duchowski, 2007). To record the eye movements, the eye tracker is positioned below a (computer) screen on which the material is displayed. During the presentation of the material, an infra-red light that is aimed at the eyes of the person sitting in front of the eye tracker produces a corneal reflection. This reflection is then recognized by the system as the brightest spot on the eye. By contrast, the pupil is identified as the darkest spot on the eye. A change in the distance between the lightest and the darkest spot indicates that a movement of the eyeball has occurred. The total distance between the eyes and the stimulus screen, the coordinates of the monitor and the change in distance between the brightest and darkest point on the eye are then used to calculate the gaze direction. Then, the eye tracking system's software aggregates the recorded data into eye tracking parameters (Duchowski, 2007; Holmqvist et al., 2011).

There are three different kinds of eye tracking parameters that are recorded: fixations, saccades, and smooth pursuit eye movements (Duchowski, 2007; Holmqvist et al., 2011). Fixations are the period of time during which the eye stabilizes the retina over a stimulus. They typically last about 200-500 ms. During this time period, the central foveal vision is held in place over the object of interest and visual information about it enters the information processing system. Saccades are rapid eye movements that reposition the eye so that a new part of the stimulus is displayed on the fovea. Saccadic eye movements can be executed voluntary or as a corrective measure. For the duration of the saccade (10 ms to 100 ms), the visual information intake of a person is suppressed (Duchowski, 2007; Rayner, 1998). In contrast to saccades, smooth pursuit eye movements are slower eye movements. Their purpose is to keep a moving stimulus on the fovea (Purves, Augustine, Fitzpatrick, Katz, LaMantia, McNamara, & Williams, 2001). Smooth pursuit eye movements are usually executed voluntary (Purves, et al., 2001). To later interpret the recorded eye data, so-called areas of interest (AOIs) are defined around elements of the (learning) material. For multimedia material, for example, an AOI can be a text segment, a picture, a graph, or a part of a picture or graph. For each AOI, one can compute different measures that describe a person's gaze behavior during a specific task. For instance, in the multimedia context, among the more important measures are the time spend on a specific AOI (fixation time), the number of fixations on a specific AOI (fixation count), the time until a specific AOI is fixated (time to first fixation), and the saccades between two AOIs (number of transitions; for other eye tracking measures in the multimedia context, see also Alemdag & Cagiltay, 2018).

According to the *eye-mind assumption* (Just & Carpenter, 1980) and the *eye-mind link* (Reichle, Pollatsek, & Rayner, 2006), fixation location (i.e., overt attention) and cognitive attention (i.e., covert attention) are closely related. It is assumed that saccades, fixations, and smooth pursuits reflect voluntary, overt visual attention (Duchowski, 2007). By showing the center of people's visual attention, these gaze parameters reflect cognitive processes and provide us with a 'window into the mind' (Van Wermeskerken, Litchfield, & Van Gog, 2018). Against this background, fixations and smooth pursuits are usually interpreted as "desire to maintain one's gaze on an object of interest" while saccades are interpreted as the "desire to voluntarily change the focus of attention" (Duchowski, 2007, p. 47). Based on the assumption that fixation duration indicates attention (Wang, 2011) as well as the amount of cognitive processing (e.g., Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Just & Carpenter, 1980; Rayner, 1998), tracking when and in what order learners attend to certain information offers a possibility to gain insight into learners' cognitive processes during learning, without interfering with the learning process itself. Furthermore, it can also give researchers the opportunity to detect inadequate use of cognitive strategies as well as information about when providing instructional support might be helpful for the learners.

Since the interpretation of the eye tracking measures (e.g., fixation time, time to first fixation) always depends on the research question, the following section gives an overview of eye tracking measures in relation to research questions and theories in the field of multimedia learning.

2.2.2 Eye tracking in multimedia learning

As stated by Johnson and Mayer (2012), there are different kinds of eye tracking measures that are closely related to the cognitive processes of selecting, organizing, and integrating information described in the CTML (Mayer, 2009, 2014a). A short overview of the measures discussed in this section is displayed in Table 1 (for a more elaborate summary and discussion of the use of different eye tracking measures in previous multimedia research see Alemdag & Cagiltay, 2018).

According to the CTML, to create coherent verbal or pictorial mental models in working memory, learners have to select and organize the information from the respective representation. In line with the eye-mind assumption (Just & Carpenter, 1980), it is assumed that eye tracking indices which are related to the amount of attention that is allocated to text or picture elements (e.g., fixation counts, fixation times) mirror cognitive selection and organization processes (Scheiter & Eitel, 2015). Thus, fixation measures such as fixation time or fixation count can be interpreted as indicators for selection and organization processes during the learning process (see Table 1).

Accordingly, more frequent and longer total fixation times on certain elements can be interpreted as indicators for more intense processing of these materials. In line with this assumption, empirical research indicates that increased attention to relevant information is positively related to learning performance in multimedia learning (e.g., Hannus & Hyönä, 1999; O'Keefe, Letourneau, Homer, Schwartz, & Plass, 2014; Scheiter & Eitel, 2015; She & Chen, 2009). For example, empirical findings of Scheiter and Eitel (2015) showed that extensive processing of certain information elements fostered learning. In a study with 55 university students, they investigated whether signals foster learning from text and pictures by examining the relationship between visual attention and learning outcomes. The results showed that learners attended to the signaled information earlier and also processed the signaled information more intensively, which in turn fostered comprehension.

An interesting finding of studies using eye tracking to study multimedia learning is the observation that learners often prefer text information over picture information (Ho, Tsai, Wang, & Tsai, 2014; Johnson & Mayer, 2012; Liu & Chuang, 2011; Schmidt-Weigand et al., 2010). This is especially important, as growing empirical evidence suggests that picture processing fosters multimedia learning (e.g., Eitel et al., 2013; Lin, Holmqvist, Miyoshi, & Ashida, 2017). For instance, Eitel et al. (2013) demonstrated that inspecting a picture even only for a short time fostered recall and comprehension performance. Hence, especially picture processing should be supported by instructional devices.

In addition to using the total fixation time as an indicator of the occurrence of selection and organization processes, there is also the possibility to split the overall fixation time in first-pass and second-pass fixation times. First-pass fixation times describe the total duration of fixations on the text AOI and the picture AOI for the initial text reading or picture inspection (Mason, Tornatora, & Pluchino, 2013, see Table 1). First-pass fixation times are usually interpreted as learners' selective attentional focus. In contrast, second-pass fixation times are usually interpreted as learners' more intentional and strategic processing (Mason et al., 2013) such as their attempt to integrate text and picture information (e.g., look-froms, see Table 1).

Integration of text and picture elements is another important cognitive process of multimedia learning (Mayer, 2009). Regarding gaze behavior, gaze shifts between different elements are interpreted as indicator for integration processes (e.g., Arndt, Schüler, & Scheiter, 2015; Hegarty & Just, 1993; Johnson & Mayer, 2012; Scheiter & Eitel, 2017). There are two eye tracking measures that are usually used as indicator for integration processes. One measure is the number of saccades between different elements (i.e., number of text-picture transitions, see Table 1). Another measure is the summed fixation duration either of picture fixation duration after rereading the text (i.e., look-from text to picture, see Table 1) or the text rereading duration after re-inspecting the picture (i.e., look-from picture to text, see Table 1).

Hegarty and Just (1993) were among the first researchers to examine the integration of text and picture using eye tracking. They investigated the gaze behavior of students when learning from an illustrated science text about a pulley system. Their results support the assumption that learners construct mental models of the learning content in a stepwise manner, as in their study learners first attended to the text, then switched to the picture, and then again back to the text. Furthermore, Hegarty and Just (1993) revealed that good learners displayed more integrative processing behavior, indicated by more frequent switches to the accompanying picture after having read a text unit.

Table 1. Eye Tracking Measures and Related Cognitive Processes (based on Alemdag & Cagiltay, 2018; Scheiter & Eitel, 2017; Johnson & Mayer, 2012; Mason et al., 2017; Mason et al., 2013)

Cognitive process	Measure	Description
Selection / Organization of pictorial information: Attentional focus on the picture information	Fixation count on the picture	Total number of fixations on the picture
	Total picture fixation time	Total duration of learner's fixation on the picture
	First-pass fixation time on the picture	Total duration of all fixations on the picture during the initial exploration
Selection / Organization of verbal information: Attentional focus on the text information	Total text fixation time	Total duration of learner's fixation on the text
	Fixation count on the text	Total number of fixations on the text
	First-pass fixation time on the text	Total duration of all fixations on the text during the initial exploration
Integration: Attempts to integrate words and pictures	Number of text-picture transitions	Number of times learners shift their eye fixation from the text to the picture and vice versa
	Duration of look-from text to picture and look-from picture to text	Summed fixation duration of picture fixation duration after rereading the text, and text rereading duration after re-inspecting the picture

Since this pivotal study, several eye tracking studies have provided further support regarding the importance of text-picture integration for multimedia learning (e.g., Mason, Pluchino, et al., 2015, Mason et al., 2016, 2017; Mason, Tornatora, & Pluchino, 2013, 2015; O'Keefe et al., 2014; Scheiter & Eitel, 2015; Schüler, 2017). For example, Mason et al. (2013) examined in a study with 49 fourth graders their processing of text and pictures while they read an illustrated science text. Based on a cluster analysis using eye tracking indices of first- and second-pass fixation times and integrative saccades, they found three patterns of gaze behavior that differed in the level of integration of text and picture. Learners in the first cluster showed very few transitions, learners in the second cluster an intermediate amount, and learners in the third cluster showed many transitions between text and picture. Further analyses revealed a significant relation between learners' gaze behavior and their performance in various learning tasks, indicating that learners who showed more integrative processing of the illustrated text also achieved higher learning outcomes. In addition, Mason, Tornatora et al. (2015) showed in another study with 43 seventh graders that differences in the frequency of transitions between text and picture as well as in the fixation duration of look-froms from text to picture (i.e., re-reading text information while re-inspecting the picture) were predictive for verbal and pictorial recall as well as for transfer performance.

To sum up, it is assumed that there are several eye tracking measures that can be used as indicators for cognitive multimedia processes (for reviews see Alemdag & Cagiltay, 2018; Lai et al., 2013). The aforementioned eye tracking studies indicated beneficial effects of integrative processing as well as picture processing on learning outcomes (e.g., Hannus & Hyönä, 1999, Mason et al., 2017; Scheiter & Eitel, 2015). These studies can be taken as evidence that especially the selection and organization of pictorial information as well as the integration of text and picture are crucial processes for learning successfully from multimedia. However, there is also empirical evidence that learners without instructional guidance are often unable to make optimal use of multimedia materials and fail in building coherent mental models, which results in poor learning outcomes (Scheiter & Eitel, 2015). As a consequence, research has addressed the question how to support learners in adequately processing multimedia materials (Scheiter & Eitel, 2015; Schlag & Ploetzner, 2011). According to Renkl and Scheiter (2015), one can distinguish between material-oriented interventions focusing on altering the learning materials (e.g., reducing complexity; cueing/signaling; physically integrating text and pictures) and learner-oriented interventions focusing on the learners themselves and how they can be enabled to draw most benefit from multimedia materials (e.g., training of learning prerequisites, pre-training, and prompting).

In the present dissertation, I used a learner-oriented intervention, namely EMME, to support learners' self-regulated use of effective multimedia processing strategies (Scheiter et al., 2017). More in detail, to investigate the influence of social cues in EMME on learners' gaze behavior, I selected two types of eye tracking measures: the total fixation time on the verbal and pictorial representations as an indicator for selection / organization processes and the number of transitions between text and picture as an indicator of integration processes.

3 Eye(s) see what you do: Eye Movement Modeling Examples

EMME consist of dynamic gaze replays of another person (usually an expert model) that are recorded while that person is performing a learning or a problem-solving task. The model's eye movements, which are overlaid onto the stimulus, are visualized using a dynamic visual cue (e.g., a colored circle; white spotlight) that can change its size depending on the model's fixation duration. The longer the model fixates specific information, the larger the cue. In contrast to static gaze displays (e.g., heat maps), EMME consist of one or more video sequences that show the dynamic change in the gaze behavior of the model as it processes the task material. These EMME videos are then shown to other learners or problem solvers as instructional materials that are supposed to help them perform a task that is either identical or similar to the task accomplished earlier on by the model.

In the following subsections, I will first provide empirical evidence that EMME are an effective instructional tool by presenting and discussing findings from previous research regarding the effectiveness of EMME across various tasks. Against this background, I will present the general theoretical foundations of EMME. Based on the theoretical foundations, I will differentiate between three perspectives that provide different possible mechanisms that might underlie the effectiveness of EMME for multimedia learning: The perceptual learning perspective, the (meta-) cognitive perspective, and the social perspective. Because the focus of the present thesis is on the social perspectives, possible mechanisms and influencing factors that can be derived from a social perspective are presented in more detail.

3.1 Empirical Evidence for the Effectiveness of EMME

The effectiveness of EMME has been investigated in different areas of research. As can be seen in Table 2, EMME have been shown to be effective for fostering visual processing of materials and / or learning outcomes in complex visual search and learning tasks (e.g., medical image diagnosis, fish locomotion patterns), digital reading, and multimedia learning.

In studies of Litchfield et al. (2010), Seppänen and Gegenfurtner (2012), and Gegenfurtner, Lehtinen, Jarodzka, and Säljö (2017), EMME were used to guide the viewers' attentional resources to task-relevant information in the to-be processed materials. For instance, Litchfield et al. (2010) examined whether novice and expert radiographers benefit from knowing another radiographer's search behavior for their diagnostic performance. By comparing the diagnostic performance of novice and experienced radiographers, they found in a first experiment no influence of the model's or the learner's expertise on task performance. In another experiment, they found that especially the diagnostic performance of novice radiographers improved when they were shown the expert's search behavior. Seppänen and Gegenfurtner (2012) revealed that EMME visualizing an expert's strategy for interpreting a computer tomography fostered not only medical students' diagnostic performance (e.g., increase in accurate diagnoses), but also their use of effective processing strategies (e.g., more fixations on task-relevant areas). More recently, Gegenfurtner et al. (2017) investigated whether using EMME could also foster adaptive expertise (i.e., the process in which experts adapt and transfer their skills to new task affordances) in medical image diagnosis. They demonstrated that EMME not only had positive effects on task performance, but also improved visual processing as well as the use of cognitive and metacognitive comprehension strategies during the task performance.

On the one hand, these aforementioned studies show that EMME fosters task performance in visual search tasks. On the other hand, there are also empirical indications that EMME not only foster the performance in the actual task, but also learning by supporting learners in developing beneficial processing and comprehension strategies for novel tasks when attention guidance is no longer present (Jarodzka et al. 2012; Jarodzka et al., 2013). In a study by Jarodzka et al. (2012), medical students received recorded eye movements superimposed onto case videos of patients. They were asked to search for symptoms and to interpret their observations. Results revealed that EMME not only improved students' visual search but also their clinical reasoning for later observations. Moreover, there is also empirical evidence outside of the field of medical imaging that EMME supports people in perceptual learning tasks. Jarodzka et al. (2013) investigated EMME as an intervention for a perceptual task with dynamic stimuli. In this study, students either received EMME demonstrating how to classify fish locomotion and additional verbal explanations or only verbal explanations. Results revealed that EMME as a pre-training intervention during a training session not only improved learners' visual search, but also supported the learners in interpreting relevant information and fostered their performance regarding novel stimuli.

In addition to perceptual search and learning tasks, EMME have proven to be effective for a number of other learning tasks in multiple fields. In a recent study, for example, Salmerón and Llorens (2018) investigated the effectiveness of EMME for digital reading in a contrasting case scenario with ninth graders. The students either evaluated pairs of EMME that displayed digital reading strategies on a hypertext or received written case examples that described other students' reading strategies. Students in the EMME condition not only spend more time reading the digital document, but also showed higher comprehension performance in a post-test one week after the instruction than students in the control condition.

Taken together, these studies show that for visual search tasks (e.g., Litchfield et al., 2010), perceptual learning tasks (e.g., Jarodzka et al., 2013) as well as for digital reading tasks (Salmerón & Llorens, 2018), EMME can be beneficial. Contrary to the above-mentioned findings, for problem-solving tasks the results for EMME were less promising (see Table 2). Although EMME led to higher investment of mental effort (Van Gog et al., 2009), there were usually no beneficial effects for learners' performance in procedural problem-solving tasks (Van Gog et al., 2009; Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016). For example, Van Gog et al. (2009) investigated whether EMME demonstrating problem-solving processes combined with a verbal description of the thought process would support learners in performing the problem-solving process on their own. Their results indicated that not only did learners not benefit from EMME in combination with verbal explanations, they actually performed even worse. More recently, Van Marlen et al. (2016) examined whether EMME are effective for learning how to solve geometry problems. They found that EMME were not more effective compared to regular video examples. There were, however, also no detrimental effects compared to learning without EMME.

Two exceptions to these findings for problem-solving tasks constitute a study by Litchfield and Ball (2011) in the medical context and a study by Van Marlen, Van Wermeskerken, Jarodzka, and Van Gog (2018) in the field of mathematics. Litchfield and Ball (2011) demonstrated that students receiving EMME visualizing the solution procedure for Duncker's radiation problem showed more effective processing strategies, faster solution times, and higher solution accuracy than students without EMME. More recently, Van Marlen et al. (2018) demonstrated that EMME can be useful for conveying knowledge on how to solve geometry problems. However, they identified learners' alleged prior knowledge as a factor influencing the effectiveness of EMME. In their study, only learners with alleged less prior knowledge (i.e., secondary students) benefitted from EMME, whereas learners with alleged more knowledge (i.e., university students) did not.

Contrary to the mixed findings for problem-solving tasks, it has been demonstrated consistently that EMME are an effective instructional tool for multimedia learning (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017; see Table 2). As the present dissertation focuses on EMME within the context of multimedia learning, the related empirical findings are described in more detail in the following.

In a first study, Mason, Pluchino et al. (2015) investigated the effectiveness of EMME for learning from multimedia material in a school context. Forty-two seventh graders studied a science learning material containing text and picture information about the food chain either with EMME support or without EMME support before the learning phase. Results of this study showed that the learners benefited from EMME support, indicated by more integrative processing (i.e., gaze switches between text and pictures) as well as better learning performance regarding verbal recall, graphical recall, and transfer performance.

In a follow-up study by Mason et al. (2016), 64 seventh graders were asked to observe eye movements of a model studying an illustrated text. Afterwards, they studied a different illustrated text on their own. Again, results indicated that students with EMME showed more integrative processing behavior (i.e., gaze switches between text and pictures) and had higher learning outcomes as compared to students without EMME. Moreover, the results indicated that EMME were especially beneficial for students with lower reading comprehension skills.

In line with the research by Mason, Pluchino et al. (2015) and Mason et al. (2016), Scheiter et al. (2017) further revealed that also adult learners can profit from EMME. They investigated whether EMME illustrating a complex processing strategy would foster multimedia learning. Fifty-three students studied multimedia learning materials containing text and picture information on cell division either with EMME support or without EMME support before the learning phase. Scheiter et al. (2017) demonstrated that learners benefited from EMME, but only as long as they possessed higher domain knowledge. Learners with less domain knowledge, however, were even hampered in their learning.

Expanding the field of research regarding EMME as instructional tool for multimedia learning, Mason et al. (2017) investigated in a study with 84 seventh graders whether the temporal sequence of demonstrating effective multimedia processing strategies influences the effectiveness of EMME. Before the learning phase, students in the three EMME conditions received EMME that showed a didactically behaving model demonstrating integrative reading strategies. The order in which pictorial and text information was processed, however, differed between the EMME conditions. In the text-first EMME condition, the model first read the whole text information before shifting between text and picture information to demonstrate integrative reading behavior. In contrast, in the picture-first EMME condition the model looked at the picture first before shifting between text and picture information. In the picture-last EMME condition, the model started reading the text first, then shifted between the text and the picture, and fixated the picture again at the end. Students in a control condition received no EMME. The results showed that the temporal sequence of the processing strategies demonstrated within the EMME was important, as only EMME demonstrating a picture-first strategy supported learners in their picture processing and in their integrative processing of text and picture information. Moreover, the results showed that EMME were beneficial for learning:

Students with picture-first EMME showed higher performance in the factual knowledge test and in the transfer knowledge test. In the recall test, all EMME conditions showed higher performance than the control group. Furthermore, these results also highlight the importance of integrative processing as a cognitive processing strategy, as the frequency as well as the duration of integrative processing mediated the effect of EMME on learning.

Taken together, these results indicate that EMME can foster learning with multimedia. It also indicates that EMME are not only beneficial for learning about specific tasks, but also for learning or triggering more general processing strategies that can be applied to different multimedia materials. Moreover, there are first indications for an influence of learner characteristics on the effectiveness of EMME such as reading comprehension skills (Mason et al., 2016) or domain knowledge (Scheiter et al., 2017).

Overall, the empirical evidence that was presented in this section indicates that EMME can be an effective instructional tool. However, in order to identify factors that can influence the effectiveness of EMME, it is important to look more closely at their theoretical foundations.

Table 2. Main Characteristics of the Previous Studies Using EMME to Foster Task Performance and Learning

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Litchfield, Ball, Donovan, Manning, and Crawford (2010)	**Medical image diagnosis:** Radiographers (Novices vs. Experts)	**Experiment 1:** Mixed design: participant expertise (novice vs. experienced), model expertise (novice vs. expert), and viewing condition (free search; image preview; EMME)	EMME visualized examination strategies of an expert / novice model on a x-ray image.	• **Solution times:** EMME > image preview > free search (longer decision times, $\eta^2_p = 0.61$) • **Diagnostic performance:** EMME > free search = image preview ($\eta^2_p = 0.41$)
		Experiment 2: 2 x 3 between-subjects design: participant expertise (novice vs. experienced) and viewing condition (image preview; expert EMME; unrelated EMME)	EMME visualized examination strategies of an expert model vs. visualizing eye-movement patterns related to a different visual task.	• **Diagnostic performance:** expert EMME > unrelated EMME = image preview depending on the level of expertise (novices: $\eta^2_p = 0.48$)
	Medical image diagnosis: Novice radiographers	**Experiment 3:** 4 x 1 between-subjects design: naïve-no-task vs. naïve-search vs. expert-search vs. incongruent-search	EMME visualized examination strategies of an expert model vs. examination strategies of a non-radiographer.	• **Diagnostic performance:** naïve-search = expert-search > naïve-no-task, incongruent-search ($\eta^2_p = 0.42$)
Jarodzka, Balslev, Holmqvist, Nyström, Scheiter, Gerjets, and Eika (2012)	**Medical image diagnosis:** Medical students	3 x 1 between-subjects design: control vs. spotlight EMME vs. circle EMME condition	EMME visualized search strategies of an expert model. Crucial features were highlighted with a circle (circle condition) vs. Features the model did not focus on were blurred (spotlight condition).	• **Attention guidance:** spotlight EMME > control ($\eta^2_p = 0.13$) • **Information processing:** spotlight EMME < control for time to relevant information (faster fixation to information, $\eta^2_p = 0.13$), EMME > control (longer fixation on relevant information, $\eta^2_p = 0.11$) • **Diagnostic accuracy:** spotlight EMME > control ($\eta^2_p = 0.11$)

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Seppänen, and Gegenfurtner (2012)	**Medical image diagnosis:** University students	2 x 1 between-subjects design: control vs. EMME condition	EMME visualized examination strategies of an expert model on a computed a CT scan.	• **Information processing:** retention > baseline for number of fixations on task relevant areas (EMME condition), retention < baseline for number of fixations on task-redundant areas (EMME condition) • **Diagnostic performance:** retention > baseline for accuracy and sensitivity (EMME condition)
Gegenfurtner, Lehtinen, Jarodzka, and Säljö (2017)	**Medical image diagnosis:** medical experts and medical students	Mixed design: expertise (PET experts vs. CT experts vs. novices) as between and time of case presentation (baseline vs. retention vs. transfer) as within factor	EMME visualized examination strategies of an expert model. Concurrent think-aloud protocols and screen actions (key strokes; mouse clicks) were provided.	• **Information processing:** transfer > baseline (number of fixations on relevant areas: for experts: *Cohens's d* = 1.23; for novices: *Cohens's d* = 0.73), retention > baseline (number of fixations on relevant areas: for experts: *Cohens's d* = 1.67; for novices: *Cohens's d* = 0.70); transfer = retention > baseline (fixation time on task-relevant areas) • **Solution accuracy** (think-aloud): correct solutions: transfer > baseline (for experts: *Cohens's d* = 0.48), retention > baseline (for experts: *Cohens's d* = 0.45); incorrect solutions: transfer < baseline (for novices: *Cohens's d* = 0.26), retention < baseline (for novices: *Cohens's d* = 0.35) • **Diagnostic accuracy:** retention > baseline (for experts: *Cohens's d* = 0.55; for novices: *Cohens's d* = 1.94)

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Jarodzka, Van Gog, Dorr, Scheiter, and Gerjets (2013)	**Perceptual task:** university students	3 x 1 between-subjects design: control vs. spotlight EMME vs. dot EMME condition	EMME visualized examination strategies of a didactically behaving expert model. Concurrent verbal explanations were provided. Crucial features were highlighted with a dot (dot condition) vs. were highlighted with spotlight (spotlight condition).	• **Attention guidance:** EMME > control ($\eta^2_p = 0.39$) • **Information processing:** EMME < control for time to relevant information (faster fixation to information, $r = 0.25$), spotlight EMME < dot EMME for time to relevant information (faster fixation to information, $r = 0.24$), EMME > control for fixation time ($r = 0.31$) • **Interpretation ability:** EMME > control ($r = 0.21$), dot EMME > spotlight EMME ($r = 0.27$) • **Mental effort:** spotlight EMME > dot EMME ($r = 0.26$)
Salmerón, and Llorens (2018)	**Digital reading:** 9th graders	Mixed design: intervention (EMME vs control) as between subjects variable and time of testing (pre-posttest) as within subjects variable	EMME visualized reading strategies of a model that tried to answer a question.	• **Information processing:** EMME > control for reading time; EMME = control for page visits; time to relevant and irrelevant pages • **Learning outcomes:** EMME > control depending on test-time for comprehension performance (post-test one week later)
Litchfield, and Ball (2011)	**Problem solving:** university students	3 x 1 between-subjects design: tumor fixation vs. natural EMME model vs. didactic EMME model condition	EMME visualized the solution strategy for Duncker's radiation problem of a didactically behaving model vs. the solution strategy of a peer model.	• **Information processing:** EMME > tumor fixation for number of saccades ($\eta^2_p = 0.79$) • **Solution accuracy:** EMME > tumor fixation • **Solution times:** didactic EMME > tumor fixation (faster solution times in the first 5 minutes); didactic EMME > natural EMME (faster overall solution times, $\eta^2_p = 0.19$)

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Van Gog, Jarodzka, Scheiter, Gerjets, and Paas (2009)	**Problem solving:** university students	2 x 2 between-subjects design: Example Type (Product vs. Process) vs. Attention Guidance (EMME vs. no EMME)	EMME visualized the solution strategy of an expert model for the Frog Leap Task. The visualized solution strategy varied between conditions.	• **Solution accuracy**: EMME < no-EMME depending on example type (process-oriented examples) • **Mental effort:** EMME > no-EMME depending on example type (process-oriented examples; $\eta^2_p = 0.08$)
Van Marlen, Van Wermeskerken, Jarodzka, and Van Gog (2016)	**Problem solving:** university students	**Experiment 1:** 3 x 1 between-subjects design: control vs. meaningful EMME vs. meaningless EMME condition.	EMME visualized eye movements of an expert model during the solution process of simple geometry problems. Model displayed an effective solution strategy vs. focused on all features of the problem in a meaningless order.	• **Solution accuracy:** EMME = control • **Solution times:** meaningful EMME < control for transfer test problems (less time, $r = 0.42$)
		Experiment 2: 2 x 1 between-subjects design: control vs. EMME condition	EMME visualized the solution strategy of an expert model for solving complex geometry problems. Concurrent verbal explanations with regard to the different solution steps were provided.	• **Information processing:** EMME < control for time to relevant information (faster fixation on information, $r = 0.27$), EMME > control for fixation time (longer fixation, $r = 0.31$) • **Solution accuracy:** EMME = control • **Solution times:** EMME > control for transfer test problems (more time, $r = -0.35$)

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Van Marlen, Van Wermeskerken, Jarodzka, and Van Gog (2018)	**Problem solving:** university students	**Experiment 1:** 2 x 1 between-subjects design: modeling examples (ME) vs. EMME condition	EMME visualized the solution strategy of an expert model for solving complex geometry problems. Concurrent ambiguous verbal explanations with regard to the different solution steps were provided.	• **Information processing:** EMME < ME for time to relevant information (faster fixation on information, $r = 0.49$), EMME > ME for number of fixations (more fixations on relevant information, $r = 0.52$), EMME = ME for fixation time • **Solution accuracy:** EMME = ME for isomorphic and transfer problems • **Solution times:** EMME > ME for isomorphic problems (more time, $r = 0.32$), EMME = Me for transfer problems
	Problem solving: secondary education students	**Experiment 2:** 2 x 2 between-subjects design: Modeling type (modeling examples (ME) vs. EMME condition) vs. Ambiguity (unambiguous vs. ambiguous)	EMME visualized the solution strategy of an expert model for solving complex geometry problems. Concurrent ambiguous vs. unambiguous verbal explanations with regard to the different solution steps were provided.	• **Solution accuracy:** EMME > ME for isomorphic ($\eta^2_p = 0.13$) and transfer problems ($\eta^2_p = 0.05$)
Mason, Pluchino, and Tornatora (2015)	**Multimedia learning:** 7th graders	2 x 1 between-subjects design: control vs. EMME condition	EMME visualized processing strategies of a didactically behaving model for processing a one-page illustrated text on the water cycle.	• **Information processing:** EMME > no-EMME for integrative processing ($\eta^2 = 0.09$ to $\eta^2 = .31$) • **Learning outcomes:** EMME > control for verbal recall ($\eta^2 = 0.09$), transfer ($\eta^2 = 0.10$), graphical recall ($r = 0.42$)
Mason, Pluchino, and Tornatora (2016)	**Multimedia learning:** 7th graders	2 x 1 between-subjects design: control vs. EMME condition	EMME visualized processing strategies of a didactically behaving model for processing a one-page illustrated text on the water cycle.	• **Information processing:** EMME > no-EMME for integrative processing ($\eta^2_p = 0.11$ to $\eta^2 = 0.22$) • **Learning outcomes:** EMME > no-EMME for transfer knowledge ($\eta^2 = 0.15$)

Authors	Area of Research / Participants	Design	EMME support	EMME effects
Scheiter, Schubert, and Schüler (2017)	**Multimedia learning:** university students	2 x 1 between-subjects design: control vs. EMME condition	EMME visualized effective multimedia processing strategies of a didactically behaving model on four pages of an illustrated text on cell division.	• **Information processing:** EMME > control for integrative processing (*adj.* R^2 = 0.53) • **Learning outcomes:** EMME effective depending on learners' prior knowledge: EMME > control for factual knowledge (stronger students, *adj.* R^2 = 0.20); EMME < control for recall (weaker students, *adj.* R^2 = 0.10)
Mason, Scheiter, and Tornatora (2017)	**Multimedia learning:** 7th graders	4 x 1 between-subjects design: text-first EMME vs. picture-first EMME vs. picture-last EMME vs. no-EMME condition	EMME visualized processing strategies of a didactically behaving model for processing a one-page illustrated text on the water cycle. The sequence of the depicted strategies varied between conditions.	• **Information processing:** picture-first EMME > no-EMME for picture processing (η^2_p = 0.11 to η^2_p = 0.15), integrative processing (η^2_p = 0.10) • **Learning outcomes:** picture-first EMME > no-EMME for factual knowledge (η^2_p = 0.14) and transfer knowledge (η^2_p = 0.09); EMME generally effective for recall (η^2_p = 0.16)

3.2 Theoretical Foundations of EMME for Multimedia Learning

EMME can be regarded as a special case of video-based modeling examples (Jarodzka et al., 2013). Thus, similar to other modeling examples, they are theoretically grounded in the tradition of example-based learning on the one hand and in observational learning on the other hand (Bandura, 1986; Van Gog & Rummel, 2010).

Learning from examples is a very effective way to acquire cognitive skills (Renkl, 2011; Renkl, 2014; Van Gog, Paas, Marcus, Ayres, & Sweller, 2009; Van Gog & Rummel, 2010). Although there are theoretically two different branches of research on the effectiveness of example-based learning, namely the cognitive load theory (e.g. Sweller, 1988, Sweller, Ayres, & Kalyuga, 2011) and the social cognitive learning theory (Bandura, 1986), both underline the importance of learning from others (for a more detailed discussion, see Van Gog & Rummel, 2010).

From a cognitive load theory perspective, providing solution steps in example-based learning helps learners to generate problem-solving schemata that are then stored in long-term memory (Sweller et al., 2011). These problem-solving schemata contain knowledge components regarding the solution-procedure such as operators, problem states, and consequences of the application of certain operators (Renkl, 2011). They can be later used to solve similar problems. With regard to the cognitive load theory, the cognitive schemata reduce the extraneous cognitive load during example-based learning. Consequently, it is assumed that worked examples provide learners with cognitive capacities for learning-relevant (i.e., germane) load (Paas & Van Gog, 2006; Renkl, 2011; Sweller, van Merriënboer, & Paas, 1998).

Similar to worked examples, modeling examples provide learners with a modeled solution procedure. Modeling examples are theoretically based on observational learning (social cognitive learning theory; Bandura, 1984). Van Gog and Rummel (2010) argue that worked examples can also be seen as a kind of observational learning against the background that the exchange of individual knowledge with others requires the externalization of cognitive activities (e.g. in written form). In contrast to worked examples, however, modeling examples provide learners with the possibility to observe a human (-like) model applying the necessary skills / solution steps (Van Gog & Rummel, 2010; Van Gog, Verveer, & Verveer, 2014), and thus entail an additional social component.

That modeling examples are effective for acquiring cognitive skills has been demonstrated empirically for a variety of tasks. They are not only effective for teaching highly structured cognitive skills such as problem solving (Hoogerheide, Van Wermeskerken, Loyens, &

Van Gog, 2016) or medical / surgical skills (Bjerrum et al., 2013; LeBel, Haverstock, Cristancho, van Eimeren, & Buckingham, 2018), but also for teaching less structured skills such as collaboration skills (Rummel & Spada, 2005; Rummel, Spada, & Hauser, 2009), creative skills (Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b) or writing skills (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004; Couzijn 1999; Zimmerman & Kitsantas, 2002).

According to Bandura (1986), learning of cognitive skills results to a large degree from observing other persons who serve as models for the to-be-learnt behavior. Moreover, learning from models is highly effective, even in the absence of rewards (Bandura, 1986). In contrast to observable skills, however, observational learning of cognitive skills requires that the model externalizes his / her cognitive processes during the process to make them observable for the learners (Collins, Brown, & Newman, 1989). Based on these assumptions, modeling examples provide learners with the possibility to observe other persons who serve as models for the to-be-learnt behavior applying the necessary skills / solution steps.

Bandura (1986) describes observational learning as an information-processing activity. By observing models applying their skill during a task, learners can expand their knowledge as well as their skills without costly errors. Moreover, modeling examples can support learners in acquiring cognitive representations of the model's behavior that outlast the current situation and enable them to demonstrate the observed behavior on later occasions (Bandura, 1986; Van Gog & Rummel, 2010). Another important aspect of modeling examples is that modeling not only fosters learning of specific behaviors. Through the process of abstract modeling, learners can also obtain judgmental skills and generalizable rules for constructing behavior as well as rules for classifying events (Bandura, 1986). This is an important assertion because it means that learners are able to generalize observed rules to other events and cases and thus can show observed behavior in new situations when they see fit. Moreover, it means that learners are not only able to reproduce observable behavior, but also are able to learn more abstract concepts such as cognitive skills by observing other people. By differentiating attention-directing functions from behavior-cuing functions of modeling, Bandura (1986) postulates that the modeled behavior not only can convey new knowledge, but also can serve as an attention-directing cue and channel the learners' attention to certain objects or stimuli. With regard to the behavior-cuing functions of modeling, Bandura presumed that "modeling influences can alter how people organize their thoughts, what type of information they seek, and how they process it" (1986, p. 103). Thereby, modeling imparts component skills and provides learners with rules on how to translate these components into new behavioral structures (Bandura, 1986).

According to Bandura (1986), observational learning is controlled by four sub-processes: attentional processes, retention processes, production processes and motivational processes (see Figure 3). Since EMME are a special kind of modeling example, it is assumed that these processes as well as some of the associated factors also apply to EMME. Therefore, the most relevant processes and factors with regard to EMME are discussed in more detail.

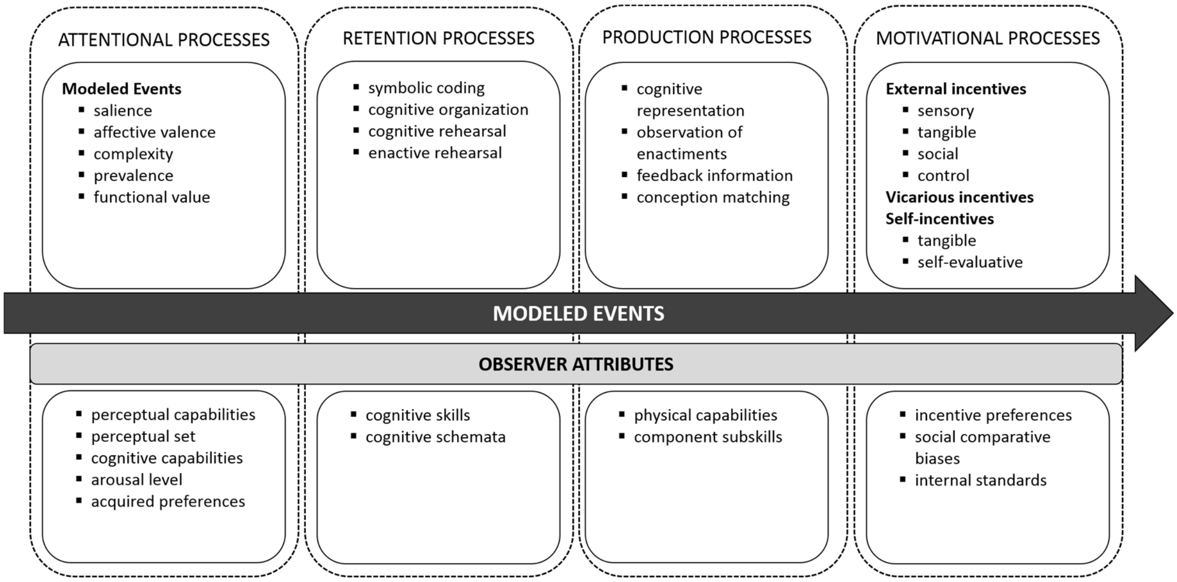


Figure 3: Sub-processes governing observational learning (adapted from Bandura, 1986, p. 52)

Attentional processes are defined as one of the most important sub-functions in observational learning (Bandura 1986). For observational learning to take place, it is crucial that learners pay attention to the relevant aspects of the modeled behavior. According to Bandura, there are factors that influence the learners' perception and selection processes (see Figure 3). On the one hand, they are influenced by the characteristics of the modeled event (e.g., salience of relevant aspects, complexity of the modeled behavior) and, on the other hand, by observer attributes (e.g., learners' cognitive capabilities and perceptual abilities). Moreover, not only cognitive factors, such as observer attributes (i.e., learner characteristics) or characteristics of the modeled event play a role in directing learners' attention, but also social factors, such as the (alleged) relation between model characteristics and learner characteristics (i.e., model observer similarity, Bandura, 1986).

Applied to EMME, this would mean that learners must have certain cognitive and perceptual prerequisites. For example, the learners need to be perceptually able to follow the model's gaze. Moreover, they need to have the cognitive abilities to interpret the eye movements as processing strategies. With regard to the influence of social factors, it is assumed that in the case that modeled behavior is sufficiently salient to draw attention by itself, social factors would only play a minor role (Bandura, 1986). However, in the case that the modeled behavior is not sufficiently salient, or learners are not sufficiently motivated, (anticipated) benefits of the modeled behavior provide incentives for paying closer attention to the model. Moreover, also social factors related to the model itself can have an influence on learners' attentional processes. According to Bandura (1986) and Schunk (1987), learners tend to pay more attention to models that are considered effective and have achieved good results with their behavior (Bandura, 1986; Schunk, 1987). Hence, it is possible that for EMME as an instructional tool, learners might particularly benefit from models that they consider or perceive as 'competent' and 'successful', as the use of these types of models motivates them to pay more attention to the modeled behavior.

The second sub-function in Banduras' model refers to retention processes. As stated by Bandura (1986), observational learning can take place only if the observers not only pay attention to the modeled behavior, but also retain it in their memory. Therefore, the information needs to be encoded. According to Bandura, the encoding of the information is a constructive process. Learners have to actively transform and restructure the information from their observation into representational models to capture the essential features and structures of the modeled event. The representational models for the modeled behavior are then coded into images and verbal symbols in the form of conceptions, rules, and propositions and integrated with the existing knowledge of the learner. As learners with higher cognitive abilities and / or already existing cognitive schemata (i.e., more prior knowledge) are assumed to be better able to interpret and retain the modeled behavior, it is generally assumed that they benefit more from modeling examples than learners with lower cognitive abilities. With regard to EMME, this would mean that learners' prior knowledge has to be taken into account when using them as instructional tool.

As a third sub-function, production processes are mentioned in Bandura's model. Production processes are (cognitive) processes that convert the representational models of the modeled behavior into an appropriate behavior by organizing behavioral responses that are spatially and temporally in accordance with the representational model. Behavioral enactment of the modeled behavior contains the following cognitive processes: organization processes (i.e., cognitive organization of response patterns), initiation processes, monitoring processes (i.e., monitoring of response enactments), and matching actions (i.e., corrective performance adjustments). Whether learners can execute the modeled behavior themselves depends, on the one hand, on the accuracy of their acquired internal representational model and, on the other hand, on the existence of the required subskills (Bandura, 1986). In the case of EMME, this would mean that for reproducing the modeled strategies themselves, learners first have to have a correct internal representational model of the modeled strategies. Then they have to implement the right strategies themselves and, at the same time, cognitively monitor their application of these strategies during the learning process, and if necessary, correct their strategy use based on their monitoring. However, as learners do not receive direct feedback on the application of the strategies, EMME has the deficit that it is difficult for learners to monitor the use of the strategies and make timely corrections. Thus, this might demand high cognitive capabilities from the learners during the learning phase.

With regard to the retention and the production processes, it becomes clear that for observational learning, learner characteristics such as cognitive capabilities play an important role. Moreover, there is empirical evidence that especially learners' prior knowledge has a major role in learning from modeling examples (Bandura, 1986; Van Gog & Rummel, 2010). Thus, considering learner characteristics for the effectiveness of modeling examples is important. With regard to EMME, there is already empirical evidence that learner characteristics such as prior knowledge (Scheiter et al., 2017) as well as reading competence (Mason et al., 2016) can influence the effectiveness of EMME.

The fourth sub-function in Bandura's model refers to motivational processes. This sub-process is important in relation to the fact that even if learners could perform the behavior themselves, this does not necessarily mean that they will actually do it. The potential discrepancies between learning and executing a modeled behavior can be explained by motivational processes (Bandura, 1986). According to Bandura, the enactment of the modeled behavior is influenced by observed and self-evaluative incentives as well as internal standards. One factor that is mentioned in the model are learners' social comparative biases (Bandura, 1986). Bandura (1977) argues that people often compare themselves with others in similar situations or with their previous accomplishments to define performance standards. In turn, these standards form the basis for further self-regulative actions (Bandura, 1977). Also, Hoogerheide, Van Wermeskerken, Van Nassau, and Van Gog (2018) claim that modeling encourages learners to engage in a social comparison with the model. Because of these social comparison processes, Hoogerheide et al (2018) presume that positive effects of modeling examples are also partly based on (alleged) model characteristics. Against this background, it is possible that learner characteristics such as social comparison orientation (Gibbons & Buunk, 1999) as a social factor might influence the effectiveness of modeling examples, and in turn also the effectiveness of EMME.

In summary, according to Bandura (1986), learners differ in their receptivity to modeling cues and therefore not all models are equally effective in generating desired behavioral results. As Bandura (1986) stated, there are three kinds of factors that influence learners' responsiveness to modeling: the characteristics of the learner (e.g., cognitive capabilities, cognitive schemata or social comparative biases), the characteristics of the model (e.g., model competence), and the characteristics of the modeled event (e.g., salience of relevant aspects). In line with this presumption, previous research indicated that learners' cognitive abilities such as prior (domain) knowledge, the perceived competence of a model, or the similarity between model and observer can affect observational learning and, in turn, the effectiveness of modeling examples (Schunk, 1987; Van Gog & Rummel, 2010).

3.3 Potential Mechanisms that underlie the Effectiveness of EMME for Multimedia Learning

Based on the theoretical considerations described above, example-based learning and observational learning offer a theoretical framework that can explain why, overall, EMME should be effective as an instructional tool. However, in order to clarify which mechanisms possibly play a role in the effectiveness of EMME for multimedia learning, it is necessary to first separate and examine different perspectives on the effectiveness of EMME. In the following subsections, I will present three perspectives on possible underlying mechanisms for the effectiveness of EMME: The perceptual learning perspective, the (meta-) cognitive perspective, and the social perspective. While the perceptual learning perspective focuses on the attention-guiding function of EMME, the (meta-) cognitive perspective focuses on generating / improving effective cognitive processing strategies, and the social perspective focuses on social aspects of EMME. As noted above, these perspectives are not mutually exclusive, but rather complementary. Although all perspectives are important for the present thesis, the main focus of this work lies on the social perspective, which has been rather neglected in previous research so far. Hence, special attention will be directed to the social side of EMME.

3.3.1 The perceptual learning perspective on EMME: Attentional guidance

EMME support perceptual learning by guiding visual attention (Goldstone et al., 2017). That is, during observing the model's gaze, the learner's visual attention is synchronized with that of the model (Jarodzka et al., 2013; Van Marlen et al., 2016). This is assumed to allow observers to infer which information is relevant to the task and highlights the order in which that information should be processed (Mason, Pluchino et al., 2015; Mason et al., 2017). For instance, Mason et al. (2017) presented EMME that illustrated different sequences of processing information from text and pictures, which in turn affected students' visual attention and learning in a subsequently performed task. Accordingly, EMME seem to teach observers the specific eye movement patterns relevant to successfully performing a task at hand (Hayhoe & Ballard, 2005).

Guiding learners' visual attention by highlighting learning-relevant information to attract learners' attention is usually referred to as signaling principle, cuing principle, or attention-guiding principle (Schneider, Beege, Nebel, & Ray, 2018; Richter, Scheiter, & Eitel, 2016; Van Gog, 2014). Since gaze cues can be regarded as a special form of attention-guiding cue (Litchfield et al., 2010; Gallagher-Mitchell, Simms, & Litchfield, 2018), it can be assumed that they follow the same mechanisms. In contrast to other attention-guiding cues that are especially designed by a task or instructional expert, however, gaze cues represent the actual eye movements of a human model. Hence, they might include an additional social dimension that is missing in other forms of attention-guiding cues (e.g., color coding, Jamet, 2014).

That attention-guiding cues per se are beneficial for multimedia learning has been shown in empirical studies where material-oriented attention-guiding cues have been used such as highlighting specific areas, or using color-changing cues (for reviews, see Richter, et al., 2016; Schneider, et al., 2018). For instance, research on signaling as material-oriented intervention indicates that using techniques such as signaling relevant information supports learners' visual as well as cognitive processing (e.g., de Koning, Tabbers, Rikers, & Paas, 2010; Jamet, 2014; Grant & Spivey, 2003; Ozcelik, Arslan-Ari, & Cagiltay, 2010).

Grant and Spivey (2003), for example, demonstrated that directing learners' attention to crucial areas can improve reasoning in a problem-solving task that is based on a diagram. Learners who received an animated diagram that drew their attention to the critical feature in the diagram performed significantly better at the problem-solving task than learners who received an animated diagram that drew their attention to a non-critical area or learners without instructional support. For multimedia material, Ozcelik et al. (2010) investigated in a study with university students the effects of signaling on eye movements and learning outcomes in the context of multimedia learning. The students either received signaled or nonsignaled multimedia material with a labeled illustration of a turbofan jet engine and a narrative instruction. For signaling, the authors used temporarily color-changing labels. The results revealed that signaling was beneficial for guiding the attention to relevant information and for improving the efficiency and effectiveness of finding this information. Furthermore, students with signaled materials performed better at the transfer test.

Likewise, Jamet (2014) demonstrated that signaling was beneficial for guiding learners' attention and for learning outcomes. Jamet investigated the influence of using color-changing items on learning with diagrams and spoken explanations. Results showed that signaling reduced the fixation time on less relevant areas. Furthermore, learners with signaled material outperformed learners without visual cues. In contrast to Ozcelik et al. (2010), however, there were no beneficial effects on learners' deep comprehension. Finally, Scheiter and Eitel (2015) demonstrated that using visual cues to guide learners' attention during the learning process fostered learning performance. They investigated the influence of visual cues on learning from multimedia material. Results showed that learners did not only change their processing of the learning material but also that this change could explain the better learning outcome for learners in the cuing condition.

According to Schneider et al. (2018), the effectiveness of attention-guiding cues for multimedia learning is theoretically based on assumptions from the CTML (Mayer, 2014a) and from the cognitive load theory (CLT; e.g., Paas & Sweller, 2014). Both theories assume that learners' overall information processing capacity is limited and that supporting learners in using effective cognitive processes facilitates multimedia learning. From a CTML perspective, attention-guiding cues support learners to actively select, organize, and integrate learning-relevant information into long-term memory (Schneider et al., 2018). From a cognitive load perspective, attention-guiding cues reduce extraneous cognitive load (ECL) that emerges from the design of the learning material and increases germane cognitive load (GCL) that supports learning-relevant processes (Amadieu, Mariné, & Laimay, 2011; Schneider et al., 2018). Thus, by drawing attention to the relevant information, highlighting important elements or the organization of the material helps learners to distinguish between relevant and irrelevant information, and facilitate goal-oriented learning (Schneider et al., 2018; Van Gog, 2014).

With regard to potential factors which can be derived from a perceptual learning perspective and that might influence the effectiveness of EMME, previous studies have, for example, identified prior knowledge / expertise of learners as to be of importance. For instance, Richter et al. (2016) conducted a meta-analysis focusing on the overall effectiveness of signaling text-picture relations in multimedia learning. They noted that especially learners with low prior knowledge benefit from signaled materials. This finding might be related to the fact that prior knowledge / expertise influences people's attention allocation. Previous eye tracking studies indicated that experts differ from novices in their visual information processing (e.g., Gegenfurtner, Lehtinen, & Säljö, 2011; Van Gog, Paas, & Van Merriënboer, 2005; Van Meeuwen, Jarodzka, Brand-Gruwel, Kirschner, De Bock, & Van Merriënboer, 2014). Gegenfurtner et al. (2011) found, for example, as a result of a meta-analysis that experts are faster in shifting their attention to relevant information, look at relevant information longer and tend to look less at irrelevant information. Thus, learners with less prior knowledge / expertise might benefit especially from instructional support such as signaled learning material that helps them to allocate their attention to relevant information.

However, even though signaled learning material and EMME may rely on the same attention-guiding mechanisms, EMME are special in the regard that the attention-guiding cue is based on human action. Hence, it is assumed that eye movements as a means of attentional guidance play a privileged role compared to other means of enhancing perceptual learning (e.g., cueing or perceptual grouping, cf. Goldstone et al., 2017).

In the chapter "The social perspective on EMME", I will further outline possible consequences of linking attentional guidance with visual (social) cues such as eye movements on the effectiveness of EMME. First, however, I will present (meta-) cognitive perspective on the effectiveness of EMME.

3.3.2 The (meta-) cognitive perspective on EMME: Conveying strategy knowledge

From a (meta-) cognitive perspective, eye movements are not only informative with regard to visual attention. Because of the close link between visual attention and cognitive processes (cf. eye-mind assumption, Just & Carpenter, 1980), EMME are also assumed to foster higher-level cognitive processes (Scheiter et al., 2017). For instance, it has been shown that by using another person's gaze as informational source, people can improve their collaborations as well as cognitive performance (Schneider & Pea, 2013). Furthermore, there is evidence that people are able to use gaze information to regulate their own cognitive processes (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Litchfield et al., 2010). On this basis, it is assumed that other mechanisms underlying the effectiveness of EMME for multimedia learning relate to the way EMME supports learners in acquiring effective processing strategies on the one hand (Mason et al. 2015, 2016) and in regulating existing strategies on the other (Mason, et al., 2017; Scheiter et al., 2017).

With respect to the theoretical foundations described in Chapter 3.2 'Theoretical Foundations of EMME', EMME can be regarded as a particular type of worked example. By progressively modeling the application of effective multimedia processing strategies to the learning material, EMME are assumed to support learners in generating a cognitive representation (c.f. cognitive schema) of the use of effective multimedia processing strategies. From a cognitive load theory perspective, an effective multimedia processing schema would be beneficial, because learners' extraneous cognitive load is reduced freeing more cognitive capacities for processing the actual content (e.g., Sweller et al., 2011). The assumption that learners generate a multimedia processing schema based on the modeled strategies in the EMME-videos is important, because this premise forms the basis for the assumption that the effectiveness of EMME can be explained not only at the perceptual level (i.e., imitation) but also at the cognitive level (i.e., acquiring cognitive schemata).

Conveying strategy knowledge via instructional support is based on the idea that cognitive strategies are mental routines that are at least partly under conscious control (Dole, Nokes, & Drits, 2009). Material-oriented interventions target these strategies by focusing on adjusting the learning materials in a way that certain cognitive processes are strongly supported (e.g., reducing complexity; cueing/signaling; physically integrating text and pictures).

Learner-oriented interventions, on the other hand, focus on the learners themselves and how they can be enabled to draw most benefit from multimedia materials (Renkl & Scheiter, 2015). Moreover, one can distinguish between different kinds of learner-oriented interventions. One of them is for example to support learners *during* the learning process through strategy prompts (Kombartzky, Ploetzner, Schlag, & Metz, 2010; Schlag & Ploetzner, 2011). Another possibility of learner-oriented interventions is to support learners *before* the learning process by providing them with strategy trainings. In contrast to strategy prompts, strategy trainings not only provide learners with the declarative knowledge of what strategies are, but also with procedural knowledge of how to apply these strategies, and with conditional knowledge of when and why the respective strategy is to be applied during a (learning) task (Dole et al., 2009; Veenman, 2011). Veenman (2011) summarized this procedure under the WWW&H rule. According to this rule, learners should be instructed and modeled when to apply what skill, why, and how in the context of a task (Veenman, 2011). Against this background, EMME can be regarded as a strategy-training intervention that provides learners with strategy knowledge before the learning phase on how to effectively process multimedia material.

As previous research on different kinds of strategy-support indicated, giving learners instructional support by providing them with the necessary strategy knowledge is beneficial for multimedia learning (e.g., Eitel et al., 2013; Hyönä, Lorch, & Kaakinen, 2002; Kombartzky et al. 2010; Larson et al., 1986; Leopold, Doerner, Leutner, & Dutke, 2015; Lorch, Lemarié, & Chen, 2013; Mason, Pluchino, et al., 2015, 2016; Mason, et al., 2017; Sanchez, Lorch, & Lorch, 2001; Scheiter et al., 2017; Schlag & Ploetzner, 2011; Stalbovs, Scheiter, & Gerjets, 2015). First evidence for the effectiveness of strategy trainings for multimedia learning was provided by Larson et al. (1986), who investigated whether a strategy training would support learners in understanding illustrated construction manuals. In their study, learners were first taught learning strategies with which they trained on example material before the start of the actual learning phase. The strategies comprised obtaining an overall idea of the function by reading the text, organizing, memorizing, and sketching the corresponding picture, memorizing the corresponding denominations from the text, and linking function and the description from the text in their mind. Results showed a beneficial effect for the strategy training, indicated by better performance in pictorial test questions.

As described in Chapter 2, for building coherent mental representations of the multimedia material, a sequence of certain cognitive processes is necessary. First, learners have to select information from both text and picture (selection processes), then organize the selected information in separate mental models of text and picture (organization processes) and subsequently relate the information from text and picture to each other (i.e., integration, Mayer, 2014a; coherence formation, Seufert, 2003). Consequently, it is assumed that inadequate cognitive processing results in incoherent mental models and therefore also in poorer learning outcomes (e.g., Scheiter & Eitel, 2015). Against this background, previous research identified various effective strategies for selecting, organizing, and integrating information from multimedia material (Eitel et al., 2013; Hegarty & Just, 1993; Hyönä et al., 2002; Lorch et al., 2013; Mason, Tornatora, & Pluchino, 2015; Mason et al., 2017; Sanchez et al., 2001; Scheiter et al., 2017). For example, there is empirical evidence that for written text, initially *inspecting the text's heading* results in better recall (Lorch et al., 2013; Sanchez et al., 2001) and a better ability to summarize the text content (Hyönä et al., 2002). Lorch et al. (2013) argue that the comprehension of expository texts requires learners to understand the structure of a text. Further, it is assumed that by *inspecting the picture before reading the text* learners gain a more comprehensive representation of the picture's visuospatial components (i.e., pictorial scaffold), which supports them in selecting relevant information from the text and, consequently, in the construction of a coherent mental model (e.g., Eitel, Scheiter, Schüler, 2013). This strategy has been found to be beneficial for picture processing as well as for integrative processing of text and picture information (Mason et al., 2017). Furthermore, there are findings that successful learners often *take a final look at a picture* before continuing (Hegarty & Just, 1993). It is assumed that learners use this strategy to assess whether the information they gleaned from the text matches the information they gained from the picture. Thus, this strategy is also considered to be a good multimedia processing strategy as it fosters the integration process. Relatedly, teaching learners to search for correspondences by *switching between text and picture* has been argued to support the integration process (see Chapter 3.1.; for empirical evidence; cf., Mason, Tornatora et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017).

Previous research indicated that conveying these above-mentioned strategies as a set has beneficial effects on multimedia learning (e.g., Scheiter et al., 2017; Stalbovs et al., 2015). For instance, Stalbovs et al. (2015) investigated whether the use of effective multimedia processing strategies also can be supported by means of implementation intentions (if-then plans). Before learning, the learners received implementation intentions regarding text processing (e.g., inspecting the text's heading), picture processing (e.g., taking an initial glance at a picture before reading the text), text-picture integration (e.g., switching between text and picture during learning) or a combination of these strategies. The results indicated that especially learners who received a combination of all three strategies benefited from the instructional support.

Since conveying these above-mentioned strategies as a set proved to be effective for multimedia learning, the above-described strategies were incorporated into the EMME-videos used in the present thesis. The EMME-model demonstrated them in the following order: For every page, the model first read the title of the page. Then the model inspected the picture on the right-hand side of the page (construction of pictorial scaffold). Thereafter, the model read the text and looked at corresponding elements of the picture (selection of relevant words and picture elements; text organization; picture organization). Concurrently, the model switched between text and corresponding picture (integration of both representations). There is first evidence that this set of multimedia processing strategies is effective in EMME interventions for adult learners, but only as long as learners possessed higher domain knowledge (Scheiter et al., 2017).

Another set of multimedia processing strategies was used by Mason et al. (2017). Here, three EMME-videos were used which differed in the temporal sequence of the demonstrated processing strategies. Results of this study indicated that the temporal sequence of the demonstrated strategies within the EMME was important, as only the strategy set illustrating a picture-first strategy supported adult learners in their picture processing and in in their integrative processing of text and picture information. A less complex set of multimedia processing strategies were used by Mason, Pluchino et al. (2015, 2016) for supporting children in multimedia learning. The model in these EMME-videos initially read the full text, before shifting between text and picture information. Results indicated that for children, this set of demonstrated strategies was effective.

Although the results of the studies described above support the assumption that strategy trainings can be effective for acquiring effective processing strategies and thus for multimedia learning, there is also empirical evidence that learners do not always benefit from strategy trainings (Scheiter, Schubert, Gerjets, & Stalbovs, 2014; Schlag, Florax, & Ploetzner, 2007).

For instance, Scheiter et al. (2014) investigated whether a strategy training conveying effective multimedia processing strategies to foster selection, organization, and integration processes before the learning phase would be beneficial for multimedia learning among ninth-grade students. The experimental group received a multimedia strategy training, the control group a training on general learning strategies. In the multimedia strategy training condition, students received printed multimedia material, a list with effective multimedia processing strategies, and a short description of each strategy. In the training phase, the same multimedia material the participants had received previously was presented on an interactive whiteboard. First, an instructor modeled one of the strategies while externalizing her cognitive processes accompanying the application of the respective strategy, then asked the students to apply the same strategy to a subsequent part of the learning material. Thereafter, the instructor provided the students with corrective feedback to their suggestions and documented all visible actions on the whiteboard. The students were also encouraged to take notes on their printed material. Following the first training phase, the students worked in small groups on a second training material on their own. Thereafter, they were again provided with an optimal solution with regard to the application of the strategies. Students in the control group received a similar training on the application of general learning strategies. Although the results indicated that the multimedia strategy training increased strategy knowledge, learners in the experimental group did not outperform learners in the control group regarding their learning performance. Moreover, results also indicated that especially learners with little prior strategy knowledge benefited more from the intervention with respect to strategy acquisition. The authors interpreted their results as indicating that learners acquired the necessary strategic knowledge but were not able to apply this knowledge in a later learning situation. Scheiter et al. (2014) attributed this to the problem that the new strategy instruction may have interfered with learners' preexisting processing strategies. This assumption is also based on findings by Schlag et al. (2007), showing that despite providing learners with strategy knowledge, they do not necessarily apply the gained strategy knowledge during learning.

Schlag et al. (2007) investigated the effectiveness of fostering learning from multimedia material by providing university students with a deep level strategy (i.e., identifying important elements in the text and in the picture; drawing connections between the text and the picture; trying to connect the new information to the prior knowledge). Although Schlag et al. (2007) expected that learning with the deep level strategy would be more effective for multimedia learning than with a surface level strategy (i.e., reading the text; examining the picture; memorizing both information), they did not find any differences regarding the learning outcome between both conditions. However, analyzing the think aloud protocols that were assessed during the learning phase revealed that in many cases, the learners applied a different strategy than they had exercised. Taking this into account, the authors could show that the deep level strategy training was beneficial for those learners who actually applied this strategy during the learning phase.

The studies by Scheiter et al. (2014) as well as by Schlag et al. (2007) provide first indications of potential difficulties with cognitive strategy trainings, namely that simply providing the strategy knowledge does not necessarily result in learners applying the acquired strategy knowledge in the learning process.

From a self-regulatory point of view, inadequate cognitive processing might be caused by learners' inability to regulate their learning (Scheiter et al., 2017). Self-regulated learning requires learners to regulate both their learning process by monitoring their own learning process and their choice of cognitive strategies (Boekaerts, 1999). According to Boekaerts (1999), a key aspect of self-regulated learning is learners' ability to select, combine, and coordinate cognitive strategies in an effective way during the learning process. Hence, the underlying problem of cognitive strategy interventions is that even though learners might have gained the necessary strategy knowledge, they still do not know how to apply the knowledge in an effective way. Moreover, the presence and the automated use of already existing (effective or ineffective) strategies can interfere with strategy trainings that try to convey new strategy knowledge (Garner, 1990; Scheiter et al., 2014; Schlag et al., 2007). This becomes especially problematic when the preexisting strategies exist for a long time and their use is highly automated (e.g. for adult learners).

EMME as a modeling example address this problem by not only providing (new) strategy knowledge, but by demonstrating learners how to regulate the use of these strategies at the same time. By enabling learners to monitor a model that applies the processing strategies effectively, EMME offer learners the opportunity to reflect implicitly and / or explicitly on their own use of processing strategies. As a consequence, EMME are effective for multimedia learning, because they encourage a more conscious use of effective strategies rather than an automated use of existing ineffective strategies. Effective strategies can thus gradually replace ineffective strategies and the use of effective strategies can then be automated over time. Against this background, another underlying mechanism for the effectiveness of EMME relates to the way EMME may help learners to better regulate already existing processing strategies by providing them with the information how to process which information and in which order (Mason et al., 2017; Scheiter et al., 2017).

However, when looking at the results of studies using EMME in a multimedia context, not all learners benefit equally from EMME (Mason et al., 2016; Scheiter et al., 2017). This leads to the assumption that there might be factors that influence whether learners are able to benefit from EMME. As described in the section on the theoretical foundations of EMME, when learning from modeling examples, both visible/derivable model characteristics (e.g. age, competence), as well as learner characteristics (e.g. previous knowledge, cognitive abilities) can influence whether learners are able to learn from modeling examples. Consequently, from a (meta-) cognitive perspective on EMME, especially factors related to learners' cognitive processing might influence the effectiveness of EMME.

According to the social cognitive learning theory, attention processes during observational learning involve “the self-directed exploration of the environment and construction of meaningful perceptions from ongoing modeled events” (Bandura, 1986, p. 53). Therefore, learners’ capability for information processing limits the amount of observational learning, especially during brief exposures to the modeled behavior (Bandura, 1986). Importantly, learners can only benefit from modeling when the modeled events occur at a rate or level of complexity that does not overtax their cognitive skills (Bandura, 1986). In the case that the modeled behavior is too complex for the learner, there is a risk that the acquired representation model of the modeled behavior is fragmentary at best. It is assumed that learners with higher cognitive skills and prior knowledge can perceive more of the subtleties in a modeled behavior and therefore might benefit more from observational learning (Bandura, 1986). Moreover, learners with more prior knowledge might be better able to detect errors in the model’s performance while novice learners do not possess the knowledge to detect these errors and might rehearse erroneous behavior (Van Gog & Rummel, 2010).

The importance of considering learners’ level of prior knowledge is also underlined by empirical evidence showing that it does not only influence the effectiveness of observational learning (Bandura, 1986), but also the learning process directly (e.g., Canham & Hegarty, 2010; Kalyuga, 2013). For instance, Canham and Hegarty (2010) demonstrated that when learning from graphics, learners’ prior domain knowledge can influence information selection and encoding as well as the interpretation of the encoded information. Another example for the influence of prior knowledge is the empirical finding that the effectiveness of instructional support can depend on learners’ prior knowledge, such as that instructional methods which are effective for learners with less prior knowledge can become ineffective for learners with more prior knowledge (expertise reversal effect; Kalyuga, 2013). Moreover, there is empirical evidence that a mere availability of prior knowledge is not always enough to improve learning performance, but that it has to be activated (Wetzels, Kester, & van Merriënboer, 2011). It is assumed that the activation of prior knowledge can provide learners with a relevant context in which they can integrate new information (Wetzels et al., 2011). In light of the finding that especially learners with lower prior knowledge benefit from prior knowledge activation (Wetzels et al., 2011), supporting learners with lower prior knowledge in creating a relevant context by activating domain-relevant prior knowledge could be a possibility to enable them to benefit from observational learning.

Based on the above-mentioned considerations prior knowledge has an important role in learning from modeling examples (Bandura, 1986; Van Gog & Rummel, 2010). In line with this assumption, Groenendijk et al. (2013a) revealed that for gaining knowledge and competences in creativity tasks only high-aptitude students benefited from observational learning. In their study, the students (61 ninth graders) were either asked to observe and evaluate videos with peers doing design tasks or to execute these design tasks themselves. The results indicated a beneficial effect of observation especially for high aptitude students. Moreover, Scheiter et al. (2017) demonstrated that only learners with higher prior knowledge benefited from EMME, whereas learners with lower prior knowledge were even hampered by EMME.

However, there is also empirical evidence that under certain circumstances, also learners with lower prior knowledge can benefit from observational learning. For instance, there are empirical indications that for novices or learners with lower prior knowledge *peer-based learning* might be beneficial (LeBel et al, 2018). In a study with students from a medical school, LeBel et al. (2018) investigated whether novices benefit more from expert models or from non-expert models for learning surgical skills. Learners either watched a video with an expert model demonstrating VR arthroscopy tasks, or a novice model (peer model) demonstrating the same tasks, or received no video. Then, they were asked to perform the same task on a VR arthroscopy simulator. The results revealed that all participants were able to improve their performance over five test sessions. Further, LeBel et al. (2018) discovered that one week after the instruction via the videos, learners in the non-expert model condition outperformed learners in the expert model condition as well as learners in the condition without video modeling. In the fourth and fifth test session, learners in both modeling conditions outperformed learners without modeling. They concluded that for novices at a very early stage of training, observing a non-expert model might be more beneficial. The results from LeBel et al. (2018) regarding a possible interaction between learner characteristics in form of prior knowledge and model characteristics in form of perceived model competence support the assumption that modeling examples create a social learning situation in which social cues can influence the effectiveness of modeling examples.

As previous EMME-research focused solely on learner characteristics such as (domain) prior knowledge for the effectiveness of EMME for multimedia learning, it is up to now an open question of whether social factors play a role. However, the research on modeling examples presented above shows that it might be important to take a closer look at the influence of social factors. Thus, in the following I will outline possible social mechanisms that could underlie the effectiveness of EMME for multimedia learning and factors that may influence their effectiveness.

3.3.3 The social perspective on EMME: EMME as a social learning situation

From a social perspective, EMME are effective because they create a social learning situation by using human eye movements as attention-guiding cue. Together with providing social cues in the introductory text, it is assumed that this increases the salience of the social situation which in turn, fosters learning (social agency theory; Mayer et al., 2003; Moreno et al., 2001). However, one of the questions that arise when EMME are regarded from a social perspective is the question of whether learners are able to regard EMME as a social learning situation. Contrary to other attention-guidance methods, in the case of EMME the visual guidance is established by showing a model's eye gaze rather than an abstract cue such as arrows. Using a model's eye gaze is assumed to be especially effective as an attentional-guidance cue because human gaze has an important communicative role (Gallagher-Mitchell et al., 2017).

People develop sensitivity to other people's gaze from an early age and use their gaze as an important source of information (Frischen, Bayliss, & Tipper, 2007; Symons, Lee, Cedrone, & Nishimura, 2004). Following another person's gaze represents a form of joint attention (Gallagher-Mitchell et al., 2017). According to Gallagher-Mitchell et al. (2018), gaze cues that represent where another person is looking can guide people's attention towards task-relevant areas and enhance their performance, regardless of whether the gaze is in real-time or pre-recorded as in videos such as EMME. Moreover, there is empirical evidence that people process even abstract stimuli (e.g., moving dots) differently when they believe that the agent behind the stimuli is human (Stanley, Gowen, & Miall, 2007), that is, when they ascribe anthropomorphic qualities to the stimulus. In fact, fMRI studies have shown that agency-instructions (i.e., human agent vs. artificial agent) can influence peoples' perception of a stimulus (Stanley, Gowen, & Miall, 2010). In addition, there is empirical evidence that people differ in their performance when they assume that abstract stimuli (red dots) have a human origin (human gaze; Gobel, Tufft, & Richardson, 2018), such as that performance is higher when anthropomorphic qualities are ascribed to the stimulus. Furthermore, people are not only capable of interpreting abstract cues as human gazes but are also able to make sense of the dynamic eye movement patterns of other persons that are visualized by abstract stimuli (i.e., moving circle; Van Wermeskerken, Litchfield, & Van Gog, 2018).

With regard to EMME, it can be hypothesized that learners connect the rather abstract spotlights used within them to human eye movements because learners are usually told that these spotlights correspond to the eye movements of another learner. Thus, learners should interpret the abstract visual cues (e.g., spotlights) in EMME as a social cue (i.e., the model's gaze). Moreover, with view to the results by Van Wermeskerken et al. (2018), they should be capable of interpreting the displayed eye movements as (cognitive) processing strategies.

In the case that learners interpret the abstract visual cues as a social cue, it is assumed that a social context (e.g., a social learning situation) emerges and social factors should come into play. For instance, the social agency theory (Mayer et al., 2003; Moreno et al., 2001) proposes that social cues prime learners to perceive the learning situation as a specific form of social interaction, thereby leading them to process the material more deeply and engage in more sense making processes which in turn foster learning outcome. In line with this assumption, social cues such as on-screen agents displaying humanlike gesturing, movement, voice, eye contact, and facial expressions have been shown to support learners in processing the learning material more deeply (e.g., Mayer, 2014b; Moreno et al., 2001; Töpper, Glaser, & Schwan, 2014; Wang, Li, Mayer, & Liu, 2018). Hence, human eye movements might serve as a social cue, thereby yielding deeper learning. Against this background, perceiving EMME as a social learning situation should foster multimedia learning. But which factors might influence the social mechanisms underlying the effectiveness of EMME?

As described in the chapter on the theoretical foundations of EMME, (alleged) model expertise / competence is presumed to be an influencing factor for observational learning. According to the social cognitive learning theory (Bandura, 1986), learners tend to pay more attention to models that are considered effective and have achieved good results with their behavior (Bandura, 1986; Schunk, 1987). Regarding the model's expertise, it can be distinguished between expertise in an absolute and in a relative sense (Van Gog, Paas et al., 2009). While domain experts are experts in their respective knowledge domain, they are not necessarily the best models to demonstrate the right behavior in a learning or problem-solving task. That means that due to their expertise, they may skip certain behavioral steps in a modeling situation that they are no longer consciously aware of (i.e., automated domain-related strategies), but that would be important for (novice) learners to achieve the same outcome (Chi, Glaser, & Farr, 1988; Van Gog, Paas et al., 2009). In comparison, expertise in a relative sense can also mean that the model itself is not an expert, but only has a higher degree of expertise compared to the learner (Van Gog, Paas et al., 2009). This can be more beneficial for learners as they perceive the model as more similar, which in turn fosters their self-efficacy beliefs and increases the likelihood that they will try to adopt the modeled behavior (Bandura, 1994; Schunk, 1987). To my knowledge, up to date there is only one study by Litchfield et al. (2010) that took model and observer expertise for the effectiveness of EMME into account. Litchfield et al. (2010) hypothesized that showing the eye movement patterns of an expert radiographer to novice radiographers would improve their decision-making performance, whereas they expected that other expert radiographers would be hampered in their performance. Importantly, participants received no explicit information about the identity or the level of expertise of the model. Results indicated neither an effect of model expertise, nor of observer expertise, nor an interaction between these two factors. Litchfield et al. (2010) argued on the basis of these results that the primary factor in learning from another person's gaze replay task-specificity of the eye movements, and not necessarily the model's expertise.

Other modeling examples often display models that show a good task performance or are instructed on how to perform a certain task instead of domain experts (e.g., Mason, Pluchino, 2015, Mason et al., 2016, Scheiter et al., 2017). For EMME, it might be important to take (alleged) model competence into account for two reasons. For one, previous research indicated that people can interpret eye movement patterns of other persons (Van Wermeskerken et al., 2018). For another, results of studies in the field of medical diagnosis revealed that experts differ in their eye movement patterns from novices (e.g., Gegenfurtner et al., 2011). In combination with the above-mentioned finding that people can interpret eye movement patterns of other persons, one could assume that learners are able to implicitly gain information about the level of the model's competence from simply observing the eye movement patterns. Yet, this might be harder for tasks where identifying effective processing strategies is more difficult for the learners (i.e., multimedia learning).

Therefore, some of the previous studies using EMME provided learners with explicit information about the model (see Table 3, for an overview).

Table 3. Differences in the Information about the Model given in the Instruction before EMME Presentation

Authors	Area of Research / Participants	Information about the Model given in the Instruction
Mason, Pluchino, and Tornatora (2015)	**Multimedia learning:** 7th graders	• Explicit information: Student who read an illustrated text and learned very well from it.
Mason, Pluchino, and Tornatora (2016)	**Multimedia learning:** 7th graders	• Explicit information: Student who read an illustrated text and learned very well from it.
Scheiter, Schubert, and Schüler (2017)	**Multimedia learning:** university students	• Explicit information: Successful learner's eye movements.
Mason, Scheiter, and Tornatora (2017)	**Multimedia learning:** 7th graders	• Explicit information: Student who read an illustrated text and learned very well from it.
Litchfield, and Ball (2011)	**Problem solving:** university students	• Explicit information: Someone else looked who solved the problem.
Van Gog, Jarodzka, Scheiter, Gerjets, and Paas (2009)	**Problem solving:** university students	*No information available*
Van Marlen, Van Wermeskerken, Jarodzka, and Van Gog (2016)	**Problem solving:** university students	**Experiment 1:** • Explicit information: Eye movements made by the model during problem solving.
		Experiment 2: • Explicit information: Eye movements made by the model during problem solving. • Implicit information: male (could be derived from the model's narration).

Authors	Area of Research / Participants	Information about the Model given in the Instruction
Van Marlen, Van Wermes-kerken, Jarodzka, and Van Gog (2018)	**Problem solving:** university students	**Experiment 1:** • Explicit information: Eye movements made by the model during problem solving. • Implicit information: female (could be derived from the model's narration).
	Problem solving: secondary education students	**Experiment 2:** • Explicit information: Eye movements made by the model during problem solving. • Implicit information: female (could be derived from the model's narration).
Jarodzka, Van Gog, Dorr, Scheiter, and Gerjets (2013)	**Perceptual task:** university students	• Explicit information: see what the expert was attending to • Implicit information: male (could be derived from the model's narration).
Salmerón, and Llorens (2018)	**Digital reading:** 9th graders	• Implicit information: participants compared a student using optimal strategies to a stu-dent using less optimal strategies (contrasting cases task)
Litchfield, Ball, Donovan, Manning, and Crawford (2010)	**Medical image diagnosis:** Radiographers (Novices vs. Experts)	**Experiment 1:** • Explicit information: Someone else examining the image
		Experiment 2: • Explicit information: Expert radiologist examining the image
	Medical image diagnosis: Novice radiographers	**Experiment 3:** • Explicit information: Expert radiologist examining the image.
Jarodzka, Balslev, Holmqvist, Nyström, Schei-ter, Gerjets, and Eika (2012)	**Medical image diagnosis:** Medical students	• Explicit information: Expert explaining the motion pattern visible in the video • Implicit information: gender (could be derived from the model's narration).
Seppänen, and Gegenfurtner (2012)	**Medical image diagnosis:** University students	*No information available*
Gegenfurtner, Lehtinen, Jarodzka, and Säljö (2017)	**Medical image diagnosis:** Medical experts and medical students	*No information available*

For instance, Mason, Pluchino et al. (2015) as well as Mason et al. (2016, 2017) described the model to the participants as “a student who read an illustrated text and learned very well from it”. In contrast, Scheiter et al. (2017) instructed their participants that they would see a “successful learner’s eye movements”. However, up to date there is a lack in research considering the consequences of explicitly providing learners with this kind of information. Closing this gap is important, as explicitly providing learners with information about the model’s competence might influence not only learners’ perception of the model, but also learners’ subsequent behavior. For this reason, it might be important for the effectiveness of EMME *how* the model is depicted with regard to his / her competence in the introductory text.

The fact that the consideration of (alleged) model competence could be important for observational learning and thus also for EMME was further underlined by empirical findings which indicated that the (alleged) similarity between learner and model (i.e., model-observer similarity; Schunk, 1987; Schunk & Hanson, 1985) can also influence the effectiveness of modeling examples (e.g. Braaksma et al., 2002; Kim, 2007). Based on the line of reasoning that modeling encourages social comparison (Berger, 1977; Hoogerheide et al. 2018; Johnson & Lammers, 2012), Schunk (1987) presumed that the perceived similarity between learner and model can serve as an important source of information for evaluating the appropriateness of the modeled behavior, for formulating appropriate outcome expectations, and for assessing one's self-efficacy for learning or performing the respective task. Therefore, it is important that learner and model are not too dissimilar from each other; otherwise, the learner cannot identify with the model and as a result does not imitate the model’s behavior (Schunk, 1987; Schunk & Hanson, 1985). Moreover, if dissimilar models are perceived as too advanced, learners might assume that they cannot achieve the demonstrated behavior on their own (Schunk & Hanson, 1985). Accordingly, it is assumed that a higher perceived model-observer similarity should improve not only the likelihood of learners paying attention to the model, but of learners themselves adopting the modeled behavior (Schunk, 1987; Schunk & Hanson, 1985). In terms of (alleged) similarity, this would mean that not necessarily models with the highest (alleged) competence are most effective, but possibly models with a higher (alleged) similarity. In line with this assumption, there is empirical evidence that for younger learners, a peer model is more effective than an adult model (e.g., Rodriguez Buritica, Eppinger, Schuck, Heekeren, & Shu-Chen & Wu, 2015; Schunk & Hanson, 1985) or that for learners with lower prior knowledge, peer models seem to be more effective than expert models (e.g., LeBel et al, 2018).

Overall, the research on the influence of model-observer similarity regarding perceived competence on observational learning produced mixed effects. Although some studies indicated that model-observer similarity was beneficial for academic tasks such as argumentative writing or gaining knowledge on abstract concepts (e.g., Braaksma et al., 2002; Kim, 2007), other studies did not find an influence of model-observer similarity on learning (e.g., Groenendijk et al., 2013b; Hoogerheide et al., 2017; Litchfield et al., 2010; Schunk & Hanson, 1985) or even found contrasting results (e.g., Hoogerheide et al., 2016).

In a study by Braaksma et al. (2002) students observed either peer models performing a writing task or performed the writing tasks themselves. Learners in the two modeling conditions watched videos showing peer-model pairs performing the writing task. In one of the modeling conditions, the learners were asked to focus on the non-competent (weak) model. In the other modeling condition, the learners were asked to focus on the competent (good) model. However, the models were not explicitly labelled as 'competent' or 'weak' in the video. Instead, it was a contrast-comparison task in which the learners compared and evaluated the displayed models in order to identify the respective model. Results of this study indicated that weak learners profited more from focusing on the non-competent model while good learners profited more from focusing on the competent model. These results were supported by findings from Kim (2007).

Kim (2007) investigated the effect of model-observer similarity on learning with an anthropomorphized pedagogical agent. First, university students were divided by their grade point averages into 'strong' and 'weak' learners. Then, they were asked to perform a learning task within an online module on concepts of instructional planning that included an interaction with the anthropomorphized pedagogical agent. In a high-competency condition, the pedagogical agent was designed to simulate an advanced peer. In a low-competency condition, the pedagogical agent was designed to simulate a novice peer with low prior knowledge in the task domain. The results for learning outcome revealed that academically strong learners showed a higher recall performance when learning more from the high-competent agent, while academically weak learners showed a higher recall performance when learning more from the low-competent agent. Furthermore, academically weak learners reported higher self-efficacy when learning more from a low-competent agent.

Contrary to the findings by Braaksma et al. (2002), Groenendijk et al. (2013b) did not find a model similarity effect, but indications that students who focused on a weaker model performed better at a creative task. Groenendijk et al. (2013b) investigated the influence of learner-observer similarity on tenth graders' learning to perform creative tasks. Similar to the procedure in the study by Braaksma et al. (2002), the learners in the study by Groenendijk et al. (2013b) were assigned either to one of two modeling conditions or to a control condition without modeling. Learners in the two modeling conditions were asked to focus on different domain independent processes that are included in a creative process. In one of the modeling conditions, the learners were asked to focus on the non-competent (weak) model. In the other modeling condition, the learners were asked to focus on the competent (good) model. As in the study by Braaksma et al. (2002), the models were not explicitly labelled as 'competent' or 'weak', but the learners compared and evaluated the displayed models in order to identify the respective model. The results of the study revealed no interaction between perceived model competence and prior knowledge. Rather, it showed that students who focused more on the 'weaker' model performed better at the creative tasks than students who focused more on the 'competent' model.

In contrast to the model-observer similarity hypothesis, Hoogerheide et al. (2016) discovered that for learning effective instructional strategies from video-modeling examples adult models were more effective than peer models. In a study with secondary education students, they investigated whether model-observer similarity regarding age and perceived competence influences the learning of effective strategies for troubleshooting parallel electrical circuits from video-modeling examples. For this purpose, learners received video-modeling examples with either an adult model or a peer demonstrating how to troubleshoot electrical circuit problems. Prior, both models were introduced either as having low or high expertise on the learning domain. Results showed that learners benefitted more from adult models than from peer models, regardless of the model's expertise. In another study, Hoogerheide, Loyens, Jadi, Vrins, and Van Gog (2017) again found no effect of model-observer similarity on the learning outcome. They investigated in two experiments the influence of model-observer similarity on learning from worked examples. In the first experiment, the model that created the worked example was introduced as either a male or a female peer student. In the second experiment the model was introduced as either a peer student or a teacher. Results revealed an effect of model-observer similarity on learning outcome for neither gender nor age.

There are different explanations for these mixed results. One possible reason is that in contrast to age or gender of a model, the effect of (alleged) competence of a model is usually a more difficult feature to investigate because it is often linked to other model characteristics (Hoogerheide et al., 2018). For example, competence can be linked to the age of the model in terms of developed knowledge or expertise. Moreover, Hoogerheide et al. (2018) argued that it is possible that (alleged) model-observer similarity plays only a subordinate role when (alleged) task appropriateness comes into play. As an example, they referred to the above-mentioned study by Hoogerheide et al. (2016), which showed that adult models were more effective than peer models. In relation to this finding, they argued that for tasks perceived as more difficult, adult models might be preferred to peer models, since more difficult tasks seem more appropriate for more experienced models. As another factor that can influence the perceived task appropriateness, Hoogerheide et al. (2018) also discussed the influence of the model's gender. According to Hoogerheide et al. (2018) for the relationship between perceived task appropriateness and the model's gender, there are mixed empirical findings with some studies indicating an influence of the model's gender on learning (e.g., Garcia-Rodicio, 2012) and others not indicating an influence (e.g., Linek, Gerjets, & Scheiter, 2010).

Against this background, it becomes clear that investigating the influence of (alleged) competence is difficult when other model characteristics such as age and gender are involved. Therefore, for EMME with acoustic elements where conclusions can be drawn about the gender and age of a model (e.g. Van Marlen et al., 2016, 2018), or when explicit information about these model characteristics is given to the learners, it is possible that age, gender and in some cases also acoustic attractiveness (i.e., agreeable voice) can play a role. For these reasons, there were no additional verbal explanations in the EMME used in the present thesis. It was assumed that under these circumstances, learners could only derive information about the model's competence in two ways, either from the information in the gaze behavior itself or from the description they received in the introductory text before watching the EMME-videos. Since the EMME used in the three studies were identical in all EMME conditions, the information participants could have derived from the model's gaze behavior itself should be identical for all participants. Consequently, it can be assumed that the only information participants received about the model's competence was via the model description that was provided in the introductory text. Moreover, because there was no further explicit information given about the model's gender, it is assumed that (alleged) task appropriateness does not play a role in the present thesis.

Another possible explanation is that there might be other learner characteristics that influence the effectiveness of model-observer similarity on the learning outcome. Hoogerheide et al. (2018) postulated, in line with the social cognitive learning theory, that positive effects of modeling examples might be partly based on model characteristics because learners are encouraged to engage in a social comparison with the model. According to Gibbons and Buunk (1999) there are individual differences regarding a person's tendency to compare oneself to others (i.e., social comparison orientation), resulting in some people tending to engage more in social comparisons and act more upon these comparisons than others (e.g., Corcoran, Crusius, & Mussweiler, 2011). Following the assumption of Schunk and Hanson (1985), modeling fosters social comparisons and therefore, individual differences regarding the learners' tendency to compare themselves to others might influence the effectiveness of model-observer similarity on observational learning. For example, Neugebauer, Ray and Sassenberg (2016) revealed that learners' predisposition for social comparison orientation can influence knowledge exchange in a learning task.

It is assumed that social comparison is a fundamental process that occurs frequently and oftentimes automatically without the full awareness of the individual engaging in the social comparison (Buunk & Gibbons, 2007). Social comparison thus describes any process in which a person relates his / her own characteristics to those of another person (Buunk & Gibbons, 2007). According to Buunk and Gibbons, also person perception is assumed to include social comparison. Although Festinger (1954) claimed that people tend to compare themselves to others similar to themselves, there is also empirical evidence that people also engage in upward and downward comparisons (Buunk & Gibbons, 2007). For instance, regarding the effect of upward comparisons (i.e., comparisons with individuals who are thought to be superior), empirical research indicates that upward comparisons are beneficial for academic success (Blanton, Gibbons, Buunk, & Kuyper, 1999). Based on findings that social comparisons can influence people's judgments, experiences, and behavior by prompting them to relate information, e.g. about other people's behavior, abilities, and success to themselves (Corcoran et al., 2011), it is possible that learners' degree of social comparison orientation also influences the effectiveness of learners' perceived model-observer similarity. Moreover, there is evidence that especially novice learners tend to engage in social comparison (Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003) and therefore it is possible that model-observer similarity is especially important for learners with lower prior knowledge (Hoogerheide, 2016).

Against this background, there are various factors from a social perspective on EMME that might influence the effectiveness of EMME for multimedia learning, such as (alleged) model competence, (alleged) model observer similarity, and learners' social comparison orientation. However, previous research has focused solely on factors such as learner characteristics (e.g., domain knowledge; reading comprehension) that can be derived from a perceptual learning and / or (meta-) cognitive perspective in order to explain the positive effects of EMME on multimedia learning. Consequently, there is still a lack of research focusing on the underlying mechanisms and factors for the effectiveness of EMME that can be derived from a social perspective on EMME. The aim of this thesis was therefore to shed more light on the social side of EMME. In the following, I will present the research questions that have guided the research conducted in this thesis. In addition, I will briefly discuss the three studies carried out to answer these questions.

4 Overview of Research Questions and Studies

EMME can be an effective instructional tool for conveying multimedia processing strategies and thus can foster multimedia learning (Mason, Pluchino et al., 2015, Mason et al., 2016, 2017; Scheiter et al., 2017). However, not all learners benefit equally from EMME, resulting in the assumption that there are several factors that might influence the effectiveness of EMME (Mason, Pluchino et al., 2016; Scheiter et al., 2017). Since previous research has focused in particular on factors that can be derived from a perceptual learning and / or (meta-) cognitive perspective to explain the effectiveness of EMME, the role of the social mechanisms and factors in this context is still unclear. To investigate whether the consideration of social mechanisms and factors contributes to explaining the effectiveness of EMME for multimedia learning, two approaches were used.

In a first approach it was investigated whether social mechanisms contribute to explain the effectiveness of EMME. For this approach, EMME using social cues (i.e., human eye movements, model description in the introductory text) were compared to another attention-guidance tool without social cues (abstract attention-guidance cue, no reference to a human model). This was based on the idea that if social mechanisms play a role, EMME with human eye-movements as prominent social cue should be more effective than a comparable instructional tool without social cues. On the one hand it is possible that merely highlighting relevant information without using a social cue (i.e., abstractly cuing relevant information) provide enough attentional guidance for learners to benefit from EMME. Support for this assumption is provided by previous research, indicating that using attention guidance techniques supports learners' cognitive as well as visual processing (e.g., de Koning et al., 2010; Jamet, 2014; Grant & Spivey, 2003; Ozcelik et al., 2010). On the other hand, it is possible that if learners consider that the model is human, they perceive the learning situation as a social situation and interpret the attentional cue in this social context. This, in turn, would yield deeper learning. This idea is based on the social agency theory (Mayer, 2014b; Moreno et al., 2001; Wang, Li, Mayer, & Liu, 2018) and previous research on the influence of (alleged) agency (Gobel et al., 2018; Stanley, Gowen, & Miall, 2007, 2010). However, if there are no measurable differences between EMME (with social cues) and a comparable attentional guidance tool (without social cues), then it would be necessary to carefully weigh the costs against the benefits of EMME against other instruments before the use of the respective instrument.

The other approach consisted of the direct manipulation of social factors to test if the effectiveness of EMME for multimedia learning is not only influenced by perceptual and / or (meta-) cognitive factors, but also by social factors. Under the assumption that EMME are perceived by learners as a social learning situation, it was of interest whether (alleged) model characteristics and / or the interaction between model and learner characteristics (i.e., model-observer similarity) influence the effectiveness of EMME. In other words, it was tested whether, in line with the social cognitive learning theory (Bandura, 1986), (alleged) model competence would influence the effectiveness of EMME or whether there would rather be an interaction between learners' prior domain knowledge and (alleged) model competence. If results would indicate that (alleged) model competence and / or (alleged) model-observer similarity influence the effectiveness of EMME for multimedia learning, then it would be necessary to pay attention to how the model is introduced in the instruction. Furthermore, it could be necessary to adapt the description of the model to the learners' level of prior knowledge in case that model-observer similarity influences the effectiveness of EMME. Arguments in favor of these assumptions are provided by previous research, indicating that for learning with modeling examples (alleged) model competence and / or (alleged) model-observer similarity influence learning outcomes (e.g., Braaksma et al., 2002; Hoogerheide et al., 2016; Kim, 2007; LeBel et al., 2018; Schunk & Hanson, 1985; Schunk, 1987).

Across both approaches, it was also investigated whether social comparison orientation (Gibbons & Buunk, 1999) plays a role in the effectiveness of EMME as well. If learners' level of social comparison orientation influences the effectiveness of EMME, this would further contribute to explaining why some learners benefit from EMME and others do not. That social comparison orientation could be an influencing factor is supported, for example, by the assumption that during learning from modeling examples, learners are encouraged to engage in a social comparison with the model (Hoogerheide et al., 2018). Moreover, it is assumed that social comparison orientation might play an important role especially for learners with lower prior knowledge (Hoogerheide, 2016). On this basis, it could be further narrowed down for which learners EMME can be an effective instructional tool for multimedia learning. Furthermore, social comparison orientation could also have an influence on whether social or non-social cues are more or less effective. For instance, the question could arise of whether learners with a higher level of social comparison orientation benefit more from an instructional tool with social cues, because they are more aware of a potential social context and therefore more responsive to social cues.

In summary, the following overarching research question guided the research conducted in this thesis: Do social mechanisms and factors contribute to explaining the effectiveness of

EMME for multimedia learning above and beyond perceptual and / or (meta-) cognitive mechanisms and factors?

For each of the two approaches that were used to investigate this question, a separate research question was formulated.

- **Research Question 1 (RQ 1): Do social mechanisms involved in EMME contribute to the effectiveness of EMME beyond the effect of perceptual and (meta-) cognitive mechanisms?**

  In the case that social mechanisms contribute to the effectiveness of EMME, using social gaze cues (i.e., human eye movements) to guide learners' attention should be more effective than using non-social cues. The reason for this is that using social gaze cues triggers a social learning context which, in turn, fosters deeper learning. In the case that both EMME and the non-social attention-guidance tool are equally effective, it could be assumed that the effectiveness of EMME for multimedia learning is mainly based on perceptual and / or (meta-cognitive) mechanisms. RQ 1 was addressed in Experiment 1.
- **Research Question 2 (RQ 2): Do social factors influence the effectiveness of EMME or is the effectiveness of EMME solely influenced by perceptual and / or cognitive factors?**

  If social mechanisms play a role in the effectiveness of EMME, social factors should influence the effectiveness of EMME for multimedia learning. In the case that social factors have no impact on the effectiveness of EMME, it could be assumed that the effectiveness of EMME for multimedia learning is mainly based on perceptual and / or (meta-cognitive) mechanisms. RQ 2 was addressed in all three experiments.

In the three experiments reported in this thesis, the EMME-videos used as instructional tool were the same as in the study by Scheiter et al. (2017). The EMME-videos showed a model's gaze replay of performing effective multimedia processing strategies on the first four pages of the later learning material. Whereas participants in the condition with EMME support received the EMME-videos before the learning phase, participants in the abstract-cuing condition received instructional videos with an abstract visual cue, and participants in the control conditions received no instructional support before the learning phase. The learning material consisted of an illustrated expository text on cell division. Variants of this learning material had been used in previous studies, indicating that learners are able to successfully learn from it regarding a variety of outcome measures (e.g., Scheiter et al., 2017; Schüler, Scheiter, & Gerjets, 2013; Stalbovs et al., 2015). In Experiment 3, participants' domain-relevant prior knowledge was additionally activated in half of the participants before the learning phase. In all studies, participants completed a knowledge test after the learning phase.

In Experiment 1, it was investigated whether social mechanisms contribute to the effectiveness of EMME beyond perceptual and (meta-) cognitive mechanisms (Chapter 5). In an eye tracking experiment with 120 university students, for participants with instructional support, the same effective multimedia processing strategies were either displayed via the eye movements of an instructed model (EMME condition) or via highlighting the sequence of processing the material using an abstract cue (abstract-cuing condition). In a third condition, participants received no instructional support before the learning phase. Experiment 2 consisted of an eye tracking experiment with 119 university students (Chapter 6). In this experiment, it was investigated whether the beneficial effect of EMME for multimedia learning was solely influenced by cognitive factors such as learners' prior domain knowledge or also on the influence of social factors such as (alleged) model competence and / or (alleged) model-observer similarity. Furthermore, it was investigated whether potential differences in learning performance could be explained by changes in learners' visual information processing. In order to investigate the influence of social factors, the description of the model was varied experimentally (competent model vs. peer model). The EMME-videos, however, were the same in both EMME conditions.

Experiment 3 was conducted with 180 university students (Chapter 7). This experiment focused on the joint role of prior knowledge, alleged model competence, and social comparison orientation regarding the effectiveness of EMME. Contrary to Experiment 2, learners' prior domain knowledge was not only assessed at the beginning of the study, but also varied experimentally before the EMME instruction. Furthermore, the potential influence of learners' tendency to compare themselves to others (i.e., social comparison orientation; Gibbons & Buunk, 1999) was explored

In the following Chapters 5 to 7, the experiments outlined above are presented and discussed in more detail. At this point, it should be noted that these three chapters were written in collaboration with my supervisor, Prof. Dr. Katharina Scheiter, and Dr. Anne Schüler in order to submit them to scientific journals for publication. This entails that in these chapters 'we' is used instead of 'I'.

5 Experiment 1

Experiment 1 aimed at investigating the role of social mechanisms compared to perceptual and (meta-) cognitive mechanisms in the effectiveness of EMME. For this purpose, it was investigated whether EMME with a social gaze cue (i.e., human eye movements) are more effective than a comparable attention guidance tool without social cues. By directly comparing these two attention guidance tools, it was attempted to clarify the question of whether social mechanisms can contribute to explaining the effectiveness of EMME for multimedia learning. Furthermore, we were also interested in the role of learners' domain knowledge and social comparison orientation on the effect.

Experiment 1 was conducted as an eye tracking experiment with a between-subjects design with three conditions. Two conditions received instructional support demonstrating effective multimedia processing strategies before the learning phase. Participants in the EMME condition received gaze replays of an instructed human model demonstrating effective multimedia processing strategies. They were also instructed beforehand that the gaze replays were derived from a human model. Participants in the abstract-cuing condition received videos where the same multimedia processing strategies were illustrated by highlighting the processing sequence using a non-social attention guiding cue (i.e., white cross). In contrast to the EMME condition, no reference was made in the instruction for participants in the abstract-cuing condition to the fact that the cueing was based on data obtained from a human model. Importantly, in both instructional support conditions, participants were shown the same effective multimedia processing strategies in the same order and for the same duration. Participants in the control condition did not receive instructional support. During the learning phase, participants' eye movements were recorded to assess their gaze behavior.

5.1 Hypotheses

The following hypotheses were formulated based on the theoretical background:

- Overall, it is expected that participants in the two instructional support conditions (EMME; abstract-cuing) show better visual processing of the learning material than participants in the control condition as revealed by longer picture fixation time and more attempts to integrate the text information and the picture information (Hypothesis 1a). Moreover, participants in both instructional support conditions are expected to show higher learning outcomes for recall and comprehension performance on the knowledge test than participants in the control condition (Hypothesis 1b).

- Moreover, if social mechanisms contribute to explaining the effectiveness of, participants in the EMME condition are expected to show better visual processing of the learning material than participants in the abstract-cuing condition as well as in the control condition as revealed by longer picture fixation time and more attempts to integrate the text information and the picture information (Hypothesis 2a). Moreover, they are also expected to show higher learning outcomes for recall and comprehension performance on the knowledge test than participants in the abstract-cuing condition as well as in the control condition (Hypothesis 2b).
- If the effect of EMME is mainly based on perceptual and (meta-) cognitive mechanisms, participants in both instructional support conditions are expected to show similar visual processing of the learning material as revealed by similar picture fixation times and similar attempts to integrate the text information and the picture information (Hypothesis 3a). Moreover, participants in both instructional support conditions are expected to show similar learning outcomes for recall and comprehension performance on the knowledge test (Hypothesis 3b).

In addition to these hypotheses, it was of interest whether learners' domain knowledge and social comparison orientation influence the effects.

5.2 Method

5.2.1 Participants and design

One hundred and twenty students from a German university took part in the experiment. The number of participants was determined by conducting a power analysis using G*Power 3.1 (Erdfelder, Faul, Buchner, & Lang, 2009) with the effect size *adjusted* R^2=.10 (derived from a prior study by Scheiter et al., 2017), a power of .90, and alpha = .05 resulting in a recommended sample size of 119 participants. Students of biology, medicine, or related fields were excluded from participating due to the learning content (i.e., mitosis). Six participants had to be excluded from data analyses due to technical problems and language problems. Due to participants dropout, the actual power of the experiment was between .88 for the comprehension performance and .99 for the recall performance, picture processing, and integrative viewing behavior.

The remaining 114 students (95 female; *M* = 23.33 years, *SD* = 3.30) were enrolled in different university courses. Participants had normal or corrected to-normal vision and were randomly assigned to one of three experimental conditions: control (no modeling), EMME condition (EMME video + preceding instruction describing a human model), and abstract-cuing condition (abstract-cuing video + preceding instruction without social cues). There were 36 students in the control condition, 37 students in the EMME condition, and 41 students in the abstract-cuing condition. Participation was voluntary and reimbursed with 10 Euro or course credits.

5.2.2 Materials

Learning material. The learning material consisted of an expository illustrated text on cell division (mitosis). It described in several text paragraphs and pictures relevant biological processes and principles on which mitosis is based (cf. Scheiter et al., 2017). The learning content was distributed across 11 pages. The whole text information had an overall length of 1,113 words. It was divided in semantically meaningful paragraphs with varying text lengths per page between 44 and 127 words. The text information was continuously presented on the left-hand side of the respective page. It was accompanied by static schematic pictures on the right-hand side of the respective page. Text and picture information were complementary. The text covered information on relevant processes during mitosis on a more abstract level. In addition to the information that was provided in the text, the pictures contained visual-spatial information about cell structures and processes during mitosis. Participants needed to process both the text as well as the picture information to build a comprehensive mental model of the learning content. For an example of the learning material see Figure 4.

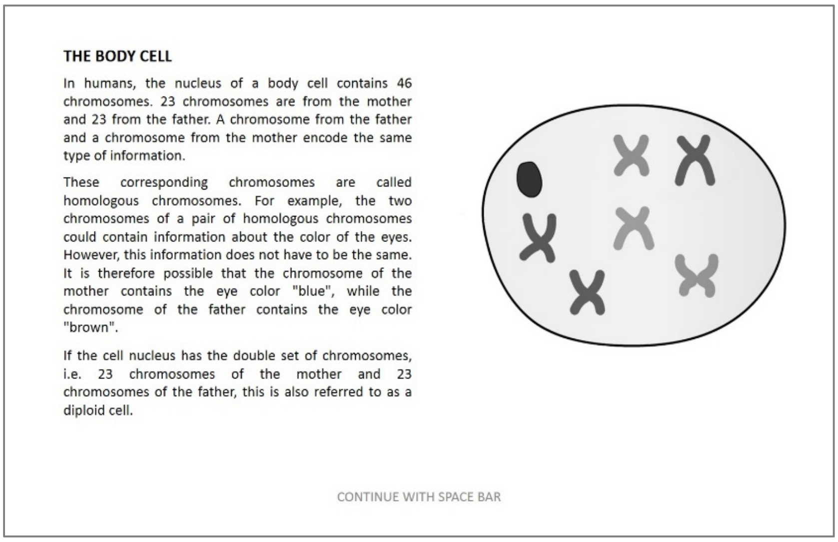


Figure 4: Example of the learning material explaining the terminology and the process in the text as well as showing spatial arrangements of the text-information in the picture.

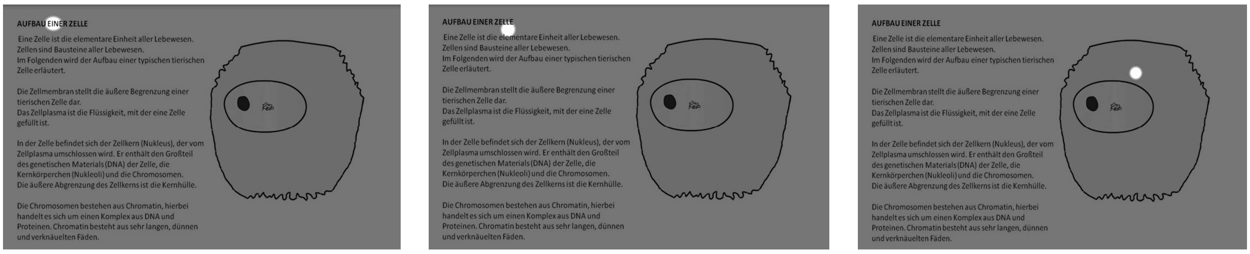
Previous research using variants of the learning material indicated that learners are able to successfully learn from it regarding a variety of outcome measures (e.g., Scheiter et al., 2017; Stalbovs et al., 2015).

For the present study, the learning material was divided into two parts (cf., Scheiter et al., 2017). Part 1 of the learning material consisted of four pages and provided the participants with information about important cell structures, chromosomes, and the concept of DNA. This information was necessary to understand Part 2 of the learning material. Part 2 of the learning material consisted of seven pages. This part of the learning material provided participants with information about basic concepts relevant to mitosis and with a detailed description of the different phases of mitosis. On the first page, participants received a short overview about the whole cell division process. On the following six pages, participants received a description of the separate steps of mitosis. First, the duplication of chromatin fibers and their development into chromosomes were described. This was followed by a description of the development of the mitotic spindle. On the next two pages, participants received information about the function of the equatorial plane and the separation of sister chromatids. The last page described the segregation of daughter cells with genetically identical material. All learning outcome measures referred only to Part 2 of the learning material.

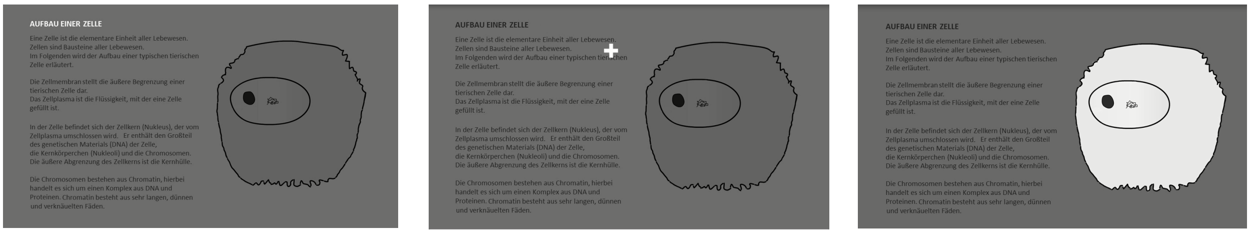
Instructional support videos. For participants in the control condition, Part 1 of the learning material was displayed without instructional support. For participants in both instructional support conditions, Part 1 of the learning material was used to demonstrate effective multimedia processing strategies that had been derived from the literature (cf. Scheiter et al., 2017). The processing strategies were either illustrated via the eye movements of a human model (EMME) or via an abstract cue that highlighted the respective parts of the learning material (abstract-cuing video). In the EMME-video as well as in the abstract-cuing video, participants were shown the same effective multimedia processing strategies in the same order and for the same duration. At the beginning of each page, the process of the construction of a pictorial scaffold was illustrated by first inspecting / highlighting the title of the page and subsequently inspecting / highlighting the picture. This was followed by the demonstration of selection and organization processes illustrated by inspecting / highlighting the text as well as corresponding elements of the picture. The integration process for the EMME condition was illustrated by continuously switching between text and corresponding picture elements. For the abstract-cuing condition it was illustrated by highlighting text and picture after another with using the abstract cue as an indicator which element would be highlighted next. At the end of each page, the process of final picture inspection was illustrated by taking a final look at the picture in case of the EMME condition or highlighting the whole picture in case of the abstract-cuing condition. It was also demonstrated to reread a text section (reaction to comprehension problems) by either using the eye movements on the respective text section to demonstrate the reading process (EMME condition) or by highlighting the text section (cuing condition).

In the EMME-videos, the multimedia processing strategies were illustrated via eye movements of a model on the learning material. For generating the EMME, the model (a student research assistant) was instructed on how to process the learning material. Moreover, she was instructed to behave didactically by demonstrating each process as explicitly as possible (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017). The model's eye movements were visualized by a white spotlight representing a gaze fixation on the otherwise shaded page. The duration of the EMME-videos varied between 51 to 115 seconds per page (M = 95 s), the overall duration was 380 seconds.

The abstract-cuing videos were created by the first author based on the model in the EMME-videos. In the abstract-cuing-videos, the multimedia processing strategies were illustrated by sequentially highlighting the respective parts of the learning material using an abstract cue instead of the eye movements of the model (for an example, see Figure 5). The duration of the strategy-videos varied between 51 to 117 seconds per page (M = 94.5 s), the overall duration was 378 seconds. In contrast to the EMME-videos, the displayed processing strategies were designed to be smoother and less noisy. Importantly, both instructional support videos differed in their instruction. Participants in the EMME condition were instructed that they would see the eye movements of a learner that participated earlier in the study. In contrast, in the abstract-cuing condition no reference was made in the instruction to the fact that the cueing was based on data obtained from a human model and that it reflected that model's eye movements.



EMME: Attention guidance using a social gaze cue (i.e., human eye movements)



Abstract-cuing video: Attention guidance using non-social attention guiding cues (highlighting; abstract cue)

Figure 5: Example of the Instructional Support Videos Illustrating the Multimedia Processing Strategies via the Human Eye or Highlighting Parts of the Material without Social Cues.

5.2.3 Measures

Domain-specific prior knowledge was assessed at the beginning of the experiment. Posttest performance and gaze data (picture fixation time, transitions between text and picture) were assessed as dependent variables. Furthermore, social comparison orientation was assessed after the learning phase. All knowledge measures were collected using the web-based survey software tool platform Qualtrics (http://www.qualtrics.com).

Domain knowledge. Two measures were used to assess participants' domain-knowledge: participants' general scientific literacy and participants' domain specific prior knowledge. The test of general scientific literacy comprised 24 items from the Life Sciences Scale of the Basic Scientific Literacy Test (Laugksch & Spargo, 1996). Participants had to rate statements about scientific processes or interrelationships between scientific concepts as either 'correct', 'incorrect', or 'unknown' (e.g., 'Many of the basic functions of an organism, such as the extraction of energy from nutrients, are carried out at the cellular level.' answer: 'correct'). For each correct answer, participants received one point; for each incorrect answer they received zero points. The maximum total score for scientific literacy was 24 points. Cronbach's alpha for scientific literacy was $\alpha = .67$.

Participants' domain specific prior knowledge about the learning content was assessed with 15 multiple-choice items with four alternatives and one correct answer including questions about cell elements, genetics and mitosis (e.g., ‘Cytokinesis…’ answer: ‘…divides the cell’). For each correct answer, participants received one point; for each incorrect answer they received zero points. The maximum total score for domain specific prior knowledge was 15 points. Cronbach’s alpha for prerequisite knowledge was with a value of $\alpha = .41$ rather low, which is likely to be due to participants partly guessing. A significant correlation of both measures with posttest performance (all *p*s $< .01$) indicated that both measures captured knowledge components that were relevant to the learning domain. Both measures were also significantly correlated with each other ($r = .34$, $p < .01$). A combination of both measures was calculated and referred to as domain knowledge using the sum of the z-standardized scientific literacy and prior knowledge scores. Cronbach’s alpha of the combined measure was $\alpha = .48$.

Learning outcome measures. To measure participants’ posttest performance participants’ recall, and comprehension performance were assessed. Both measures comprised text- and picture-based multiple choice items with four alternatives and one correct answer (e.g., ‘Nucleoli…’ answer: ‘are visible in the interphase.’) and of text- and picture-based forced-choice verification items for which participants should state if the statements or pictures were either true or false (e.g., ‘In the anaphase, the spindles become longer, while the two-chromatid chromosomes migrate to the spindle poles.’ answer: ‘incorrect’). For each correct answer, participants received one point; for each incorrect answer they received zero points. Recall performance was assessed with three multiple-choice items and 17 forced-choice verification items. The maximum total score for recall performance was 20 points (Cronbach’s alpha $\alpha = .54$). Comprehension performance was assessed with nine multiple-choice items and four forced-choice verification items. The maximum total score for comprehension performance was 13 points (Cronbach’s alpha $\alpha = .47$). For both measures, the percentage of correct answers was calculated. No time limit was given for answering the posttest. All items referred only to Part 2 of the learning content for which no instructional support had been displayed.

Social Comparison Orientation. Participants’ social comparison orientation was assessed with 11 items of the Iowa-Netherlands Comparison Orientation Scale (INCOM; Gibbons, & Buunk, 1999). For each item, participants stated on a five-point Likert scale (I disagree strongly – I agree strongly) if they compare themselves to others (e.g., ‘If I want to find out how well I have done something, I compare what I have done with how others have done’). Questions six and ten were reverse coded. The social comparison orientation score was calculated by summing the responses to each question with a higher score indicating a higher tendency to social comparison behavior. Cronbach’s alpha was $\alpha = .81$.

Alleged anthropomorphism. Participants in the two instructional support conditions were asked to rate on a scale from 1 to 7 whether the strategies in the videos were simulated with the help of a computer (1) or whether the strategies were recorded with the help of a human model (7). Subsequently, participants were asked to state in a multiple-choice item with four alternatives whether the strategies were (a) human eye movements, (b) simulated eye movements, (c) human processing strategies, or (d) simulated processing strategies. Moreover, participants in the EMME condition were asked to rate the model's perceived competence in relation to their own competence on a scale ranging from -10 (less competent than themselves) to +10 (more competent than themselves).

Gaze data. Participants' gaze behavior was assessed in Part 2 of the learning phase. Gaze data from Part 1 of the learning material was excluded from the analysis because for participants in both instructional support conditions their gaze behavior on those slides was externally guided either by EMME or by the abstract-cuing videos. For analyzing the gaze data, Areas of Interests (AOIs) were defined for each page of the learning material of Part 2. On each page one AOI encompassing the text as a whole and one AOI encompassing the picture were created. The size of the AOIs differed between the pages with regard to the text length or the size of the picture. However, the size of the respective AOIs did not differ between participants. According to Johnson and Mayer (2012), there are eye-tracking measures that are closely related to the cognitive processes described in the CTML (Mayer, 2009). These include the total fixation time on text / diagram' as indicator for attentional focus on words or pictures as well as integrative transitions as indicator for attempts to integrate words and pictures (Johnson & Mayer, 2012). Against this background, the measures described below were used as an indicator for participants' use of adequate multimedia processing strategies.

Previous research indicated that increased attention to picture information was positively related to learning outcomes (e.g., Eitel et al., 2013; Lin, et al., 2017; Mason et al., 2017). Thus, picture processing was assessed for each participant and each page as the sum of the participants' fixation time (in milliseconds) within the AOI that covered the respective picture. The obtained measure was interpreted as participants' attempt to select and organize picture information. Moreover, there is empirical evidence that the integrative processing of text and picture is crucial for multimedia learning (e.g., Hannus & Hyönä, 1999; Mason et al., 2015a, 2017). Therefore, text-picture integration was assessed as the number of transitions (i.e., the frequency of gaze shifts) between the text and picture AOIs and vice versa. Subsequently, the measure was computed by summing up the total number of transitions between the text and picture AOIs and vice versa. The resulting measure was interpreted as an indicator of the participants' attempts to integrate words and pictures.

Further Measures. The following measures were assessed for exploratory reasons in the experiment but were not included in the analyses: participants’ judgement of learning (JOL), and intrinsic motivation (in the instructional support conditions). JOL was assessed directly after the learning phase. Participants stated on a zero to 100 scale how confident they were that they could answer the following questions about cell division based on the knowledge they gained in the learning phase (c.f., Schleinschok, Eitel, & Scheiter, 2017). For the JOL score, participants’ JOL ratings as percent value were used with higher percentage scores indicating a higher confidence in learning success. For the accuracy of participants’ judgments, we computed the difference between participants’ overall posttest performance and their JOL score (c.f., Schleinschok et al., 2017).

Participants’ intrinsic motivation regarding the instructional videos was assessed using 10 items based on items of the intrinsic motivation short scale (KIM; Wilde, Bätz, Kovaleva, Urhahne, 2009). The items assessed on a five-point Likert scale (doesn't apply at all– applies completely) participants’ experiences with the instructional videos (i.e., if they had fun watching the EMME-videos, if watching the EMME-videos was pleasant for them). Questions two, six, seven, nine and ten were reverse coded. For the intrinsic motivation score, we used an averaged sum of the ratings. Cronbach’s alpha was $\alpha = .84$. Means and standard deviations for these measures can be found in Table 8 in the Appendix.

5.2.4 Apparatus

A SensoMotoric Instruments (SMI) remote eye mobile tracker (RED 250) with a sampling rate of 250 Hz, iViewX and ExperimentCenter 3.6 software was used to record participants’ eye movements. SMI’s BeGaze 3.6 software (www.smivision.com) was used for editing and preparing the eye tracking data for statistical analysis. Settings for fixation and saccade detection were set to default settings (peak velocity threshold = 40 degree/seconds; minimum saccade duration = 22 milliseconds; minimum fixation duration = 50 milliseconds).

5.2.5 Procedure

Participants were tested individually or in small groups up to four participants. Each participant group was randomly assigned to one of the experimental conditions (EMME, abstract-cuing, control). At the beginning, participants received written information on the experimental procedures and signed a consent form. Afterwards, they were seated individually in front of a laptop with the mobile eye tracker. The distance between the participant and the screen was approximately 0.70 m. The first phase was conducted without eye tracking. Participants answered demographic questions, the scientific literacy test and the domain-specific prior knowledge test.

Before the learning phase, we used a 9-point calibration for calibrating the eye tracker for each participant. After the calibration, participants received an onscreen instruction. All participants were informed that they would be allowed to learn at their own pace during the learning phase and proceed to the next page of the learning material by pressing the space bar after the word ‘next’ appeared below the learning content. They were also informed that they would not be able to go back to previous pages. Additionally, participants in the EMME condition were instructed as follows: ‘On the first four pages of the learning material you will see the eye movements of a learner who has participated in this study earlier. This learner used learning strategies that are particularly effective for learning.’ Furthermore, participants in the EMME condition were informed that the eye movements would be illustrated by a white spotlight on a grey shaded page with the size of the spotlight illustrating the learner’s fixation time (e.g., larger spots illustrating longer fixation times).

In contrast, participants in the cuing condition were instructed as follows: ‘On the first four pages of the learning material, you will see various learning strategies that are particularly effective for learning.’ Furthermore, participants in the cuing condition were informed that the strategies would be illustrated by bright markings on a grey shaded page with the highlighted element indicating the location on the learning material to which the strategy currently refers.

For participants in both instructional support conditions the learning phase started with watching either EMME or the abstract-cuing videos on the first four pages (i.e., Part 1) of the learning material. Thereafter, they received Part 1 of the learning material again without instructional support to give them the opportunity to study the material again at their own pace. For participants in the control condition no instructional support was displayed on Part 1 of the learning material. After Part 1, participants entered Part 2 of the learning material, which was identical in all three conditions. To ensure a minimum learning time, each page was displayed for 50 seconds before participants could decide to continue. Participants’ eye movements were recorded during the learning phase. After finishing Part 2 of the learning material, participants completed the posttest regarding the learning content and a questionnaire assessing participants’ social comparison orientation. Thereafter, participants in the EMME condition were asked to rate their perceived similarity with the model with regard to competence. Then participants in both instructional support conditions were asked to assess whether the depicted processing strategies in the respective videos were actual human eye movements, simulated eye movements, actual human processing strategies, or simulated processing strategies. At the end of the experiment, participants were debriefed and paid. Each session lasted about 75 min.

5.2.6 Data Analyses

To investigate the effects of EMME and the abstract-cuing videos on learning outcome and viewing behavior, and their potential moderation by domain knowledge and social comparison orientation, multiple regression analyses with two continuous and a polytomous predictor were conducted. For reporting overall effects, unweighted effect coding was used for the experimental conditions (coding [0.5; 0; -0.5] and [0; 0.5; -0.5] for the conditions EMME, abstract-cuing and control, respectively). Both variables were multiplied with the z-standardized domain knowledge score and the z-standardized social comparison orientation score to obtain the respective interaction terms. For comparing the effects of the respective conditions, experimental conditions were dummy coded with the control condition as the reference category (coding [1; 0; 0] and [0; 1; 0] for the conditions EMME, abstract-cuing and control, respectively). These were multiplied with the z-standardized domain knowledge score and the z-standardized social comparison orientation score to obtain the respective interaction terms.

Experimental condition, the z-standardized domain knowledge score, the z-standardized social comparison orientation score and the generated interaction terms were simultaneously entered as predictors in the multiple regression analyses (Aiken, West, & Reno, 1991). To follow up on significant interaction terms, complex and simple slope analyses were conducted to estimate the size of the effect of the condition at different levels of participants' domain knowledge and / or at different levels of participants' social comparison orientation (Aiken et al., 1991). To determine the effect of the experimental condition for lower values, the effect was estimated at –1 SD relative to the mean of the moderator variable. For higher values, the effect was estimated at +1 SD relative to the mean of the moderator variable.

For all statistical analyses the α level was set to $\alpha = .05$. R^2 values were reported as measure of effect size.

5.3 Results

5.3.1 Learner characteristics

To determine whether the experimental conditions differed regarding participants' domain knowledge and / or social comparison orientation, two regression analyses were conducted with condition (EMME vs. abstract-cuing vs. control) as independent variable and the z-standardized domain knowledge score and the z-standardized social comparison orientation score as dependent variables, respectively.

For domain knowledge, results indicated no significant differences between the EMME ($M = 0.19$, $SD = 1.12$), the abstract-cuing ($M = 0.01$, $SD = 0.90$), and the control condition ($M = -0.20$, $SD = 0.96$), $F(2,111) = 1.39$, $p = .254$, *adjusted* $R^2 = .01$. For social comparison orientation, results also indicated no significant differences between the EMME ($M = 0.17$, $SD = 0.88$), the abstract-cuing video ($M = -0.05$, $SD = 1.12$), and the control condition ($M = -0.12$, $SD = 0.97$), $F < 1$.

5.3.2 Visual processing

Picture fixation time was not normally distributed and therefore was submitted to a log-transformation prior to analysis. Due to missing eye-tracking data, three participants had to be excluded from the analysis for the number of transitions as dependent variable. To investigate whether participants would adopt and use effective visual multimedia processing strategies, two separate multiple regression models were conducted with either picture fixation time or number of transitions as dependent variables and with experimental condition, the z-standardized domain knowledge score, the z-standardized social comparison orientation score and the respective interaction terms simultaneously entered as predictors (see Table 4 for means and standard deviations).

For *picture fixation time*, results revealed no effect for domain knowledge, $F < 1$, social comparison orientation, $F(1, 99) = 2.60$, $p = .110$, *adjusted* $R^2 = -.002$, or experimental condition, $F < 1$. None of the two-way interactions were significant (all Fs < 1, except for the two-way interaction between modeling condition and social comparison orientation, $F(2,99) = 1.21$, $p = .303$, *adjusted* $R^2 = -.002$). Furthermore, also the three-way interaction was not significant, $F(2, 99) = 1.35$, $p = .264$, *adjusted* $R^2 = -.002$. Taken together, these results suggest that the participants in the instructional support conditions did not change their processing strategy regarding picture processing. Moreover, the results indicated that the use of picture processing strategies was not influenced by participants' domain knowledge or social comparison orientation.

For *transitions between text and picture*, results revealed no significant effect for domain knowledge, $F < 1$, or experimental condition, $F < 1$. For social comparison orientation the effect was significant, $F(1, 99) = 9.24$, $p = .003$, *adjusted* $R^2 = .12$, suggesting that overall participants with higher social comparison orientation ($M = 132.76$, $SE = 8.63$) made significantly more transitions between text and picture than participants with lower social comparison orientation ($M = 96.01$, $SE = 8.45$). There was neither a significant interaction between experimental condition and domain knowledge, $F(2, 99) = 1.66$, $p = .195$, *adjusted* $R^2 = .12$, nor between experimental condition and social comparison orientation, $F(2, 99) = 2.21$, $p = .115$, *adjusted* $R^2 = .12$, nor between domain knowledge and social comparison orientation, $F(1, 99) = 1.05$, $p = .309$, *adjusted* $R^2 = .12$. There was also no significant three-way interaction, $F(2, 99) = 2.10$, $p = .128$, *adjusted* $R^2 = .12$. Overall, these results indicated that while participants' social comparison orientation influenced their use of integrative viewing strategies, instructional support condition or domain knowledge had no effect.

5.3.3 Learning outcome

To investigate the effect of experimental condition, domain knowledge and social comparison orientation on learning outcomes, two multiple regression analyses were computed for recall performance and comprehension performance separately. Experimental condition, the z-standardized domain knowledge score, the z-standardized social comparison orientation score and the related interaction terms were entered simultaneously as predictors (see Table 4 for means and standard deviations).

For *recall performance*, results revealed a significant effect of domain knowledge, $F(1, 102) = 12.37$, $p = .001$, *adjusted* $R^2 = .08$, indicating that participants with higher domain knowledge ($M = 71.06$, $SE = 1.96$) outperformed participants with lower domain knowledge ($M = 61.41$, $SE = 1.81$). Furthermore, there was a significant effect of social comparison orientation, $F(1, 102) = 4.52$, $p = .036$, *adjusted* $R^2 = .08$, indicating that participants with higher social comparison orientation ($M = 69.19$, $SE = 1.92$) performed better than participants with lower social comparison orientation ($M = 63.28$, $SE = 1.88$). However, there was no significant effect of experimental condition, $F < 1$. Moreover, none of the two-way interactions were significant (all Fs < 1, apart for the two-way interaction between experimental condition and domain knowledge, $F(2, 102) = 1.05$, $p = .354$, *adjusted* $R^2 = .08$). Results revealed no three-way interaction, $F < 1$. In sum, these results indicated that participants' characteristics such as domain knowledge or social comparison orientation influenced their recall performance. However, instructional support had no beneficial influence on the learning outcome.

For *comprehension performance*, results showed a significant effect of domain knowledge, $F(1, 102) = 4.56$, $p = .035$, *adjusted* $R^2 = .01$, indicating that participants with higher domain knowledge ($M = 55.46$, $SE = 2.30$) performed better than participants with lower domain knowledge ($M = 48.60$, $SE = 2.12$). There was neither a significant effect of experimental condition, $F < 1$, nor of social comparison orientation, $F(1, 102) = 2.48$, $p = .118$, *adjusted* $R^2 = .01$. Results revealed no significant two-way interaction (all Fs < 1, except for the two-way interaction between condition and domain knowledge, $F(2, 102) = 1.03$, $p = .362$,

R^2 = .11). There was also no significant three-way interaction, $F < 1$. Taken together, only participants' domain knowledge influenced their comprehension performance, whereas social comparison orientation or instructional support had no influence on the learning outcome.

Table 4. Means and standard errors (in parentheses) for the learning performance measures and the eye tracking measures as a function of experimental condition.

	EMME (n = 37)	**Abstract cuing** (n = 41)	**Control** (n = 36)
	M (SE)	*M (SE)*	*M (SE)*
Learning outcome			
Recall performance (in %)	66.60 (2.29)	65.65 (2.15)	66.45 (2.30)
Comprehension performance (in %)	50.71 (2.68)	54.16 (2.52)	51.22 (2.70)
Visual processing			
Overall picture fixation duration (ms)	62337.44 (6480.49)	56836.69 (6048.65)	51893.45 (6557.94)
Overall picture fixation duration (log)	10.84 (0.21)	10.47 (0.20)	10.62 (0.22)
Overall number of transitions	120.87 (10.59)	115.77 (9.88)	103.79 (10.71)

5.3.4 Explorative analyses

Effect of visual processing on the learning outcome. To investigate whether participants' visual processing of the learning material influenced their learning performance independent of the experimental condition, we explored the effects of visual processing on participants' recall and comprehension performance. Because picture fixation time and transitions between text and picture were strongly correlated ($r(109) = .72$, $p < .001$), we conducted four separate regression models with either picture fixation time or transitions between text and picture as predictors and with either recall or comprehension performance as dependent variables.

Recall performance. For the first regression model with picture fixation time as predictor, results showed no effect of picture fixation time, $F < 1$. This result suggests that picture fixation duration had no positive effect on recall performance. For the second multiple regression model with number of transitions as predictor, results showed a significant effect of the number of transitions on the learning outcome, $F(1, 109) = 12.18$, $p < .001$, *adjusted* $R^2 = .09$, suggesting that participants with more transitions ($M = 70.54$, $SE = 1.80$) outperformed participants with fewer transitions ($M = 61.62$, $SE = 1.80$). This result indicated that the increased use of integrative processing strategies had a positive effect on recall performance.

Comprehension performance. For the first regression model with picture fixation time, the model was not significant, $F(1, 109) = 1.49$, $p = .225$, *adjusted* $R^2 = .04$. For the regression model with the number of transitions as predictor, results revealed a significant effect, $F(1, 109) = 17.10$, $p < .001$, *adjusted* $R^2 = .13$, suggesting that participants with more transitions ($M = 57.49$, $SE = 2.00$) outperformed participants with less transitions ($M = 45.77$, $SE = 2.00$). Again, this result showed that with regard to comprehension performance, the use of integrative processing strategies had a beneficial effect.

Assessment of the Model. First, it was tested whether the two instructional support conditions differed in terms of the degree of anthropomorphism attributed to the model. As an indicator, participants' assessment with regard to the extent to which the processing strategies presented in the videos were computer animated or human were used (simulated with the help of a computer (1) / strategies were recorded with the help of a human model (7)). The results of a one-way ANOVA indicated a significant difference between the EMME condition and the abstract-cuing condition in the expected direction, $F(1, 74) = 7.47$, $p = .008$, $\eta_p^2 = .09$. Participants in the EMME condition ($M = 3.77$, $SE = 0.26$) rated the model as significantly 'more human' than participants in the abstract-cuing condition ($M = 2.77$, $SE = 0.25$). Participants' domain knowledge and social comparison orientation were used as covariates in the analysis to control for a potential influence. For further analyses one-sample t-tests were conducted for each instructional support condition to investigate whether the respective mean score differed significantly from the mean of the scale ($M = 4.0$). These analyses revealed that only for participants in the abstract-cuing condition, the mean score differed significantly from the mean of the scale, $t(40) = -5.01$, $p < .001$. For participants in the EMME condition, the mean score did not differ significantly from the mean of the scale, $t(36) = -0.91$, $p > .05$. These results suggest that although participants in the abstract-cuing condition perceived their 'model' less human, participants in the EMME condition did not necessarily perceive the model as 'human' as intended.

5.4 Discussion

The aim of Experiment 1 was to investigate the role of social mechanisms in the effectiveness of EMME for multimedia learning. It is assumed that EMME imply a social learning situation as they use human eye movements to cue effective multimedia processing strategies. The human eye movements can be seen as a 'social cue'. According to the social agency theory (Mayer et al., 2003; Moreno et al., 2001) social cues lead learners to perceive a learning situation as a specific form of social interaction. This encourages them to process learning materials more deeply and to engage in more sense making processes, which in turn fosters learning outcomes (Mayer et al., 2003; Moreno et al., 2001). Against this background, human eye movements as social gaze cues could encourage deeper learning and also lead to better learning outcomes. Accordingly, this resulted in the question of whether triggering social mechanisms by using a social attention-guiding cue (i.e., human eye movements) and social cues in the introductory text (i.e., eye movements from a learner who has participated earlier) would result in EMME being more effective than a comparable instructional tool without social cues. For this purpose, EMME were compared to instructional videos with a non-social attention guidance cue and without social cues in the introductory text (i.e., abstract-cuing videos). Importantly, the EMME-videos and the abstract-cuing videos demonstrated the same effective multimedia processing strategies in the same order and for the same duration.

Hypotheses 1a and 1b assumed that overall learners with instructional support would show more intensive visual information processing (longer picture fixation time; more frequent transitions between text and picture) and better learning performance (recall; comprehension) than learners without instructional support. Hypotheses 2 and 3 were formulated as competing hypotheses. Hypotheses 2a and 2b assumed that in the case that social mechanisms contribute to the effectiveness of EMME for multimedia learning, learners in the EMME condition would show more intensive visual information processing and better learning performance than learners in the abstract-cuing condition. Otherwise, it was assumed that in the case that the EMME condition and the abstract-cuing condition do not differ significantly in terms of their effect on visual processing (Hypothesis 3a) and learning outcome (Hypothesis 3b), mainly perceptual and (meta-) cognitive mechanisms underlie the effectiveness of EMME. Furthermore, it was also of interest whether domain knowledge and / or social comparison orientation would influence the effectiveness of the respective instructional tool.

Contradicting Hypotheses 1a and b, results revealed no beneficial effect for either instructional support tool, indicating that participants with instructional support performed at a similar level as participants without instructional support. In line with Hypothesis 3 (and contradicting Hypothesis 2), results showed no significant difference between the EMME condition and the abstract-cuing condition with regard to the use of effective multimedia processing strategies and the learning outcomes. At first glance, this finding speaks in favor of the assumption that mainly perceptual and (meta-) cognitive mechanisms underlie the effectiveness of EMME. However, as all learners performed at a similar level, it cannot simply be concluded that social mechanisms do not play a role in the effectiveness of EMME.

One possible explanation why the three conditions did not differ significantly from each other regarding their learning outcome is that all learners used effective multimedia processing strategies without instructional support. As the exploratory analyses indicated, integrative viewing behavior, however, was predictive for the learning outcome. Whereas in contrast to previous research (e.g., Eitel et al., 2013; Lin et al., 2017) the present results revealed no beneficial effect of picture processing on the learning outcome, the findings for the integrative viewing behavior are in line with previous research, showing that in particular text-picture integration fosters multimedia learning (e.g., Mason, Pluchino, 2016; Mason et al., 2017). Thus, in the case that learners in all conditions showed similar effective multimedia processing, it is not surprising that they also would not differ regarding their learning outcomes. Another possible explanation could be that domain knowledge influenced the effectiveness of the instructional support tools, as the results from previous research suggest (e.g., Scheiter et al., 2017). However, even though results revealed an overall effect of domain knowledge on the learning outcome, there was neither a moderating effect of domain knowledge on the effectiveness of EMME, nor on the effectiveness of the abstract-cuing videos. Another factor that was assumed to potentially moderate the effectiveness of the instructional videos was social comparison orientation. It would have been possible for learners with a higher social comparison orientation to benefit more from EMME, as EMME provide an (alleged) human model these learners could have compared to. However, results did not indicate an influence of social comparison orientation on the effectiveness of either instructional video.

Another possible explanation for the finding that EMME were not effective in Experiment 1 is offered by the exploratory analyses of learners' assessment of the respective model. These indicated that while learners in the EMME condition perceived the model as 'more human' than in the abstract-cuing condition, the mean response was still relatively low. This might be problematic because in order to trigger social mechanisms, learners need to perceive the learning situation as a social interaction. On this basis, it is possible that the social mechanisms were not triggered to such an extent that they could have had a beneficial effect. Moreover, if mainly social mechanisms are responsible for the effectiveness of EMME, then this would also explain why neither instructional support condition outperformed the control condition. However, this assumption is only speculative and based on exploratory analyses.

5.4.1 Limitations and implications for future studies

A limitation of Experiment 1 is that the abstract-cuing videos were used in Experiment 1 for the first time. We attempted to make them as similar as possible to the EMME-videos with the difference that they contained no social cues. However, it is unclear whether they are generally suitable for conveying strategy knowledge. Hence, their effectiveness should be tested in future studies. Another limitation is that due to the lack of effectiveness of the manipulation of the (alleged) anthropomorphism in the instructional videos, the role of social mechanisms in the effectiveness of EMME remained unclear. Moreover, even though social comparison orientation was not relevant in the present experiment, it is possible that there are social factors that influence the effectiveness of EMME. Against this background, further research is needed to investigate if social aspects play a role in the effectiveness of EMME, and whether there are social factors that might influence the effectiveness of EMME for multimedia learning.

5.4.2 Conclusions

Taken together, Experiment 1 does not allow drawing conclusions about the role of social mechanisms in the effectiveness of EMME for multimedia learning. Nonetheless, it provides further indication that EMME might be not generally effective for multimedia learning. This suggest that aside from prior knowledge as shown by Scheiter et al. (2017), there might be also other factors that influence the effectiveness of EMME for multimedia learning. To draw firmer conclusions on this aspect, however, further research is necessary.

6 Experiment 2

Experiment 2[1] focused on the question of whether (alleged) model competence and / or (alleged) model-observer similarity as social factors influences the effectiveness of EMME for multimedia learning.

For measuring not only learning outcome, but also the use of multimedia processing strategies during the learning phase, Experiment 2 was conducted as an eye tracking experiment with a between-subjects design with three conditions. Two of these three conditions received EMME before the learning phase. These EMME consisted of eye movement recordings of an instructed model demonstrating effective multimedia processing strategies. The eye movements were overlaid onto the first four pages of the to-be-processed learning material. To investigate the influence of model-descriptions with regard to model competence on the effectiveness of EMME, (alleged) model competence was experimentally varied between the two EMME conditions. Learners received either the competent-model description or the peer-model description before watching the EMME. Importantly, the EMME itself were the same across conditions. A third condition, the control condition, did not receive EMME.

6.1 Hypotheses

With regard to the question of whether also social factors can influence the effectiveness of EMME for multimedia learning, two sets of hypotheses were formulated. The first set of hypotheses assumes that only cognitive factors explain the effectiveness of EMME. The second set of hypotheses assumes that also social cues such as model-observer similarity may influence the effectiveness of EMME.

If the effect of EMME was solely based on cognitive factors, we hypothesize that

- Participants in both EMME conditions show better visual information processing than participants in the control condition as revealed by longer picture fixation time and more attempts to integrate the text information and the picture information (*Hypothesis 1)*

[1] Experiment 2 was published as Krebs, M.C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: The influence of model-observer similarity on Eye Movement Modeling Examples. *Learning and Instruction, 61,* 126–137. https://doi.org/10.1016/j.learninstruc.2018.10.005). In accordance with the granted author rights for scholarly purposes by Elsevier, the subchapters are adopted from the original article.

- Participants in both EMME conditions show higher learning outcomes for recall and comprehension performance on the knowledge test than participants in the control condition (*Hypothesis 2)*
- Based on previous research by Scheiter et al. (2017), we expected EMME to improve learning outcomes for learners with high prior knowledge, but not for those with low prior knowledge (*Hypothesis 3)*

However, if social cues, such as model competence and thus model-observer similarity, influence participants' readiness to follow the model's gaze on the material, then EMME effects depend on the information about the model relative to learners' prior knowledge. In line with the model-observer similarity hypothesis, we hypothesized that

- Prior knowledge moderates the influence of alleged model competence on visual information processing, such that participants with *higher* prior knowledge show better visual information processing as revealed by longer picture fixation time and more attempts to integrate the text information and the picture information when the EMME are described as resulting from a 'successful learner' (i.e., high model-observer similarity) than when the EMME are described as resulting from 'another participant' (i.e., low model-observer similarity) or without EMME. *(Hypothesis 4a)*
- Furthermore, we expect participants with *lower* prior knowledge to show better visual information processing as revealed by longer picture fixation time and more attempts to integrate the text information and the picture information when the model is described as 'another learner' and therefore perceived as a peer learner (i.e., high model-observer similarity) compared to when the model is described as a 'successful learner' (i.e., low model-observer similarity) or without EMME. *(Hypothesis 4b)*
- Prior knowledge moderates the influence of alleged model competence on learning outcome such that participants with *higher* prior knowledge show higher learning outcomes for recall and comprehension performance on the knowledge test when the EMME are described as resulting from a 'successful learner' (i.e., high model-observer similarity) than when the EMME are described as resulting from 'another participant' (i.e., low model-observer similarity) or without EMME. *(Hypothesis 5a)*
- Furthermore, we expect participants with *lower* prior knowledge to show higher learning outcomes for recall and comprehension performance on the knowledge test when the model is described as 'another learner' and therefore perceived as a peer learner (i.e., high model-observer similarity) compared to when the model is described as a

‘successful learner’ (i.e., low model-observer similarity) or without EMME. *(Hypothesis 5b)*

We used mediation analyses to test the extent to which differences in participants’ visual information processing - as a function of EMME-condition and participants’ prior knowledge - explain possible differences in learning outcomes.

6.2 Method

6.2.1 Participants and design

One hundred and nineteen students from a university in the southwestern part of Germany took part in the study. The number of participants was determined by conducting a power analysis using G*Power 3.1 (Erdfelder et al., 2009) with the effect size R^2=.10 (derived from a prior study by Scheiter et al., 2017), a power of .90, and alpha = .05 resulting in a recommended sample size of 118 participants. Students of biology, medicine, or related fields were excluded from participating due to the learning content (i.e., mitosis). Eight participants had to be excluded from data analyses due to the above-mentioned exclusion criteria, technical problems, language problems, or missing data (due to participants’ drop-out the actual power of the study was .88). The remaining 111 students (23.02 years, *SD* = 2.71; 92 female) were enrolled in different university courses. Participants had normal or corrected to-normal vision. The design was a between-subjects design with experimental condition as between-group variable. Participants were randomly assigned to one of three conditions: control (no EMME), competent model (EMME + preceding instruction describing a competent model), or peer model (EMME + preceding instruction describing a peer model). There were 37 students in each condition. Participation was voluntary and reimbursed with 12 Euro or course credits.

6.2.2 Materials

Variants of the learning material had been used in previous studies, indicating that students are able to successfully learn from it regarding a variety of outcome measures (e.g., Scheiter et al., 2017; Schüler et al, 2013; Stalbovs et al., 2015). The learning material consisted of an expository illustrated text on cell division as one would also find in a textbook on the same matter. It was distributed across 11 pages and described relevant biological processes and principles on which mitosis is based and provided learners with professional terminology. The whole text had an overall length of 1,113 words and was divided into semantically meaningful paragraphs. Text lengths per page varied between 44 and 127 words. The text was accompanied by static schematic pictures on the right-hand side of the respective page. The text described the processes during mitosis on a more abstract level, while the pictures provided additional visuo-spatial information about cell structures and processes during mitosis. Text and picture were complementary and both necessary to understand mitosis. For an example of the learning material see Figure 6.

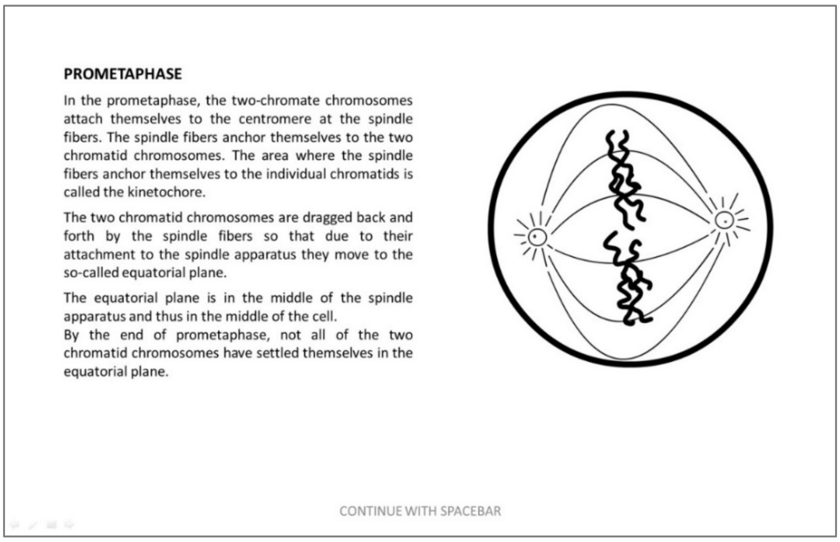


Figure 6: Example of the learning material explaining the terminology and the process in the text as well as showing spatial arrangements of the text-information in the picture.

For the present study, the learning material was divided into two parts (cf., Scheiter et al., 2017). Part 1 of the learning material consisted of four pages providing information about important cell structures, chromosomes and the concept of DNA. This information was necessary to understand Part 2. Part 2 consisted of seven pages. On these seven pages, basic concepts relevant to mitosis as well as the phases of mitosis were described in more detail. First, participants received a short overview about the process. Subsequently, the duplication of chromatin fibers and their development into chromosomes were described. This was followed by the description of the development of the mitotic spindle, the function of the equatorial plane, and the separation of sister chromatids. On the final page the segregation of daughter cells with genetically identical material was described. Learning outcome measures referred only to Part 2 of the learning material.

For participants in the control condition, Part 1 of the learning material was displayed without visual guidance via EMME. For participants in both EMME conditions, Part 1 of the learning material was used to create EMME showing effective multimedia processing strategies that had been derived from the literature (cf., Scheiter et al., 2017).

For every page, the model first read the title of the page. This text processing strategy was selected based on previous research indicating that the inspection of the text's headings as global text processing strategy, for example, fostered learners' recall of the content (Lorch et al., 2013; Sanchez et al., 2001) as well as their ability to summarize the content (Hyönä et al., 2002). After inspecting the title, the learners inspected the picture on the right-hand side of the page (construction of pictorial scaffold). Empirical evidence indicated that by inspecting the picture first, learners gain a more comprehensive representation of the picture's visuospatial components, which supports learners in the construction of a cohesive mental model (e.g., Eitel et al., 2013). Furthermore, there is first evidence that supporting this processing strategy is both beneficial for picture processing as well as integrative processing of text and picture information (Mason et al., 2017). Thereafter, the model read the text and looked at corresponding elements of the picture (selection of relevant words and picture elements; text organization; picture organization). The model switched between text and corresponding picture elements while reading (integration of both representations). It has been shown that this is an effective processing strategy for learning with multimedia (Hegarty & Just, 1993; Mason, Tornatora, et al., 2013, 2015). At the end of each page, the model took a final look at the picture (final picture inspection) and in some cases reread a text section (reaction to comprehension problems). Based on findings that learners' often take a final look at a picture before continuing (Hegarty & Just, 1993), it is assumed that learners use this strategy to assess whether the information they gleaned from the text matches with information they gleaned from the picture.

The model's eye movements on the learning material were visualized by a white spotlight representing a gaze fixation on the otherwise shaded page. The model was a student research assistant who was given instructions beforehand regarding the order in which text and images should be processed. In line with previous research using EMME, the model was instructed to process the materials didactically by demonstrating each process as explicitly as possible (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017). Duration of the EMME-videos varied between 52 to 115 seconds per page (M = 95.5 s), the overall duration was 382 seconds.

6.2.3 Measures

We assessed participants' domain knowledge as potential moderator at the beginning of the experiment. Posttest performance and gaze data (picture fixation time, transitions between text and picture) were assessed as dependent variables. All knowledge measures were collected using the web-based survey software tool platform Qualtrics (http://www.qualtrics.com).

Domain knowledge. To assess participant's domain knowledge, two different measures were used: a test of participants' prerequisite knowledge on cell division to measure the domain-specific prior knowledge about the learning content and a test of general scientific literacy to measure general knowledge about relevant concepts in the life sciences. Prerequisite knowledge was assessed with 15 multiple-choice items with four alternatives and one correct answer including questions about cell elements, genetics and mitosis (e.g., 'What are chromatids?' answer: 'The longitudinal hemispheres of a chromosome'). Correct answers were given one point; incorrect answers were given zero points. The maximum total score for prerequisite knowledge was 15 points. Cronbach's alpha was with a value of $\alpha = .40$ rather low. This, however, could be attributed to the fact that in the prior knowledge test participants were probably partly guessing.

Participants' scientific literacy was assessed with 24 items from the Life Sciences Scale of the Basic Scientific Literacy Test (Laugksch & Spargo, 1996). Participants had to rate statements about scientific processes or interrelationships between scientific concepts as either 'correct', 'incorrect', or 'unknown' (e.g., 'Each gene is one - or more than one - specific segment of a DNA molecule.' answer: 'correct'). Correct answers were given one point; incorrect answers were given zero points. The maximum total score for scientific literacy was 24 points. Cronbach's alpha was $\alpha = .74$.

Both measures were correlated with posttest performance (all *p*s < .01), indicating that both measures captured knowledge components that were relevant for the learning domain. We calculated a combined measure of both measures referred to as domain knowledge consisting of the sum of the z-standardized scientific literacy and prior knowledge scores. The correlation between both measures was significant ($r = .33$, $p < .01$). Cronbach's alpha of the combined measure was $\alpha = .40$.

Learning outcomes. The posttest consisted of recall and comprehension items. Both measures consisted of text- and picture-based multiple choice items with four alternatives and one correct answer (e.g., 'How many chromosomes does each of the two daughter cells contain at the end of the telophase?' answer: '46 One-Chromatid Chromosomes') and of text- and picture-based forced-choice verification items for which participants had to state if the statements or pictures were either true or false (e.g., 'At the end of mitosis, a new nuclear envelope is formed from the cell plasma.' answer: 'false'). Correct answers were given one point; incorrect answers were given zero points. For both measures the percentage of correct answers was calculated. Recall performance was assessed with three multiple-choice items and 17 forced-choice verification items. The maximum total score for recall performance was 20 points

(Cronbach's α = .58). Comprehension performance was assessed with nine multiple-choice items and four forced-choice verification items. The maximum total score for comprehension performance was 13 points (Cronbach's α = .45). No time limit was given for answering the posttest. All items referred only to Part 2 of the learning content for which no EMME had been displayed. This was done to ensure that the posttest would assess learning effects of EMME rather than visual guidance effects.

Gaze data. To test the assumption that for learners in the EMME condition social cues have an influence on their multimedia processing strategies, we assessed participants' gaze behavior in Part 2 the learning phase. Fixations and saccades as eye-tracking parameters are assumed to reflect voluntary, overt visual attention (Duchowski, 2007). According to Duchowski (2007), fixations are usually interpreted as the desire to focus attention on an object of interest. In contrast, saccades are interpreted as the desire to change the focus of attention to another object.

For analysis, we defined Areas of Interests (AOIs) for each page of the remaining learning material (Part 2). One AOI encompassing the text as a whole and one AOI encompassing the picture were created on each page. The size of the AOIs differed between the pages with regard to the text length or the size of the picture but did not differ between participants. As we were interested in differences between conditions, but not between pages, the obtained gaze measures were not adjusted for differences in AOI size and exposure times.

According to Johnson and Mayer (2012) there are eye-tracking measures that are closely related to the cognitive processes described in the CTML (Mayer, 2009). Johnson and Mayer (2012) described 'total fixation time on text / diagram' as indictor for attentional focus on words or pictures and 'integrative transitions' as indicator for attempts to integrate words and pictures.

Based on previous eye tracking studies indicating that the integrative processing of text and image is crucial for multimedia learning (e.g., Hannus & Hyönä, 1999; Mason, Pluchino, et al., 2015; Mason et al., 2017; Mason, Tornatora et al., 2015; O'Keefe et al., 2014), we assessed for each participant the number of transitions (i.e., the frequency of gaze shifts) between the text and picture AOIs and vice versa for each page for Part 2 of the learning material. The measure was computed by summing up the number of transitions of fixations from the text AOI to the picture AOI and vice versa. The measure received was used as indictor for participants' attentional shift between the text and picture AOIs. Based on previous research, it was interpreted as an indicator of the participants' attempts to integrate words and pictures.

Furthermore, previous research has found that increased attention to picture information is positively related to learning outcomes (e.g., Eitel et al., 2013; Lin et al., 2017; Mason et al., 2017). We therefore assessed for each participant the picture fixation time (in milliseconds) for each page for Part 2 of the learning material. Picture fixation time for each page described the sum of the learners' fixation time within the AOI that covered the respective picture. The measure received was interpreted as learners' attention to picture elements and in correspondence to the CTML (Mayer, 2009) interpreted as learners attempt to select and organize picture information.

Both measures served as an indicator for participants' use of adequate multimedia processing strategies.

6.2.4 Apparatus

A SensoMotoric Instruments (SMI) remote eye tracker (RED 250) with a sampling rate of 250 Hz, iViewX and ExperimentCenter 3.4 software was used to record the participants' eye movements. The learning material was presented on a 22" widescreen monitor. For editing and preparing the eye tracking data for statistical analysis we used SMI's BeGaze 3.6 software (www.smivision.com). Settings for fixation and saccade detection were set to default settings (peak velocity threshold = 40 degree/seconds; minimum saccade duration = 22 milliseconds; minimum fixation duration = 50 milliseconds).

6.2.5 Procedure

Participants were tested in individual sessions and were randomly assigned to one of the experimental conditions (competent model, peer model, control). At the beginning, they received written information on the experiment procedures and signed a consent form. Then, participants answered demographic questions, the domain-specific prior knowledge test, and the scientific literacy test. Afterwards they were seated in front of the eye tracker, which was calibrated using a 9-point calibration. After the calibration, participants were informed that during the learning phase they could learn at their own pace and proceed to the next page of the learning material by pressing the space bar after the word 'next' appeared below the learning content but that they could not return to previous pages. Additionally, participants in both EMME conditions were informed that they would see short videos with a learner's eye movements, which would be illustrated by a white spotlight on a grey shaded page with the size of the spotlight illustrating fixation duration. Furthermore, participants in the 'competent model condition' were informed as follows: 'On the first four pages of the learning material you will see the recorded eye movements of a learner who used successful strategies and therefore scored well on the knowledge test.' In contrast, participants in the 'peer model condition' were informed as follows: 'On the first four pages of the learning material you will see the recorded eye movements of a learner who participated earlier in the experiment'. The EMME themselves were, however, identical in both conditions.

For participants in the EMME conditions the learning phase started with watching EMME on the first four pages (i.e., Part 1) of the learning material. Thereafter, they were shown the first four pages of the learning material again without EMME to give them the opportunity to study the material again at their own pace. For participants in the control condition no EMME were displayed on the first four pages (Part 1). After Part 1 had been finished, participants entered Part 2 of the learning material, which was identical for all three groups. To ensure a minimum learning time in all conditions each page was displayed for 50 seconds before participants could decide on their own to continue by pressing the space bar. Participants' eye movements were recorded during the learning phase. After the learning phase, participants answered the posttest at their own pace and were debriefed and paid. A single session lasted about 60 min.

6.2.6 Data Analyses

To test the effects of EMME on learning outcome and viewing behavior, and their potential moderation by model-observer similarity, multiple regression analyses with a continuous and a polytomous predictor were conducted. For reporting overall effects, we used unweighted effect coding for the experimental conditions (coding [0.5 0 -0.5] and [0 0.5 -0.5] for the conditions competent model, peer model and control, respectively). Both variables were furthermore multiplied with the standardized domain knowledge score to obtain two interaction terms.

To compare the effects of the respective conditions, experimental conditions were dummy coded with one reference category (coding [1 0 0] and [0 1 0] for the conditions competent model, peer model and control, respectively), which were additionally multiplied with the standardized domain knowledge score for two interaction terms.

Experimental conditions, the z-standardized domain knowledge score, and the generated interaction terms were simultaneously entered as predictors in the multiple regression analyses (Aiken, West, & Reno, 1991). To follow up on significant interaction terms, simple slope analyses were conducted to estimate the size of the effect of the condition at different levels of participants' domain knowledge (Aiken et al., 1991). To determine the effect of the experimental condition for lower values, the effect was estimated at – 1 SD relative to the mean of domain knowledge. For higher values, the effect was estimated at + 1 SD relative to the mean of domain knowledge.

To follow-up on possible significant findings for learning outcomes as well as visual processing, moderated mediation models were constructed to test whether a possible relation between experimental condition, domain knowledge and learning outcome might be mediated by participants' visual processing (PROCESS Model 8; Hayes, 2013). For all statistical analyses the α level was set to $\alpha = .05$.

6.3 Results

To determine whether the experimental conditions differed regarding participants' domain knowledge an analysis of variance (ANOVA) was conducted. Results indicated no significant difference between the competent model condition ($M = -0.14$, $SD = 1.00$), the peer model condition ($M = 0.17$, $SD = 1.09$), and the control condition ($M = -0.03$, $SD = 0.91$), $F < 1$.

6.3.1 Visual processing

Based on the assumption that participants would change their use of visual processing strategies, we investigated the effect of condition and the influence of domain knowledge on visual processing behavior. To this end, we conducted two separate multiple regression models with either picture fixation time or number of transitions as dependent variables and with experimental condition, the z-standardized domain knowledge score and the respective interaction terms simultaneously entered as predictors. The dependent variables were not normally distributed and therefore were submitted to a log-transformation prior to analysis. Means and standard deviations for picture fixation time and number of transitions as a function of condition and domain knowledge are displayed in Table 5.

For *picture fixation time*, the results showed no main effect for domain knowledge, $F < 1$. There was a marginal main effect of condition, $F(2, 101) = 2.94$, $p = .057$, *adjusted* $R^2 = .07$ (Cohen's $f^2 = .08$), which was qualified by a significant interaction between domain knowledge and condition, $F(2, 101) = 3.12$, $p = .049$, *adjusted* $R^2 = .07$ (Cohen's $f^2 = .08$). Simple slope analyses revealed that for participants with relatively high domain knowledge the effect of condition was not significant, $F < 1$. For participants with relatively low domain knowledge, there was a significant effect of condition, $F(2, 105) = 6.00$, $p = .003$. Participants in the competent model condition, $b = 0.52$, $SE = 0.17$, $\beta = .46$, $p = .002$ as well as participants in the peer model condition, $b = 0.52$, $SE = 0.18$, $\beta = .46$, $p = .004$, looked significantly longer at pictures than participants in the control condition. There was no significant difference between both EMME conditions, $p = .998$.

For the *number of transitions between text and picture*, results showed no significant influence of participants' domain knowledge, $F < 1$, condition, $F(2, 101) = 1.85$, $p = .163$, *adjusted* $R^2 = .01$, or an interaction, $F < 1$.

6.3.2 Learning outcome

To investigate the effect of condition and the influence of domain knowledge on learning outcomes, multiple regression analyses were computed separately for recall performance and comprehension performance. Condition, the z-standardized domain knowledge score, and the related interaction terms were entered simultaneously as predictors (see Table 5 for means and standard deviations).

For *recall performance*, the results revealed a main effect of domain knowledge, $F(1, 105) = 5.98$, $p = .016$, *adjusted* $R^2 = .03$ (Cohen's $f^2 = .03$), indicating that participants with higher domain knowledge ($M = 70.34$, $SD = 17.18$) performed better than those with lower domain knowledge ($M = 64.69$, $SD = 16.98$). There was neither a significant main effect of condition, $F < 1$, nor an interaction between domain knowledge and condition, $F(2, 105) = 1.00$, $p = .368$, *adjusted* $R^2 = .03$.

For *comprehension performance*, the results showed a significant main effect of domain knowledge, $F(1, 105) = 10.23$, $p = .002$, *adjusted* $R^2 = .10$ (Cohen's $f^2 = .11$), indicating that participants with higher domain knowledge ($M = 61.38$, $SD = 21.86$) performed better than those with lower domain knowledge ($M = 51.97$, $SD = 21.61$). There was no main effect of condition, $F < 1$. Results revealed a significant interaction effect between domain knowledge and condition, $F(2, 105) = 3.34$, $p = .039$, *adjusted* $R^2 = .10$ (Cohen's $f^2 = .11$). To investigate the effect of condition for participants with lower and higher domain knowledge, simple slope analyses were conducted. For participants with relatively high domain knowledge the omnibus effect of condition was not significant, $F < 1$. However, for participants with relatively low domain knowledge, results revealed a significant effect of condition, $F(2, 105) = 3.38$, $p = .038$. Results indicated that participants in the peer model condition scored marginally higher on comprehension items than participants in the control condition, $b = 9.77$, $\beta = .29$, $p = .063$ and significantly higher than participants in the competent model condition, $b = 12.44$, $\beta = .37$, $p = .013$ (see Figure 7). Participants in the competent model condition showed similar performance as participants in the control condition, $p = .589$.

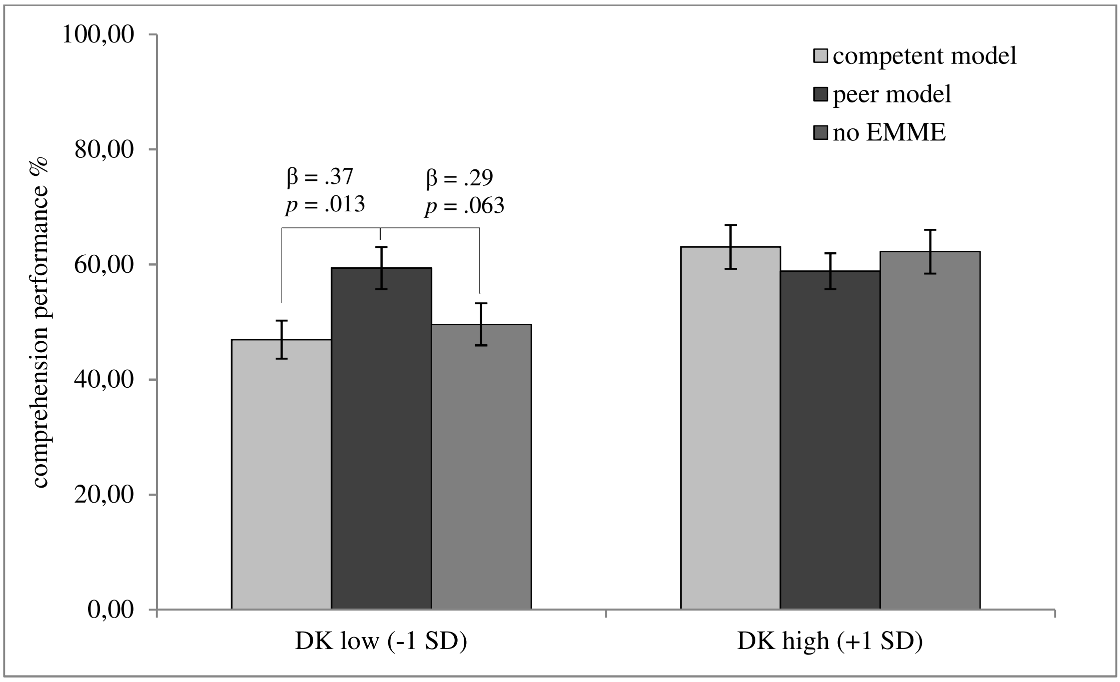


Figure 7: Comprehension performance in percent as a function of condition and domain knowledge (DK).

6.3.3 Moderated mediation

We tested whether the difference in comprehension performance was due to a change of picture-related gaze behavior (i.e., picture processing) moderated by domain knowledge. Because of the experimental design with a polytomous predictor with three categories, two separate analyses were necessary. Based on the results revealing no significant differences between the competent model condition and the control condition regarding the learning outcome, we compared in two separated moderated mediation models only the effects for peer model condition versus control condition and the effects for peer model condition versus competent model condition. To test for specific patterns, we used orthogonal contrast coding for the experimental conditions to create the independent variable for the respective analysis. Because we did not find an effect of condition and learners' domain knowledge for the number of transitions between text and pictures, the moderated mediation models were only constructed for picture-related gaze behavior.

In the first moderated mediation model, the effects for peer model condition versus control condition were compared. For this purpose, the conditions were recoded into two orthogonal contrasts comparing the peer model condition (coding [1]) to the competent model condition (coding [0]) and to the control condition (coding [-1]) (i.e., condition-contrast). The second contrast (coding [-1, 2, -1]) was treated as a covariate in the analysis (i.e., covariate-contrast).

The condition-contrast as independent variable, the covariate-contrast as covariate, domain knowledge as moderator variable, picture fixation time as mediator variable, and comprehension performance as dependent variable were simultaneously entered in the analysis. Results of the moderated mediation analysis revealed a marginally significant total effect of the interaction between the condition-contrast and domain knowledge on comprehension performance for participants with lower domain knowledge, *b* = 5.04, *SE* = 2.62, *p* = .057, 95% CI [-0.14 | 10.23]. This effect was reduced to non-significance when picture fixation time was entered in the analysis as a mediator, *b* = 2.24, *SE* = 2.55, *p* = .382, 95% CI [-2.82 | 7.30]. The indirect effect of the interaction between condition and domain knowledge via picture fixation time on comprehension performance was significant, *b* = -1.62, *SE* = 0.73, 95% CI [-3.35 | -0.38]. The indirect effect was only significant for participants with relatively low domain knowledge, *b* = 2.86, *SE* = 1.02, 95% CI [1.11 | 5.01]. For participants with average, *b* = 1.23, *SE* = 0.84, 95% CI [-0.10 | 3.12] or relatively high domain knowledge, *b* = -0.40, *SE* = 1.21, 95% CI [-2.54 | 2.07] the indirect effect was not significant. These results suggest that the difference in learning outcome for learners with lower domain knowledge can be explained by longer picture fixation time in the peer model condition (see Figure 8).

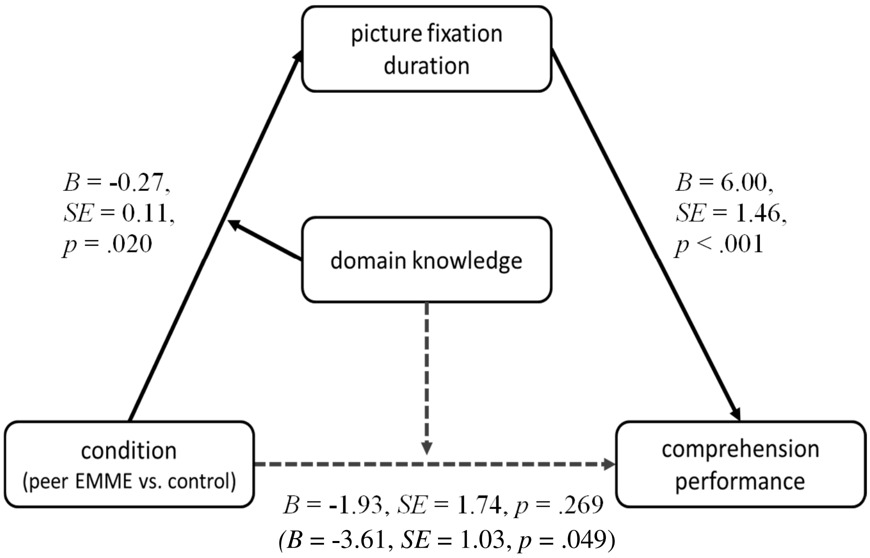


Figure 8: Moderated mediation model with the unstandardized regression coefficients for the direct effect and the total effect (in parentheses) of condition and domain knowledge on comprehension performance, as well as the path via picture fixation duration (indirect effect). The independent variable 'condition' represents the condition-contrast, with the coefficients 1 (peer model), 0 (competent model) and -1 (control).

In the second moderated mediation model, the effects for peer model condition versus competent model condition were compared. For this purpose, the conditions were recoded into two orthogonal contrasts. The first contrast compared the peer model condition (coding [1]) to the competent model condition (coding [-1]) and to the control condition (coding [0]) (i.e., condition-contrast). The second contrast (coding [-1, -1, 2]) was treated as a covariate in the analysis (i.e., covariate-contrast). The condition-contrast as independent variable, the covariate-contrast as covariate, domain knowledge as moderator variable, picture fixation time as mediator variable, and comprehension performance as dependent variable were simultaneously entered in the analysis. Results of the moderated mediation analysis revealed a significant total effect of the interaction between the condition-contrast and domain knowledge on comprehension performance for participants with lower domain knowledge, $b = 6.18$, $SE = 2.51$, $p = .014$, 95% CI [1.29 | 11.06]. The analysis revealed no indirect effects of the interaction between condition and domain knowledge via picture fixation time on comprehension performance, $b = -0.46$, $SE = 0.59$, 95% CI [-1.52 | 0.80]. These results suggest that for learners in the peer model condition and the competent model condition, the difference in learning outcome for learners with lower domain knowledge was not due to differences in picture fixation time in the peer model condition.

Table 5. Means and standard errors (in parentheses) for the domain knowledge measures, the learning performance measures and the visual processing measures as a function of experimental condition and domain knowledge.

Learning Performance

	Competent model (*n* = 37)		**Peer model** (*n* = 37)		**Control** (*n* = 37)	
Domain knowledge	**low** *M (SE)*	**high** *M (SE)*	**low** *M (SE)*	**high** *M (SE)*	**low** *M (SE)*	**high** *M (SE)*
Domain knowledge (z-standardized)	-1.86 (0.27)	1.40 (0.27)	-1.35 (0.29)	1.92 (0.29)	-1.68 (0.24)	1.58 (0.24)
domain-specific prior knowledge	4.57 (0.37)	9.11 (0.37)	5.51 (0.36)	10.05 (0.36)	4.70 (0.37)	9.24 (0.37)
scientific literacy	13.15 (0.64)	20.69 (0.64)	13.53 (0.67)	21.07 (0.67)	13.61 (0.56)	21.15 (0.56)
Learning outcome						
Recall performance (in %)	66.06 (2.59)	70.21 (3.00)	67.20 (2.89)	69.67 (2.46)	60.80 (2.88)	71.15 (2.98)
Comprehension performance (in %)	46.93 (3.30)	63.08 (3.81)	59.37 (3.68)	58.82 (3.13)	49.60 (3.67)	62.23 (3.80)
Visual processing						
Overall picture fixation duration (ms)	71018.81 (6744.82)	63681.95 (7797.94)	70987.81 (7526.67)	63108.43 (6407.04)	45147.52 (7494.17)	70549.22 (7761.99)
Overall picture fixation duration (log)	11.11 (0.11)	11.02 (0.13)	11.11 (0.12)	10.89 (0.11)	10.59 (0.12)	10.98 (0.13)
Overall number of transitions	91.26 (9.91)	81.03 (11.46)	79.00 (11.06)	74.73 (9.41)	61.20 (11.01)	81.33 (11.40)
Overall number of transitions (log)	4.39 (0.14)	4.31 (0.16)	4.28 (0.15)	4.07 (0.13)	3.97 (0.15)	4.18 (0.16)

6.4 Discussion

The study investigated to what extent the effectiveness of Eye Movement Modeling Examples (EMME) as an instructional tool for learning with multimedia is influenced by social cues. In particular, we focused on the question of whether alleged model-observer similarity as social cue influences how well students learn from EMME. Furthermore, we examined whether differences in learning performance are explained by changes in learners' visual information processing. We formulated two sets of hypotheses (see section 1.3). The first set of hypotheses was based on the assumption that the effect of EMME is based on cognitive factors only (Hypothesis 1-3). The second set of hypotheses was based on the assumption that social cues in the form of model-observer similarity may also influence the effectiveness of EMME (Hypothesis 4 and Hypothesis 5).

Hypotheses 1 and 2 assumed that learners of both EMME conditions would show more intensive visual information processing (longer picture fixation time; more frequent transitions between text and picture) and better learning performance (recall; comprehension) than the control group. Hypothesis 3 assumed EMME to support learning outcomes only for learners with higher prior knowledge (Scheiter et al., 2017). Overall, we were not able to confirm Hypotheses 1-3, indicating that the beneficial effect of EMME might not only depend on cognitive factors. Hypothesis 4 and Hypothesis 5 assumed that next to cognitive factors also social cues in the form of model-observer similarity may influence the effectiveness of EMME. Based on the model-observer similarity hypothesis, we assumed that domain knowledge moderates the influence of alleged model competence on visual information processing (Hypothesis 4) and performance (Hypothesis 5), so that learners who perceive high model-observer similarity show more intensive visual processing and higher performance.

Contrary to Hypothesis 4, watching EMME prolonged the picture fixation time of learners with lower domain knowledge independent of the information given about the model, whereas there was no effect of watching EMME for learners with higher domain knowledge. In line with previous research by Scheiter et al. (2017), the results regarding picture processing substantiates the claim that EMME enhance visual processing of multimedia material for adult learners. In contrast to previous research, this effect depended on learners' domain knowledge.

According to Siegler (2007), strategy change can be described by means of four component processes: acquisition of new strategies, increasing use of the most advanced existing strategies, increasingly efficient execution of strategies, and improved choice among strategies. Performing new or less familiar strategies requires a shift in learners' strategy use, in the sense that there is a higher probability for them to choose a more effective strategy from a broad range of available strategies (e.g., Garton, 2008). It is possible that in our study learners with lower domain knowledge increased their use of beneficial already existing strategies, whereas learners with higher domain knowledge were probably able to use them adequately, also without EMME. This assumption is supported by our findings that learners with higher domain knowledge in the control condition displayed longer picture fixation time during the learning phase.

Regarding Hypothesis 5 concerning comprehension performance, the results provide partial support and speak in favor of the assumption that the effect of EMME is also based on social factors such as (alleged) model competence. In particular, there was an interaction between learners' domain knowledge and alleged model competence. In line with Hypothesis 5b, learners with lower domain knowledge benefitted from EMME, but only when they received the information that the model was a 'peer learner' (i.e., high model-observer similarity). When they were confronted with an alleged 'high competent' model (i.e., low model-observer similarity), watching EMME had no effect (Hypothesis 5a). Consistent with previous research showing beneficial effects of interventions primarily for transfer knowledge (e.g., Ozcelik et al., 2010; Richter et al., 2016), we found the influence of alleged model-observer similarity only on learners' comprehension performance but not on recall performance.

Our results confirm assumptions from social cognitive theory regarding the influence of model-observer similarity (Bandura, 1986) as well as previous findings from Braaksma et al. (2002). Although in both EMME conditions the model was introduced as a learner, there is empirical evidence that people's perceptions of their own characteristics guide their estimates of others (e.g., Krueger & Stanke, 2001). Therefore, without further information about the model's knowledge, learners with lower domain knowledge should tend to perceive the model as a learner with similarly low domain knowledge.

Moreover, it is possible that learners with less domain knowledge found the gap between themselves and the 'high competent' model to be too wide. Thus, they could not identify with the model and assumed that they could not achieve the demonstrated behavior alone (Schunk & Hanson, 1985) which in turn prevented them from benefitting from the 'high competent' model. In contrast, when learners with less domain knowledge are not actively pushed into a social comparison with the model, they might feel less 'threatened' by the model and thus are also able to benefit from a 'competent' model. Therefore, emphasizing the model's peer learner status might be one way of circumventing the detrimental effect of EMME for learners with lower domain knowledge found by Scheiter et al. (2017).

On the other hand, model-observer similarity did not play a role for learners with relatively high domain knowledge, which is not in line with either Hypothesis 5a or Braaksma et al. (2002) or Hoogerheide, Van Wermeskerken et al. (2016). It might well be that in the present study, students with relatively high domain knowledge were able to sufficiently rely on their existing schemata and were hence less receptive for social cues, whereas for learners with lower domain knowledge the social cues might have been more salient and therefore had a greater influence (cf. Mayer, 2014a).

To investigate whether the differences in comprehension performance for learners with lower domain knowledge were due to a change in picture processing moderated by learners' domain knowledge, a moderated mediation analysis was conducted. For learners with lower domain knowledge, changes in picture processing could explain differences in learning outcome between the peer-model and the control condition. Accordingly, in line with previous research demonstrating that intense picture processing is an important processing strategy (e.g., Scheiter & Eitel, 2015), picture fixation time explained the effects of model-observer similarity.

However, our results also suggest that learners with lower domain knowledge increased their picture processing when they were confronted with a competent model – without this being mirrored by an increase in their learning performance. This finding implies that there might be other factors involved that affect the relation between learners' gaze behavior and performance. One possibility is that the perceived difference between model and learner lead to differences in social activation, which in turn resulted in different learning outcomes (Mayer, 2014a). It is possible that these learners only looked at the pictures without processing them actively and therefore failed in building a coherent mental model. However, social activation was not assessed, and the question of which factors are exactly involved still remains open.

6.4.1 Limitations and implications for future studies

On the one hand, the results of our study suggest that also adults with lower domain knowledge can benefit from EMME, when it is suggested that these were generated by a peer learner. On the other hand, we were not able to replicate prior findings showing that EMME were especially helpful for adult learners with higher domain knowledge (Scheiter et al., 2017). Therefore, more research is necessary to determine the level of learners' domain knowledge that is required for learners to be able to benefit from EMME.

Another limitation of our study is that in contrast to prior studies we were not able show an influence of EMME on learners' integrative viewing behavior (Mason, Pluchino et al., 2015a, 2015b; Mason, Scheiter et al., 2017; Scheiter et al., 2017). The question yet remains open whether learners were incapable to identify this particularly processing strategy as helpful or whether they just did not need this strategy to process this specific learning material. There is empirical evidence that more integrative viewing behavior is not always predictive for better learning outcomes (e.g., Scheiter & Eitel, 2015). It is possible that at a certain point an increase in integrative viewing behavior mirrors no longer integration, but comprehension problems or informational overload. Thus, more research regarding the role of integrative viewing behavior in learning performance is needed.

Although previous studies show that EMME can be generally effective in different areas of learning (see Table 2), more research is also needed to unravel the underlying mechanisms of EMME. Knowledge of the underlying mechanisms can then provide us with information on how EMME must be designed to be effective and who can benefit from EMME. With our study we took a first step in this direction showing that social cues such as perceived model-observer similarity might influence the effectiveness of EMME for learners with lower prior knowledge. However, further research is needed to investigate the conditions under which social factors underlie the effectiveness of EMME and should therefore be taken into account when designing EMME in the learning context.

6.4.2 Conclusions

Our results imply that using EMME to provide learners with insights into a model's cognitive processing can alter adult learners' general information processing of the learning material and foster their learning performance. These findings contribute to the research regarding EMME as an instructional tool for learning with multimedia. Moreover, we extend this research by providing first indications that the effectiveness of EMME is not solely based on cognitive factors such as the level of learners' domain knowledge but also social cues such as alleged model competence. Our findings hence provide support for grounding effects of EMME and video-based modeling more generally in social cognitive theories of learning (Van Gog & Rummel, 2010). In particular for learners with lower domain knowledge the framing of the model as 'peer' learner is important for altering their processing strategies and increasing their learning performance.

7 Experiment 3

Experiment 3 aimed at investigating the joint role of prior knowledge, (alleged) model competence, and social comparison orientation for the effectiveness of EMME. Moreover, Experiment 3 also focused on replicating the results from Experiment 2 regarding the influence of (alleged) model-observer similarity on the effectiveness of EMME. In particular, the effect of prior knowledge activation was investigated. Here it was focused on the question whether learners with lower domain knowledge would benefit from EMME independently of the alleged model observer similarity if their domain prior knowledge had been activated beforehand.

Experiment 3 was conducted as a laboratory experiment with a 2 x 3 between-subjects design. The first factor 'domain-relevant prior knowledge' (PK) referred to the activation of learners' domain-relevant prior knowledge. Before the learning phase, half of the participants received a text containing domain-relevant information (i.e., information about cell biology) to activate their domain-relevant prior knowledge. The other half received a text containing domain-irrelevant information (i.e., information about the Mayan history). By reading the text and subsequently recalling as many facts as possible from the text, domain-relevant prior knowledge was assumed to be either activated (cell biology text) or not activated (Mayan history text). The second factor 'modeling condition' was also experimentally varied between participants. Two-third of the participants received EMME directly before the learning phase. These EMME consisted of recorded eye movements of an instructed model demonstrating effective processing strategies on the first four pages of the to-be-processed learning material. As in Experiment 2, participants were given different information about the model before the EMME presentation (i.e., that the model was introduced as being either highly competent or a peer model) to manipulate the (alleged) model competence. Importantly, whereas the instructions differed between EMME conditions, the EMME itself were the same between conditions. The remaining participants did not receive EMME (control condition).

7.1 Hypotheses

Against the theoretical background and the questions derived from it, the hypotheses listed below were formulated (see Figure 9).

- Participants with more domain knowledge in both EMME conditions are expected to outperform participants in the control condition regardless of whether their prior knowledge had been previously activated (Hypothesis 1). Moreover, no significant difference in learning outcomes is expected between the two EMME conditions for learners with more domain knowledge (EMME effect irrespective of alleged model competence; Hypothesis 2).

- For participants with less domain knowledge and without prior knowledge activation it is expected that model-observer similarity would influence the effectiveness of EMME. More specifically, in line with the model-observer similarity hypothesis and the findings from Experiment 2, it is hypothesized that participants with less domain knowledge show higher learning performance when the model is described as ‘another learner’ (i.e., high model-observer similarity) compared to when the model is described as a ‘successful learner’ (i.e., low model-observer similarity) or without EMME (Hypothesis 3) (see Figure 9, left).
- For participants with less domain knowledge and with prior knowledge activation, however, it is hypothesized based on positive effects of prior knowledge activation for learners with less domain knowledge (Wetzels et al., 2011), that they would benefit from EMME regardless of (alleged) model-observer similarity. Thus, participants with less domain knowledge and with prior knowledge activation in both EMME conditions are expected to outperform participants in the control condition (Hypothesis 4). Moreover, no significant difference in learning outcomes is expected between the two EMME conditions (Hypothesis 5) (see Figure 9, right).

In addition to the hypotheses regarding the effect of prior knowledge activation and model observer-similarity for learners with more or less domain knowledge, it is of interest whether learner’s tendency to compare themselves with others (i.e., social comparison orientation) would influence possible effects.

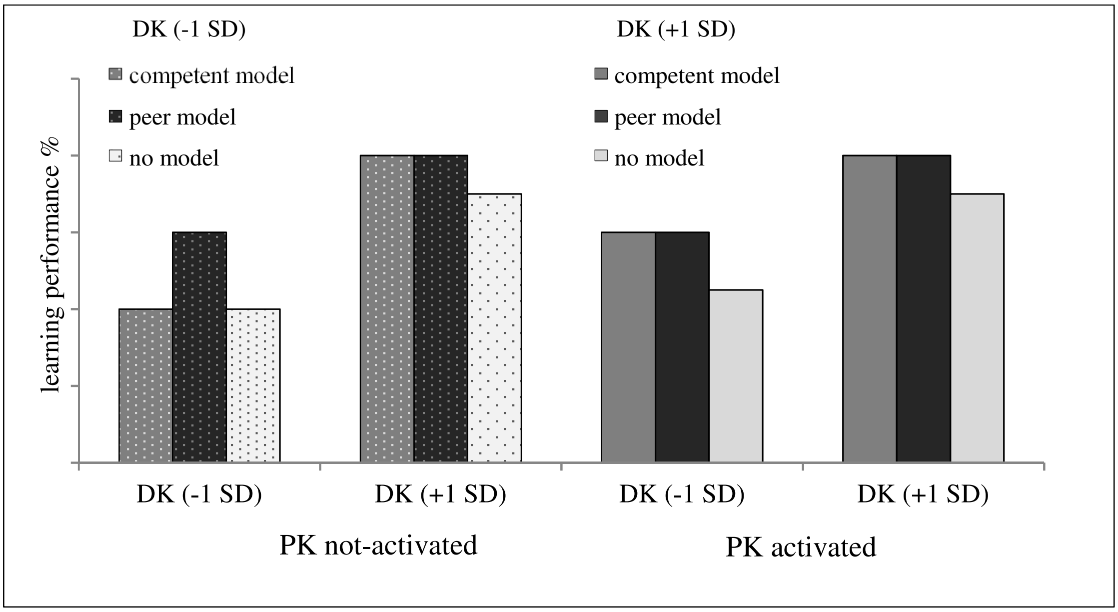


Figure 9: Hypothesized relation between modeling condition, prior knowledge condition (PK) and domain knowledge (DK) against the backdrop of the model-observer similarity hypothesis.

7.2 Method

7.2.1 Participants and design

One hundred and eighty students from a German university took part in the experiment. The number of participants was determined by conducting a power analysis using G*Power 3.1 (Erdfelder et al., 2009) with the effect size *adjusted* R^2 = .10 (derived from Experiment 2), a power of .90, and alpha = .05 resulting in a recommended sample size of 174 participants. Students of biology, medicine, or related fields were excluded a priori from participating due to the learning content (i.e., mitosis). Six participants had to be excluded from data analyses due to technical and language problems. The remaining 174 students (122 female; *M* = 22.48 years, *SD* = 3.62) were enrolled in different university courses.

The design of the experiment was a 2 x 3 between-subjects design with domain-relevant prior knowledge (PK activated vs. PK not-activated) and modeling condition (competent vs. peer vs. control) as independent variables. Participants were randomly assigned to one of six conditions: PK not-activated / competent-model condition (n = 29), PK not-activated / peer-model condition (n = 29), PK not-activated / control condition (n = 30), PK activated / competent-model condition (n = 28), PK activated / peer-model condition (n = 29), PK activated / control condition (n = 29). Participation was voluntary and reimbursed with either 10 Euro or course credits.

7.2.2 Materials

Activation of domain-relevant prior knowledge (PK). To activate PK, half of the participants received domain-relevant information before the learning phase. The other half received domain-irrelevant information instead. Participants in the condition with 'PK activated' received an explanatory text about cell biology with information regarding the history of its scientific discovery, description of cell processes and relevance for daily life. Importantly, the text did not contain the same information as conveyed during the learning phase (overall text length: 396 words). Based on the German curriculum for biology courses, which includes the learning topic 'cell division', we assumed that all participants had some prior knowledge that could be activated. Participants in the condition 'PK not-activated' received an explanatory text about the Mayan people with information about their history, culture and religion (overall text length: 400 words).

Learning material. The material consisted of an expository illustrated text that described relevant biological processes and principles on which mitosis is based (cf. Krebs, Schüler, & Scheiter et al., 2019; Scheiter et al., 2017). Variants of this learning material had been used in previous studies, indicating that students are able to successfully learn from it regarding a variety of outcome measures (e.g., Krebs et al., 2019; Scheiter et al., 2017; Schüler et al., 2013; Stalbovs et al., 2015). The learning content was distributed across 11 pages. The text was divided in semantically meaningful paragraphs with an overall length of 1,113 words. Text lengths per page varied between 44 and 127 words. On each page, the text was accompanied by a static schematic picture on the right-hand side of the respective page. Text and picture were complementary. The text described the processes during mitosis on a more abstract level and the pictures provided additional visual-spatial information about cell structures and processes during mitosis. To build a comprehensive mental model of the content, participants needed to process the text information as well as the picture in-formation.

For the present study, the learning material was divided into two parts (cf., Krebs et al., 2019; Scheiter et al., 2017). Part 1 of the learning material consisted of four pages. In Part 1, the participants received basic information about important cell structures, chromosomes and the concept of DNA. Information from Part 1 was necessary to understand Part 2 of the learning material. Part 2 of the learning material consisted of seven pages. These seven pages contained more detailed information on the basic concepts relevant to mitosis as well as on the phases of mitosis. First, participants were provided with a short overview about the whole process during mitosis. On the next page, the duplication of chromatin fibers and their development into chromosomes were described. Subsequently, the participants were provided with a description of the development of the mitotic spindle, the function of the equatorial plane, and the separation of sister chromatids. The final page referred to the segregation of daughter cells with genetically identical material. All learning outcome measures referred only to Part 2 of the learning material.

For participants in conditions without modeling, Part 1 of the learning material was displayed without external visual guidance via EMME. For participants in conditions with modeling, on Part 1 of the learning material effective multimedia processing strategies were displayed via EMME. The multimedia processing strategies displayed in the EMME videos had been derived from the literature and the same EMME had been used in previous studies (cf., Krebs et al., 2019; Scheiter et al., 2017). In order to generate the EMME, the model (a student research assistant) was instructed on how to process the learning material and behave didactically by demonstrating each process as explicitly as possible (Krebs et al., 2019; Mason et al., 2016; Mason et al., 2017; Scheiter et al., 2017). First, the model read the title of the respective page and inspected the picture on the right-hand side of the page (construction of pictorial scaffold). This was followed by reading the text carefully and looking at corresponding elements of the picture (selection of relevant words and picture elements; text organization; picture organization). During reading the model switched between text and corresponding elements in the picture (integration). Before moving to the next page, the model took a final look at the picture (final picture inspection) and in some cases reread a text section (reaction to comprehension problems). A white spotlight on the otherwise grey shaded page represented a gaze fixation. The overall duration of the EMME videos was 382 seconds and varied between 52 to 115 seconds per page (M = 95.5 s).

7.2.3 Measures

Participants' domain knowledge and social comparison orientation were assessed be-fore the activation of learners' prior knowledge and the learning phase. Posttest performance was measured as dependent variable. Furthermore, we examined participants' judgement of learning before as well as participants' interest and motivation after the posttest. All variables were recorded using the web-based survey software tool platform Qualtrics (http://www.qualtrics.com).

Domain knowledge. Participants' domain knowledge was assessed using a test of participants' prerequisite knowledge on cell division and a test of general scientific literacy. Participants' prerequisite knowledge was assessed with 15 multiple-choice items with four alternatives and one correct answer. The items included questions about cell elements, genetics and mitosis (e.g., 'What are microtubules?' answer: 'Components of the spindle fibres'). For each correct answer participants received one point. The maximum total score for prerequisite knowledge was 15 points. Cronbach's alpha was with a value of α = .43 rather low, which is likely to be due to participants partly guessing.

Participants' scientific literacy was measured by 24 items from the Life Sciences Scale of the Basic Scientific Literacy Test (Laugksch & Spargo, 1996). For each item participants had to rate statements about scientific processes or interrelationships between scientific concepts as either 'correct', 'incorrect', or 'unknown' (e.g., 'The elements that form the molecules of living beings are continuously recycled.' answer: 'correct'). For each correct answer participants received one point. The maximum total score for scientific literacy was 24 points. Cronbach's alpha was α = .68.

A significant correlation of both measures with posttest performance (all *p*s < .05) indicated that both measures captured knowledge components that were relevant to the learning domain. Both measures were also significantly correlated with each other ($r = .31$, $p < .01$). As in Experiment 2, a combined measure of both measures was calculated referred to as domain knowledge using the sum of the z-standardized scientific literacy and prior knowledge scores. Cronbach's alpha of the combined measure was α = .44.

Social comparison orientation. Social comparison orientation was assessed by 11 items of the Iowa-Netherlands Comparison Orientation Scale (INCOM; Gibbons, & Buunk, 1999). For each item participants had to state on a 5- point Likert scale ('I disagree strongly' – 'I agree strongly') if they tend to compare themselves with others (e.g., 'If I want to find out how well I have done something, I compare what I have done with how others have done'). Participants' social comparison orientation score was calculated by summing the responses to each question. A higher score on the INCOM scale indicates a higher tendency to social comparison behaviors. Cronbach's alpha was α= .80.

Learning outcome measures. Posttest performance measure comprised participants' recall and comprehension performance. To assess recall and comprehension performance, text- and picture-based multiple-choice items as well as text- and picture-based forced-choice verification items were used. The multiple-choice items had four alternatives and one correct answer (e.g., 'What is not true about mitosis?' answer: 'Cytokinesis is the division of the cell nucleus.'). For the forced-choice verification items participants had to state if the presented statements or pictures were either true or false (e.g., 'The kinetochore check whether the chromosomes are pulled from both sides with the same force.' answer: 'incorrect'). Correct answers were given one point and incorrect answers zero points. For both measures the percentage of correct answers was calculated. Recall performance was assessed with 3 multiple-choice items and 17 forced-choice verification items (Cronbach's alpha α = .53). Comprehension performance was assessed with 9 multiple-choice items and 4 forced-choice verification items (Cronbach's alpha α = .45). There was no time limit for answering the posttest. All posttest items referred only to Part 2 of the learning material for which no EMME had been displayed to avoid direct influence of the external guidance.

Further Measures. Following measures were assessed for exploratory reasons in the experiment but were not included in the analyses: participants' judgement of learning (JOL), interest and overall motivation, and intrinsic motivation (in the EMME-conditions). JOL was assessed after the learning phase. Participants rated on a zero to 100 scale how confident they were that they could answer the following questions about cell division based on the knowledge they gained in the learning phase (c.f., Schleinschok, Eitel, & Scheiter, 2017). For the JOL score participants' ratings as percent value were used. Higher percentage scores indicated a higher confidence in their success. For the correctness of their judgment the difference between participants' overall test score and their JOL score was computed (c.f., Schleinschok et al., 2017). Participants' interest and motivation were assessed with 10 items after the posttest (e.g., Rotgans & Schmidt, 2011; 2014; Tanaka & Murayama; 2014). Participants rated on a five-point scale (doesn't apply at all– applies completely) their states during the learning phase (i.e., 'I had fun during the learning phase'), their interest in the learning content (i.e., 'I want to know more about the learning content'), the learning process (i.e., 'I was focused during the learning phase'), and the effort they put in the learning process (i.e., 'I made an effort to understand the learning content'). For the score an averaged sum of the ratings was used. Cronbach's alpha for the interest and motivation scale was α= .88. Participants' intrinsic motivation regarding the EMME-videos was only assessed in the EMME conditions using 10 items based on items of the intrinsic motivation short scale (KIM; Wilde, Bätz, Kovaleva, Urhahne, 2009). Here, participants should state on a five-point scale (doesn't apply at all– applies completely) which experiences they made with the EMME-videos (i.e., if they had fun watching the EMME-videos, if watching the EMME-videos was pleasant for them). For the motivation score an averaged sum of the ratings was used. Cronbach's alpha for this scale was α= .79. Means and standard deviations for these measures can be found in Table 8 in the Appendix.

7.2.4 Procedure

Participants were tested in small groups of two to five participants. Each participant was seated individually and was randomly assigned to one of six conditions. At the beginning, they received paper-based written information on the experimental procedures and signed a consent form. Subsequently, they answered computer-based demographic questions, the test on prior knowledge regarding cell division, the scientific literacy test, and the questionnaire on social-comparison orientation. Afterwards the prior knowledge activation took place. Participants were asked to read the upcoming information carefully and were instructed as follows: ‘Please memorize important information from the text in a way that you are able to provide this information to a good friend later’. They were also informed that after four minutes they would be automatically forwarded to the next page. In the following, participants in the conditions with PK activated received information about cell biology. Participants in the conditions with PK not-activated received information about the Mayan people. After four minutes, participants were automatically forwarded to the next page, where they were asked to recall the information by describing the newly learned content to a good friend in 4 minutes time. Then, participants received the instruction for the learning phase. All participants were informed that during the upcoming learning phase they would be able to learn at their own pace and proceed to the next page after the word ‘next’ appeared. Furthermore, they were instructed that they would not be able to go back to previous pages. Moreover, participants in the EMME conditions were informed that they would see short videos with a learner’s eye movements. Thereby, the eyes of the learner would be illustrated by a white spotlight on a grey shaded page with the size of the spotlight illustrating the learner’s fixation time (i.e., larger spots illustrating longer fixation times).

Additionally, participants in the ‘competent-model’ conditions were informed as follows: ‘On the first four pages of the learning material you will see the recorded eye movements of a very successful learner. This learner used learning strategies that are particularly effective for learning. Accordingly, this learner performed very well in the subsequent knowledge test.’ In contrast, participants in the ‘peer-model’ conditions were informed as follows: ‘On the first four pages of the learning material you will see the recorded eye movements of a learner who participated earlier in the experiment’. The EMME themselves were, however, identical in all EMME conditions. After the instruction, participants in the EMME conditions watched the EMME-videos on the first four pages (i.e., Part 1) of the learning material. Thereafter, they were shown the first four pages of the learning material again without EMME to give them the opportunity to study the material again at their own pace. For participants in the conditions without EMME, the first four pages (Part 1) of the learning phase were displayed directly without external guidance.

After Part 1 of the learning material, participants entered Part 2 of the learning material, which was identical for all conditions. To ensure a minimum learning time each page was displayed for 50 seconds before participants could decide on their own to continue by pressing the space bar. After the learning phase, participants' posttest performance was assessed. Finally, participants were debriefed and paid. Each session lasted about 60 minutes.

7.2.5 Data analyses

To investigate the effects of PK activation and alleged model competence on the effectiveness of EMME and the influence of learner' social comparison orientation and domain knowledge results were analyzed in different steps.

First, participants' social comparison orientation and domain knowledge were analyzed to determine whether the experimental conditions differed regarding these characteristics. Subsequently, to analyze the effects of modeling condition and PK activation on learning outcome and their potential moderation by participants' social comparison orientation and / or domain knowledge, multiple regression analyses with two continuous, a dichotomous predictor and a polytomous predictor were conducted.

For reporting overall effects, unweighted effect coding was used for the polytomous predictor (modeling condition) and the dichotomous predictor (PK activation). The modeling condition was coded [0.5 0 -0.5] and [0 0.5 -0.5] for the competent-model condition, the peer-model condition and the control condition, respectively. The PK activation condition was coded [0.5 -0.5] for the conditions domain-relevant knowledge activation and domain-irrelevant knowledge activation, respectively. The obtained variables were furthermore multiplied with the z-standardized domain knowledge score and the z-standardized social comparison orientation score to compute the respective interaction terms.

For comparing the effects of the respective conditions, the modeling condition was dummy coded with one reference category (coding [1 0 0] and [0 1 0] for the competent-model condition, the peer-model and the control condition, respectively). The prior knowledge activation condition was dummy coded (coding [1 0] and [0 1] for the PK activated condition and the PK not-activated condition, respectively). To obtain the respective interaction terms, these variables were also multiplied with the z-standardized domain knowledge score and the z-standardized social comparison orientation score.

For analyzing the effects on participants' recall and comprehension performance, multiple regression analyses with the respective coded modeling conditions, PK activation conditions, the z-standardized domain knowledge score, the z-standardized social comparison orientation score, as well as the generated interaction terms for all possible interactions as predictors were conducted (Aiken, West, & Reno, 1991). To follow up on significant three-way interactions, complex slope analyses were conducted probing significant two-way interactions of modeling condition and domain knowledge / or social comparison orientation for learners with either activated PK vs. not-activated PK. To follow further up on significant interaction terms simple slope analyses were conducted to estimate the size of the effect of the modeling condition at different levels of participants' domain knowledge and / or social comparison orientation (Aiken et al., 1991). To determine the effect of the experimental conditions for lower values, the effect was estimated at – 1 SD relative to the mean of the respective continuous variable. For higher values, the effect was estimated at + 1 SD relative to the mean of the respective continuous variable.

For all statistical analyses the α level was set to $\alpha = .05$. R^2 values were reported as measure of effect size.

7.3 Results

To determine whether there were a priori differences between experimental conditions regarding participants' domain knowledge, social comparison orientation, interest, motivation or learning time, regression analyses with conditions as predictors were conducted. There were no significant differences among conditions with all *p*s > .05. Further results for interest, motivation or learning time are not included in the results section. Table 6 shows the correlations among all measured variables as well as means and standard deviations at baseline.

Table 6. Correlations Among All Measured Variables as well as Means and Standard Deviations at Baseline

Measure	**1[a]**	**2[a]**	**3[b]**	**4[b]**
1. Recall[a]	-			
2. Comprehension[a]	$.382^{***}$	-		
3. Domain knowledge[b]	$.260^{**}$	$.255^{**}$	-	
4. Social comparison orientation[b]	-.029	-.008	.069	-
M	68.19	52.56	0.00	0.00
SD	13.48	15.54	1.00	1.00

Note. *N* = 174. Since domain knowledge and social comparison orientation were z-standardized, the mean value was 0 and the standard deviation 1 for both measures.
[a]Percentage correct. [b]z-standardized.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

To test the effects of modeling condition and PK activation on learning outcome as well as their potential moderation by learners' prior knowledge and social comparison orientation, we conducted two multiple regression analyses with recall performance and comprehension performance as dependent variables, respectively. Modeling condition (competent model vs. peer model vs. no model), domain-relevant prior knowledge (activated vs. not-activated), the z-standardized domain knowledge score and the z-standardized social comparison orientation score as well as the interaction terms for the respective two-way interactions and three-way interactions were entered as independent variables into the analyses. To keep the analyses as parsimonious as possible, interactions comprising both z-standardized continuous variables were excluded from the regression model, since they are difficult to interpret and were not backed up by hypotheses.

For recall performance, there was a significant effect of domain knowledge, $F(1, 156) = 8.12$, $p = .005$, *adj.* $R^2 = .08$, indicating that participants with relatively high domain knowledge ($M = 71.14$ %, $SE = 1.45$) showed a better recall performance than participants with relatively low domain knowledge ($M = 65.18$ %, $SE = 1.46$). Furthermore, there was a significant effect of modeling condition, $F(2, 156) = 4.54$, $p = .012$, *adj.* $R^2 = .08$. Pairwise comparisons revealed that participants in the competent-model condition, $b = 6.95$, $SE = 2.54$, $\beta = .242$, $p = .007$, 95% CI [1.94, 11.95], as well as in the peer-model condition, $b = 6.09$, $SE = 2.48$, $\beta = .213$, $p = .015$, 95% CI [1.19, 10.98] performed better than participants without EMME. There was, however, no significant difference between both EMME conditions, $b = -0.86$, $SE = 2.51$, $\beta = -.030$, p = .732, 95% CI [-5.82, 4.10] (see Table 7). Results showed no effect of prior knowledge activation, $F < 1$, or social comparison orientation, $F < 1$. None of the two-way interactions were significant (all Fs < 1, except for the interaction between modeling condition and social comparison orientation, $F(2, 156) = 1.52$, $p = .221$, *adj.* $R^2 = .08$). Furthermore, none of the three-way interactions were significant (all Fs < 1).

For comprehension performance, results revealed a significant effect of domain knowledge, $F(1, 156) = 8.38$, $p = .004$, *adj.* $R^2 = .11$, indicating that participants with relatively high domain knowledge ($M = 56.30$ %, $SE = 1.66$) performed significantly better than participants with relatively low domain knowledge ($M = 49.46$ %, $SE = 1.65$). There was no effect of PK activation condition, $F(1, 156) = 2.71$, $p = .102$, *adj.* $R^2 = .11$, or social comparison orientation, $F < 1$. Furthermore, results showed a significant main effect of modeling condition, $F(2, 156) = 3.46$, $p = .034$, *adj.* $R^2 = .11$. Pairwise comparisons revealed that participants in the competent-model condition ($M = 54.82$ %, $SE = 2.05$), $b = 6.22$, $SE = 2.86$, $\beta = .188$, $p = .031$, 95% CI [0.56, 11.88], as well as in the peer-model condition ($M = 55.23$, $SE = 1.96$), $b = 6.64$, $SE = 2.80$, $\beta = .202$, $p = .019$, 95% CI [1.11, 12.17], performed better than participants without EMME ($M = 48.59$, $SE = 2.00$). There was no significant difference between the two EMME conditions, $b = 0.28$, $SE = 2.88$, $\beta = .009$, $p = .923$, 95% CI [-5.42, 5.98].

In addition, results indicated also a significant interaction between modeling condition and social comparison orientation, $F(2, 156) = 3.58$, $p = .030$, *adj.* $R^2 = .11$. To investigate the effect of modeling condition at different levels of social comparison orientation, simple slope analyses were conducted. The omnibus effect of modeling condition for participants with relatively high social comparison orientation failed to be statistically significant, $F(2, 156) = 3.03$, $p = .051$, *adj.* $R^2 = .11$. There was, however, a significant omnibus effect of modeling condition for participants with relatively low social comparison orientation, $F(2, 156) = 3.63$, $p = .029$, *adj.* $R^2 = .11$. Results indicated that participants in the peer-model condition performed significantly better than participants in the control condition, $b = 9.94$, $SE = 3.84$, $\beta = .302$, $p = .011$, but only marginally better than participants in the competent-model condition, $b = 7.58$, $SE = 4.06$, $\beta = .230$, $p = .064$. Participants in the competent-model condition did not perform better than participants in the control condition, $b = 2.36$, $SE = 4.10$, $\beta = .071$, $p = .567$. None of the other two-way interactions were significant (all *F*s < 1, except for the interaction between PK activation condition and social comparison orientation, $F(1, 156) = 1.51$, $p = .221$, *adj.* $R^2 = .11$). Moreover, there were no significant three-way interactions (all *F*s < 1, except for the three-way interaction between modeling condition, PK activation condition and domain knowledge, $F(2, 156) = 2.50$, $p = .085$, *adj.* $R^2 = .11$).

Table 7. Means and Standard Errors for the Learning Performance Measures as a Function of Modeling Condition

	Competent model ($n = 57$)		**Peer model ($n = 58$)**		**Control ($n = 59$)**	
Measure	*M*	*SE*	*M*	*SE*	*M*	*SE*
Recall (in %)	70.76	1.82	69.90	1.73	63.82	1.77
Comprehension (in %)	54.82	2.05	55.23	1.96	48.59	2.00

7.4 Discussion

The aim of Experiment 3 was to investigate the relative influence of cognitive and social factors and a possible interaction between them on the effectiveness of EMME. For one, it was attempted to replicate the findings from Experiment 2 regarding the influence of (alleged) model-observer similarity on the effectiveness of EMME indicating that only learners with lower domain knowledge benefit from (alleged) model-similarity. For another, it was of interest whether prior knowledge activation would enable learners with lower prior knowledge to benefit from

EMME independently of (alleged) model-observer similarity. Furthermore, it was investigated whether differences in learners' social comparison orientation would explain differences in learners' ability to benefit from EMME.

The results of Experiment 3 reflected the assumptions that for learners with higher domain knowledge neither model-observer similarity nor prior knowledge activation influence the beneficial effect of EMME presentation (Hypotheses 1 and 2). With regard to model-observer similarity these findings are in line with Experiment 2 and further underlines the premise that social factors (i.e., information about the model's competence) are less decisive for learners with higher domain knowledge. The results indicating that learners with higher domain knowledge did not benefit from prior knowledge activation were also in line with the hypotheses. For learners with higher domain knowledge it was assumed that they already had more elaborate cognitive representations of learning content into which they could more easily integrate (new) information.

For learners with lower domain knowledge, however, it was assumed in line with Wetzels et al. (2011) that they would benefit from prior knowledge activation, because they would be provided with a relevant context which would help them to integrate new information. Further, it was assumed that this would enable learners with lower domain knowledge to benefit from EMME even if the (alleged) model-observer similarity is low, because providing them with a relevant context would reduce the cognitive demands with regard to the learning content. In contrast to these assumptions, results revealed no beneficial effect of prior knowledge activation for learners with lower domain knowledge, thereby confirming neither Hypotheses 4 nor 5.

Moreover, contrary to the hypotheses and previous findings from Experiment 2, results of Experiment 3 indicated no influence of (alleged) model-observer similarity for learners with lower domain knowledge. That is, in Experiment 3 all learners (irrespective of their prior knowledge) performed better when learning with EMME than without, an effect that was independent of the way the model's competence level had been characterized (not confirming Hypothesis 3). Thus, Experiment 3 was not able to replicate the findings from Experiment 2 that learners with lower prior knowledge only benefit from EMME when the (alleged) model-observer similarity is high.

Taken together, in contrast to previous studies, Experiment 3 did not indicate that only learners with more domain knowledge could benefit from EMME (Scheiter et al., 2017), nor was there evidence of an influence of model-observer similarity on the effectiveness of EMME as in Experiment 2. Importantly, however, the results of Experiment 3 supported previous research that EMME are a beneficial instructional tool for multimedia learning (e.g., Mason et al., 2016; Mason et al., 2017). Results revealed that overall learners with instructional support by EMME outperformed learners without instructional support with regard to their recall as well as comprehension performance – regardless of learners' level of domain knowledge, prior knowledge activation, and model-observer similarity.

With view to a possible influence of learners' social comparison orientation, results indicated an interaction between alleged model competence and learners' social comparison orientation on learners' comprehension performance. For learners with higher social comparison orientation, results revealed only a marginal influence of (alleged) model competence. Moreover, although learners with a lower social comparison orientation are less willing to compare themselves with others from the outset and thus the alleged competence of the model should not actually have an influence on learners' comprehension performance, the results of Experiment 3 pointed to the opposite. That is, learners with lower social comparison orientation only benefitted from EMME for their comprehension performance when they received the instruction that the model was a 'peer learner'. It is possible that when presenting a 'competent' model, a potential comparison situation was more salient, and this automatically triggered internal defense mechanisms for these learners in order to avoid a comparison with this model. Maybe the 'peer' model was perceived as less threatening in this direction, so they were actually able to benefit from the model. Yet, these are only speculations referring to results from exploratory analyses that require further investigation in future studies.

7.4.1 Limitations and implications for future studies

Since activating learners' prior knowledge by providing them with a domain-relevant prior knowledge background before the learning phase was not helpful, it is possible that prior knowledge activation may not be the means of choice to promote the effectiveness of EMME - at least not the type of prior knowledge activation that was used in Experiment 3. It is important, however, that Experiment 3 was able to demonstrate that EMME are suitable for multimedia learning not only for younger children (e.g., Mason et al., 2016), but also for adult learners (university students) - regard-less of their domain knowledge or (alleged) model-observer similarity.

Yet, there are certain limitations to the results of Experiment 3. On the one hand, all participants completed a pretest at the beginning of the experiment. The pretest itself contained domain-relevant prior knowledge questions that already could have activated their prior knowledge to a certain degree. On the other hand, all participants were randomly assigned to the activation conditions without taking their actual prior knowledge into account. Therefore, it is a possibility that learners' existing prior knowledge collided with the intended prior knowledge activation. Furthermore, it is also possible that the method was not appropriate to activate learners' prior knowledge and that learners with lower domain knowledge were in turn not able to form a domain-relevant knowledge context before the learning phase and therefore could not benefit from the activation intervention. Furthermore, contrary to Experiment 2, learners' information processing was not assessed in Experiment 3. Therefore, it could not be investigated whether learners' prior knowledge, social comparison orientation or the respective model descriptions influence learners' information processing, albeit not their learning performance.

Taking these limitations into account, further research is needed to follow up on the question of the influence of individual differences such as prior knowledge and / or social comparison orientation, as well as on the influence of social cues such as model-observer similarity on the effectiveness of EMME.

7.4.2 Conclusions

In sum, results of Experiment 3 support the assumption that using EMME can foster multimedia learning, thereby further confirming EMME's function as an effective instructional tool. Although the findings do not provide direct evidence for the influence of social factors such as alleged model-observer similarity they provide first indications that individual differences such as social comparison orientation can influence the effect of model instruction on the effectiveness of EMME. In addition, results show that the role of learners' prior knowledge remains unclear and requires further research. Taken together, further research is needed to follow up on the influence of individual factors as well as social cues on the effectiveness of EMME as instructional tool for multimedia learning.

8 General Discussion

8.1 Summary of Results

Multimedia material is ubiquitous in the educational context today, whether in textbooks, worksheets, or online learning environments. It is therefore more necessary than ever for learners to acquire and use strategies that allow them to adequately process multimedia materials (e.g., illustrated text). EMME as a form of gaze-based learner-oriented intervention provide learners with effective processing strategies that are expressed through the eye movements of an instructed model. By observing when, where, and how long the model has looked at certain parts of the task material, learners may deduce the underlying processing strategies that in turn help them to perform a similar or identical task (Veenman, 2011). Accordingly, previous research indicated that EMME are generally effective for multimedia learning (Mason, Pluchino et al., 2015, Mason et al. 2016, 2017; Scheiter et al., 2017). However, the question of possible mechanisms underlying the effectiveness of EMME for multimedia learning has remained open so far.

For the present thesis, I differentiated between three perspectives regarding possible mechanisms underlying the effectiveness of EMME: The perceptual learning perspective, the (meta-) cognitive perspective, and the social perspective. Whereas theoretical claims have referred to all three perspectives (cf. Van Gog & Rummel, 2010), empirical research so far has relied mainly on mechanisms and factors that can be derived from a perceptual learning and / or (meta) cognitive perspective to explain positive effects of EMME (e.g., Jarodzka et al., 2013; Mason, Pluchino, et al. 2015; Mason et al., 2016, 2017; Scheiter et al., 2017). As a consequence, the role of social mechanisms in the effectiveness of EMME for multimedia learning is hitherto unknown. Therefore, the overarching research question of this thesis was whether the consideration of social mechanisms and factors provides an additional explanation for the effectiveness of EMME that goes beyond a mere perceptual and / or cognitive explanation.

Two different approaches were used to investigate two research questions that resulted from this overarching research question. The research question that guided the first approach was whether social mechanisms involved in EMME contribute to the effectiveness of EMME beyond the effect of perceptual and (meta-) cognitive mechanisms. Following the first approach, Experiment 1 compared EMME to another attention-guidance tool without social cues (i.e., abstract-cuing videos). If social mechanisms contribute to explaining the effectiveness of EMME for multimedia learning, EMME with a social gaze cue (i.e., human eye movements) and social cues in the introductory text should be more effective than a comparable attention guidance tool with a non-social attention-guiding cue and without social cues in the introductory text. Results of Experiment 1, however, were inconclusive. Unexpectedly, there was no positive effect on information processing and / or learning outcomes for either of the two kinds of instructional support tools, nor were there differences between EMME and the attention-guidance tool without social cues.

The research question that guided the second approach was whether social factors influence the effectiveness of EMME or whether the effectiveness of EMME is solely influenced by perceptual and / or cognitive factors. If social mechanisms underlie the effectiveness of EMME, social factors should moderate the effectiveness of EMME. In contrast, if the effectiveness of EMME is solely based on perceptual and / or (meta-) cognitive mechanisms, social factors should not influence their effectiveness. To investigate these hypotheses underlying the second approach, the three experiments involved either an experimental manipulation of the salience or an assessment of social factors that might influence learning with EMME.

Experiments 2 and 3 aimed at investigating the question of whether (alleged) model competence and / or (alleged) model-observer similarity influence the effectiveness of EMME for multimedia learning. Experiments 1 and 3 investigated the question of whether social comparison orientation can influence the effectiveness of EMME. In Experiment 2 learners with lower prior knowledge benefitted from EMME, if the model was introduced as a ‘peer’ learner, but not if it was introduced as a ‘competent’ model. Learners with higher prior knowledge did not benefit from EMME. Thus, the findings of this experiment point to the possibility that (alleged) similarity between model competence and learner competence might be one social factor that can influence the effectiveness of EMME. Experiment 3 aimed to replicate Experiment 2 as well as to extend it by addressing further social factors, namely learners’ social comparison orientation. In contrast to Experiment 2, Experiment 3 did not only use learners’ naturally occurring differences in prior knowledge to study its impact on the effectiveness of EMME (related to alleged model competence). Instead prior knowledge was also activated experimentally. Contrary to the findings from Experiment 2, alleged model-observer similarity had no influence on the effectiveness of EMME, neither with regard to the existing prior knowledge, nor with regard to the manipulated prior knowledge. Regarding social comparison orientation, based on the literature one would have expected model competence to have an influence on learning performance only for learners with higher social comparison orientation. However, the results pointed to the opposite: Only learners with lower social comparison orientation benefitted from EMME that allegedly had been obtained from a peer model. For Experiment 1, there was no effect of social comparison orientation on the effectiveness of EMME. To conclude, the results from Experiments 2 and 3 point towards a relevance of social factors for the effectiveness of EMME, but the evidence across the experiments is inconsistent.

In the following I will further discuss the results from the three studies in more detail. To this end, I will first briefly address the question of whether the present experiments were able to confirm that EMME are an effective instructional tool that improves learners' processing of illustrated text and, in turn, learning outcome. Subsequently, I will summarize and discuss the present results' implications for the question of which mechanisms contribute to and which factors moderate the effectiveness of EMME.

8.2 Effectiveness of EMME as Instructional Tool for Multimedia Learning

Whereas most previous EMME-research found that EMME are generally an effective instructional tool for multimedia learning (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; see Table 2), the results of the three experiments conducted in the present dissertation paint a more differentiated picture. While Experiment 1 revealed no effect of EMME (or an abstract attention guidance tool) on learning, Experiments 2 and 3 demonstrated that learners with EMME support used more effective processing strategies (Experiment 2) as well as had better learning outcomes (Experiments 2 and 3). However, especially Experiments 2 and 3 also indicate that there are factors which can moderate the effectiveness of EMME. For one, results of Experiment 2 indicated in line with findings from Scheiter et al. (2017) that cognitive factors such as prior knowledge can play a role. Moreover, across the experiments there were first indications that also social factors (which will be discussed in more detail below) might influence the effectiveness of EMME.

That guiding visual attention to support learners in inferring important information from multimedia material is effective has been shown not only in previous EMME-research (e.g., Mason, Pluchino et al., 2015, Mason et al., 2016, 2017), but also in studies using material-oriented attention-guiding cues (for reviews, see Richter, et al., 2016; Schneider, et al., 2018). Thus, the findings from Experiments 2 and 3 strengthen the assumption that gaze cues may follow at least in part similar mechanisms as other attention-guiding cues that are designed by a task or instructional expert (e.g., color coding, Jamet, 2014). However, Experiment 1 did not confirm the assumption that eye movements play a privileged role compared to other means of enhancing perceptual learning (e.g., cueing or perceptual grouping, cf. Goldstone et al., 2017).

From a (meta-) cognitive perspective, EMME are not only assumed to guide visual attention but also to foster higher-level cognitive processes (Scheiter et al., 2017). On this basis, previous EMME-research assumes that other mechanisms which may underlie the effectiveness of EMME relate to the way EMME support learners in acquiring effective processing strategies (Mason, Pluchino et al. 2015; Mason et al., 2016) and in regulating existing strategies (Mason et al., 2017; Scheiter et al., 2017). The basic idea from a (meta-) cognitive perspective is that EMME can serve as a strategy training intervention. In contrast to other means of instructional support such as material-oriented interventions (e.g., reducing complexity; cueing/signaling; physically integrating text and pictures) or other learner-oriented interventions such as strategy prompts (Kombartzky et al., 2010; Schlag & Ploetzner, 2011), strategy trainings do not only provide learners with declarative strategy knowledge, but additionally with procedural and conditional knowledge of how, when, and why to apply the respective strategy during a learning task (WWW&H rule; Veenman, 2011). The findings from Experiments 2 and 3 support (in line with findings from Scheiter et al., 2017) the assumption that EMME can be an effective strategy training intervention not only for younger children (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017), but also for adult learners. Moreover, Experiments 2 and 3 replicated the findings by Scheiter et al. (2017) that fostering the use of a specific set of multimedia processing strategies has beneficial effects on multimedia learning. Experiment 2, however, also implies in line with previous findings by Scheiter et al. (2017) that EMME might not be equally effective for all learners, but that, instead, cognitive factors such as prior knowledge can moderate their effectiveness.

Prior knowledge as a potential moderator for the effectiveness of EMME can be derived from a perceptual as well as from a (meta-) cognitive perspective. The fact that it might be important to consider learners' prior knowledge as a potential influencing factor is underlined by previous empirical evidence that prior knowledge does not only influence the effectiveness of observational learning (e.g., Groenendijk et al., 2013a; LeBel et al., 2018), but also learners' ability to process information (e.g., Canham & Hegarty; Gegenfurtner et al., 2011; Van Gog et al., 2015) as well as the effectiveness of instructional support measures (expertise reversal effect; Kalyuga, 2013). Moreover, Scheiter et al. (2017) were the first to reveal that learners' prior knowledge can influence the effectiveness of EMME. Yet, their finding that only learners with higher prior knowledge benefit from EMME was not replicated in the present experiments. Instead, results from Experiment 2 indicated that learners with lower prior knowledge were able to benefit from EMME as well, but only as long (alleged) model-observer similarity was high.

Results of Experiments 2 and 3 suggest that EMME are effective in guiding learners' attention to foster the use of effective (multimedia) processing strategies (e.g., increased picture processing). This is in accordance with the assumption that the effectiveness of EMME both perceptual and (meta-) cognitive mechanisms underlie the effectiveness of EMME for multimedia learning. Hence, these results are in line with the signaling / cuing / attention-guiding principle (Schneider et al., 2018; Richter et al., 2016; Van Gog, 2014). Moreover, they also contribute to the research on the effectiveness of instructional support for multimedia learning, indicating that providing learners with necessary strategy knowledge is beneficial for multimedia learning (e.g., Eitel et al., 2013; Kombartzky et al., 2010; Mason, Pluchino, et al., 2015; Mason, et al., 2016, 2017; Stalbovs et al., 2015).

Nevertheless, results from Experiment 2 also suggest that cognitive factors can moderate the effectiveness of EMME, as not all learners benefitted equally from EMME. This is line with previous research by Scheiter et al. (2017). In contrast to their results, however, results from Experiment 2 showed in line with previous research that learners with lower prior knowledge can also benefit from EMME, but only if the (alleged) similarity with the model in terms of (alleged) competence is high (e.g., Braaksma et al., 2002; Kim, 2007). This finding indicated that not only cognitive factor such as prior knowledge, but also social factors might influence the effectiveness of EMME. These will be discussed in more detail in the next subchapter.

Taken together, the results of the three experiments of this dissertation did not provide a clear picture with regard to the effectiveness of EMME for multimedia learning. Whereas the findings from Experiments 2 and 3 support the assumption that EMME as in instructional tool can foster multimedia learning, the results from Experiment 1 were less promising. Moreover, even though there was further support for previous EMME-research showing that learners' prior domain knowledge can be a potential moderator (Scheiter et al., 2017), this finding was only limited to Experiment 2.

8.3 Social Aspects of Learning with EMME

Despite the fact that earlier EMME-research generally claimed that EMME is based on the of social cognitive learning theory (e.g. Jarodzka et al., 2013; Mason, Pluchino et al., 2015, Mason et al., 2016, 2017), the role of social mechanisms in the overall effectiveness of EMME for multimedia learning has not been empirically investigated until now. With regard to the influence of social factors such as model competence and model-observer similarity in video-based modeling in general, some empirical studies already exist (e.g., Braaksma et al., 2002; Groenendijk et al., 2013b, Hoogerheide et al., 2016; Hoogerheide et al., 2017; Hoogerheide et al., 2018; LeBel et al., 2018; Litchfield et al., 2010). However, the findings of these studies were rather inconclusive. Against this background, this dissertation sought to design experiments that specifically investigate the role of social mechanisms in the effectiveness of EMME as well as a potential influence of social factors on the effectiveness of EMME for multimedia learning.

Experiment 1 tested whether social mechanisms contribute to the effectiveness of EMME by comparing EMME with a non-social attention guidance tool. Previous research indicates that if people ascribe anthropomorphic qualities to abstract cues as those used in the EMME-videos (i.e., spotlights), a social context emerges which, in turn, yields deeper learning (social agency theory; Mayer et al., 2003; Moreno, et al., 2001). However, in Experiment 1, the two instructional tools did not differ in their effectiveness and neither instructional tool proved to yield superior learning outcome compared to a control condition. Hence, Experiment 1 does not allow to draw definite conclusions about the role of social mechanisms in the effectiveness of EMME.

The social cognitive learning theory (Bandura, 1986), however, provide a more indirect way of testing whether social aspects play a role in the effectiveness of EMME. According to this theory there are three kinds of factors that influence the effectiveness of modeling examples: the characteristics of the learner (e.g., cognitive capabilities, social comparative biases), the characteristics of the model (e.g., model competence), and the characteristics of the modeled event (e.g., salience of relevant aspects). Both Experiments 2 and 3 focused on the interplay between model and learner characteristics in form of (alleged) model-observer similarity and found, similar to previous research, mixed effects (e.g., Braaksma et al., 2002; Groenendijk et al., 2013b, Hoogerheide et al., 2016, 2018; LeBel et al., 2018; Litchfield et al., 2010). Whereas results from Experiment 2 indicated an influence of (alleged) model-observer similarity on the effectiveness of EMME, this finding could not be replicated in Experiment 3. The findings from Experiment 2 are in line with the assumption that learners' prior knowledge has an important role in learning from modeling examples (e.g., Bandura, 1986; Van Gog & Rummel, 2010), indicating that the influence of social cues could be particularly important when learners are cognitively more challenged with a task, for example if learners have lower prior knowledge.

With regard to a potential influence of (alleged) model competence, results of both experiments clearly indicated that (alleged) model competence had no direct effect on the effectiveness of EMME. This finding contradicts the assumption that experts or persons with higher (alleged) competence are better models per se (Bandura, 1986) and are also in line with the findings from Litchfield et al. (2010), indicating no influence of the model's level of expertise on diagnostic performance. In contrast to Experiment 2, however, Litchfield et al. (2010) found also no effect of model-observer similarity. However, importantly, there are two major differences to the study by Litchfield et al. (2010). First, in Experiments 2 and 3 learners received explicit information about the model in the introductory text to manipulate (alleged) model competence. In contrast, in the study by Litchfield et al. (2010), participants received no information about the model's expertise or performance Second, the EMME-videos in Experiments 2 and 3 were identical, only the introductory text varied between the conditions. In contrast, in the study by Litchfield et al. (2010) the model (expert vs. novice), and thus the EMME-video varied between the conditions. Moreover, Litchfield et al. (2010) investigated the use of EMME for supporting task performance in a medical field and not for multimedia learning.

A second social factor that was investigated in Experiments 1 and 3 based on Bandura's (1986) assumption that also learner characteristics influence learning from modeling examples, was social comparison orientation (Gibbons & Buunk, 1999). Social comparison orientation should be particularly influential as during learning from modeling examples, learners are encouraged to engage in a social comparison with the model (Hoogerheide et al., 2018). In contrast to model-observer similarity, social-comparison orientation was not manipulated in the experiments. It was assumed that learners with a higher social comparison orientation are more sensitive to cues indicating that the current situation offers a possibility to compare with someone else and thus be more attentive when presented with a human model. This assumption could not be supported by the results of Experiments 1 and 3. However, results of Experiment 3 indicated an interaction between learners' social comparison orientation and the way the model was depicted in the introductory text (i.e., model competence). This finding is in line with the assumption that modeling examples encourage learners to engage in a social comparison with the model (Hoogerheide et al., 2018; Schunk & Hanson, 1985) and supports the assumption that social factors might play a role in the effectiveness of EMME. However, the results for an influence of a social factor on the effectiveness of EMME were again rather mixed.

In summary, the results of the three experiments were inconclusive with regard to the role of social mechanisms in the effectiveness of EMME. Whereas results from Experiments 2 and 3 gave first indications that social mechanisms might play a role, these findings were rather fragile. In contrast to Experiments 2 and 3, results from Experiment 1 gave no indication that social mechanisms can influence the effectiveness of EMME. I will discuss possible reasons for these inconclusive findings in the following.

As already mentioned in the beginning of this section, results of previous research on the effect of (alleged) model competence and (alleged) model-observer similarity are rather instable across the research on (video) modeling examples in general. Whereas Braaksma et al. (2002) and LeBel et al. (2018) for example found rather beneficial effects of model-observer similarity, other studies did not find an effect of model-observer similarity (e.g., Groenendijk et al., 2013b; Hoogerheide et al., 2017; Litchfield et al., 2010), or even found contrasting results (e.g., Hoogerheide et al., 2016). With regard to EMME, there is up to date only one study by Litchfield et al. (2010) that took model and observer expertise into account when investigating the effectiveness of EMME for supporting radiographers' diagnostic performance. Results of this study indicated an influence of neither model expertise, nor observer expertise, nor an interaction between those two factors on task performance. Importantly, however, learners in this study received no explicit information about the level of the model's expertise or and had to infer this information solely from the model's gaze behavior. This might have been too difficult, as results indicated that both the expert model and the novice model fixated similar areas. Yet, without further information, learners might not have been able to deduce the model's expertise. Hence, there may have been too little social cues to trigger a social comparison process, which in turn may have led to the fact that social factors had no influence in this case.

A second issue for investigating the effects of model-observer similarity or model competence is how (alleged) model competence can be experimentally manipulated, as noted by Hoogerheide (2016). When model-observer similarity is varied experimentally, not only the models differ between conditions (e.g., with regard to age or gender), but often also the content of the modeling examples. This was also the case for the study by Litchfield et al. (2010). In their study, participants in the different conditions received instructional support from different models (expert vs. novice radiographer), as well as different materials (expert's gaze replay vs. novice's gaze replay). Hoogerheide et al. (2016) addressed this problem by varying the (ostensible) age and the expertise of the model, while keeping the content of the modeling examples the same. Results of this study showed no influence of model-observer similarity for age as a model characteristic, indicating that for their sample of younger learners, adult models were more effective than peer models, irrespective of the model's reported expertise. But although the peer as well as the adult models both presented themselves as persons with lower or higher expertise, the quality of the adult model's explanation was rated as higher by the participants – even though the explanations were identical. This finding points to another problem when investigating the effect of (alleged) model competence and model-observer similarity: (Alleged) model characteristics are often interlinked, such as age and expertise (Hoogerheide et al., 2018). Thus, it is usually difficult to separate them for (video) modeling examples which, in turn, might also contribute to the mixed findings regarding the influence of (alleged) model-observer similarity.

For (video) modeling examples that do not show the model at all or only partially (e.g., the hands; Fiorella & Mayer, 2016), model characteristics such as attractiveness are less ostensible. However, learners are still able to gain information about the model not only from the provided information (e.g., introductory text, short self-presentation), but also from other cues (e.g., voice, male/female hands). For investigating the effect of (alleged) model-observer similarity, it should therefore be beneficial to reduce those outside cues as much as possible. Hence, the EMME-videos in the present dissertation included no verbal comments. Moreover, in accordance with Hoogerheide et al. (2016), the content of modeling examples was kept equal across the conditions by using the same EMME-videos for learners with instructional support, and only varying the information about the model's competence in the introductory text. However, results for the influence of model-observer similarity for Experiments 2 and 3 were still mixed, even though the information in the introductory text was kept similar and the EMME-videos in both experiments were identical. These mixed results could relate to the fact that people might have to first engage in social comparison processes (i.e., compare themselves with an (hypothetical) other person), in order for (alleged) similarity to be perceived at all. There are three possible explanations for why learners in the present experiments might have failed to engage in social comparison processes in the experiments of the present dissertation.

The first explanation relates to the subtlety of the manipulations employed. In contrast to other (video) modeling examples (e.g., Hoogerheide et al., 2016, 2018; Van Gog et al., 2014), the manipulation of the (alleged) characteristics of the model in Experiments 2 and 3 was very subtle, as the information about the model containing social cues was solely in one paragraph in the introductory text. This required the participants to carefully process the contents of the introductory text. It is therefore possible that Experiment 3 did not replicate the results from Experiment 2 due to participants processing the information in the introductory text less attentively. Unfortunately, as Experiment 3 was no eye tracking study, there is no data available that might support this assumption.

The second explanation relates to the problem that the laboratory-context of the present experiments might have prevented the activation of a sufficiently strong social context. As in the three experiment the only social information participants in the EMME conditions received was in the introductory text, the actual social richness of the observational learning situation in the EMME-videos was very low. In contrast, other (video) modeling examples provide considerable more social richness either by showing the human model (Fiorella & Mayer, 2016; Hoogerheide et al., 2016, 2018; Van Gog et al., 2014), by providing additional verbal explanations (e.g., Van Gog et al., 2009; Van Marlen et al., 2016, 2018), or by encouraging learners to discuss the contents of the EMME-videos (Salmerón, & Llorens; 2018). Moreover, other studies that investigated the influence of social cues based on the social agency theory (e.g., Töpper, Glaser, & Schwan, 2014; Wang, Li, Mayer, & Liu, 2018), also provided stronger social contexts to prime a social response in the learner, for example showing a (human) agent, or using personalization cues (e.g., “you”). According to Mayer (2014b), this social response increases learners’ active cognitive processing, which in turn results in better learning outcomes. Moreno and Mayer (2004) claim that a social response requires the feeling of a social presence (i.e., feeling of interacting with another social being; Mayer, 2014b). Therefore, it might have been problematic that the social context in the experiments of the present dissertation was mainly based on participants’ imagination. They neither saw the model (e.g., Hoogerheide et al., 2016, 2018; Töpper, Glaser, & Schwan, 2014; Van Gog et al., 2014), nor heard her voice (e.g. Van Marlen et al., 2016, 2018), nor could they observe the model’s real eyes (e.g., Symons et al. 2014).

Findings by Gobel et al. (2018) suggest that this much sociality might not be necessary, as they demonstrated that people differ in their performance even if they only assume that abstract stimuli have a human origin. The authors argued that for visual attention, mainly the social relevance of a cue is important, and not its social appearance. What needs to be noted, however, is that the stimuli used by Gobel et al. (2018) still activated a social context since in the social cue condition, participants were led to believe that the location of the cue was connected to the gaze location of an (unseen) partner. Moreover, the (alleged) interactive task to be accomplished with a partner might have additionally also increased the relevance of the task. It is possible that participants in the three experiments of the present dissertation might have perceived the (learning) task to be less personally relevant for two reasons. For one, since students of biology, medicine, or related fields were excluded from participating, the learning content itself (i.e., mitosis) was not relevant for the learners (e.g., for an exam). For another, the learning task required no (alleged) interaction with another person.

Given that the social context created by the subtle manipulation in the introductory texts was rather weak, a high perceived task relevance might, however, been necessary for (alleged) model-observer similarity and social comparison orientation to come into effect. This idea rests on the assumption that, if participants did not perceive the learning situation to be personally relevant, they might not have been motivated enough to compare themselves to another (fictional) person which, in turn, could have prevented a potential influence of the social factors on the effectiveness of EMME. Against this background, it might be necessary for the social context to be sufficiently activated for social mechanisms and factors to come into play (social agency theory; Mayer et al., 2003; Moreno et al., 2001). If the manipulation in Experiments 1 and 3 failed to sufficiently activate the social context, it might explain why EMME were not effective in Experiment 1, and why the influence of (alleged) model-observer similarity from Experiment 2 could not be replicated in Experiment 3. With regard to this possibility, it might be necessary to find the right balance between providing enough social information to sufficiently activate a social context versus providing too much information about the model so that the effect of (alleged) model competence and / or (alleged) model observer similarity cannot be investigate independently from other model characteristics.

The third explanation relates to the assumption that people differ in their disposition to compare themselves to others as well as in the way they interpret the information they gain during the comparison process (Buunk & Gibbons, 2007; Gibbons & Buunk; 1999). According to Festinger (1956), the theoretical basis for social comparisons is the individual's need for self-knowledge in order to judge and evaluate themselves in comparison to others. For example, in classical experiments with a rank order design, participants are given information about the performance of others and are asked to choose with whom they want to compare their own performance (e.g., Wheeler, 1966). In more recent work, however, the social comparison theory has been related more and more to research on social cognition, which emphasized also the need of investigating cognitive processes during the social comparison process (Buunk & Gibbons, 2007). For Experiments 1 and 3, the possibility was considered that learners with a higher social comparison orientation (Gibbons & Buunk; 1999) might be more inclined to engage in a social comparison with the model as they are more susceptive for social cues. Results of Experiments 1 and 3 for learners' social comparison orientation as potential moderator of the effectiveness of EMME, however, were mixed. But one has to keep in mind that social comparison orientation as a construct comes from a different field of psychological research. Thus, it is possible that the INCOM scale is only conditionally suited to measure what is defined as 'tendency for social comparison orientation' in the present thesis, because the items of the scale refer more to a general context. To my knowledge, there are up to date no other studies on modeling examples that included this scale to assess the influence of learners' social comparison on the effectiveness of modeling examples. Thus, further research is necessary to investigate whether social comparison orientation proves to be useful in the observational learning context.

In summary, the studies conducted in this thesis have not succeeded in conclusively clarifying the question of whether social factors influence the effectiveness of EMME. However, they successfully provided first indications that they can play a role, for example in the form of (alleged) model-observer similarity. In addition, there are initial indications that there may be learner characteristics other than domain knowledge, such as a social comparison orientation, which have the potential to influence the effectiveness of EMME. Overall, the results of the first two studies allow initial conclusions to be drawn that if EMME are framed as a social learning situation, a possible influence of social factors on their effectiveness should be considered.

8.4 Strengths, Limitations, and Future Directions

A strength of the present dissertation is that the research was conducted based on assumptions that were derived from well-established theories and theoretical frameworks from different fields. Thereby, the focus lay on their individual contribution to explaining which mechanisms of action may underlie as well as which factors may influence the effectiveness of EMME for multimedia learning.

Moreover, by considering also theories and frameworks from social psychology such as social comparison, the present thesis offers an innovative contribution to previous EMME-research and further broadens the theoretical perspective on the effectiveness of EMME. Even though there are several theories and theoretical frameworks, including social theories, that contribute to explaining why EMME are effective for multimedia learning, no one has yet attempted to incorporate theories and theoretical frameworks from social psychology. However, including theories and theoretical frameworks from other areas of psychological research also raises challenges, because the question arises to what extent the concepts and the assumptions derived from them can be adapted to another research area. Therefore, a sound theoretical reasoning prior to their inclusion into an experiment is just as important as subsequently reflecting on their effects regarding the results. Including concepts from other research areas can broaden the own theoretical basis but can also contribute to the further advancement of research in the other areas as well as link different research areas more closely.

By separating and including different explanations for the effectiveness of EMME from different theories and theoretical frameworks, the presented research offers a first approach towards the development of a theoretical framework for the effectiveness of EMME for multimedia learning. This may provide the basis for future research to further investigate not only the general effectiveness of EMME, but also specific mechanisms of action and factors arising from different theories and theoretical frameworks. Particularly, previous and current EMME-research highlights the importance of examining the underlying mechanisms for their effectiveness as well as factors that can influence their effectiveness for different learners. An objective of future EMME-research should therefore be the investigation of mechanisms of action, influencing factors and boundary conditions for the effectiveness of EMME.

Furthermore, the presented research contributes to the previous EMME-research by investigating the role of social mechanisms in the effectiveness of EMME. For the first time, it was empirically examined whether the social mechanisms that are usually theoretically assumed to contribute to the effectiveness of EMME (e.g., Jarodzka et al., 2013; Mason, Pluchino et al., 2015, 2016; Salmeron et al., 2017) can also be detected empirically. This contributes to defining mechanisms of action as well as boundary conditions for the effectiveness of EMME. For instance, the presented research provides first empirical indications that under certain conditions, social factors can influence the effectiveness of EMME. Moreover, the presented research also fosters in line with first findings from Scheiter et al. (2017) the assumption that not only younger children (Mason, Pluchino et al., 2015, 2016), but also adult learners (university students) can benefit from EMME that demonstrate a more complex processing strategy for illustrated texts containing several effective multimedia processing strategies. In contrast to earlier research, however, the results also show that the positive effect is not limited to learners with higher prior knowledge (Scheiter et al., 2017). These results may not only contribute to broadening the theoretical framework for the effectiveness of EMME, but also provide further indications for the practical use of EMME as an instructional tool for multimedia learning.

Another strength of the presented research applies to the use of eye tracking as a process-related measure in addition to learning outcome measures in two of the three experiments. This offers the opportunity to take a closer look at the influence of potential mechanisms and factors on the learning outcome, but also on the use of (cognitive) processing strategies. At this point, however, it is important to note that even though eye movements are assumed to reflect cognitive processes (eye-mind assumption, Just & Carpenter, 1980), the relationship between eye movements and cognitive processes is not always unambiguous (Kok & Jarodzka, 2017a). For one, looking at information does not necessarily mean that learners (consciously) process the information (Kok & Jarodzka, 2017b). For another, there is empirical evidence that even if learners change their processing behavior, that does not necessarily translate into their task performance or learning outcome (e.g., Kok, Jarodzka, De Bruin, BinAmir, Robben, Van Merriënboer, 2016; Kok & Jarodzka, 2017b; Ozcelik, Arslan-Ari, Cagiltay, 2010; Scheiter, et al., 2017). Nevertheless, eye tracking can be used as an objective measure of which information has entered the cognitive system and is therefore more likely to be processed (Kok & Jarodzka, 2017b). Against this background, the present research was also able to offer in two of the three experiment further insights into the use of processing strategies during multimedia learning using eye tracking a rather objective measure.

With regard to the practical use of EMME, however, there are certain limitations. First, regarding the generalizability of the findings from the presented research one has to keep in mind that the same learning material was used in all experiments. Therefore, it is possible that the effects that were found might be limited to the domain area (biology), to the topic (mitosis), or even to the specific learning material. With a view to the latter, however, it should be considered that in previous research, different learning materials were used and that this previous research was even able to show transfer effects from one learning material to another (Mason, Pluchino et al., 2015, 2016, Mason et al., 2017). In future studies, however, it would be important to replicate the results of the presented research for other learning materials, especially since the effects in the three studies have proven to be rather unstable, for example regarding the effect of (alleged) model-observer similarity. Another limitation also relates to the external validity of the findings. All three conducted experiments were laboratory experiments with a sample from a restricted population (university students). Hence, the ecological validity of the experiments is also limited in this regard. One objective of future research could therefore be to test the effects outside the laboratory, for example, by testing the effect of various model-descriptions on the effectiveness of EMME in a classroom-setting.

In summary, the innovative and integrative theoretical approach as well as the use of eye-tracking as a process-related measure contribute to the development of a theoretical framework for EMME. However, the empirical results must be interpreted with some caution, as they are not very stable, and the material does not allow conclusions to be drawn about their generalizability.

8.5 Theoretical and Practical Implications

The findings from the present dissertation have theoretical as well as practical implications. In the following I will focus on the theoretical implications first, before continuing with the practical implications.

8.5.1 Theoretical implications

The results from the three experiments have theoretical implications in different respects. First, they add to the research on multimedia learning. By showing that supporting learners in using multimedia processing strategies that can be derived from multimedia learning theories (e.g., integration of text and picture elements; CTML; Mayer, 2009, 2014a; ITCP; Schnotz, 2002;

2014), the results from Experiments 2 and 3 further underline the relevance of these strategies for multimedia learning. Furthermore, the findings from the dissertation also strengthens theoretical assumptions from theories such as observational learning (Bandura, 1986) or example-based learning (Renkl, 2011; Renkl, 2014; Van Gog, Paas, & al., 2009; Van Gog & Rummel, 2010; Van Gog, Verveer, & Verveer, 2014), by showing that offering learners the possibility to observe in the EMME-videos a human(-like) model applying effective solution procedures supports them in applying these strategies themselves.

Most importantly, the present thesis also contributes to theory development regarding the mechanism underlying the effectiveness of EMME for multimedia learning. More specifically, the present thesis attempted to disentangle the various mechanisms resulting from the broad theoretical foundation EMME are built upon. For this purpose, three theoretical perspectives were introduced that shed light on the effectiveness of EMME from different angles. Against the broad theoretical background consisting of different theories and theoretical frameworks, it has so far only been possible to draw conclusions about the general effectiveness of EMME. To either investigate the influence of or to draw conclusions about specific mechanisms and / or individual factors that can be derived from these mechanisms, however, it is necessary to separate the various explanations from different theories. Even if the theoretical work in the present thesis represents only a first attempt to separate the different explanations and does not yet result in a consistent theoretical model for the effectiveness of EMME, it is nevertheless a first step in this direction. Later research can use the present thesis as a first basis for testing assumptions arising from the different perspectives and to successively build a theoretical model based on its findings.

Especially relevant for theory development on the effectiveness of EMME is that the current dissertation addressed the question of the role social mechanisms and factors play for the effectiveness of EMME. From a theoretical point of view, this is of interest because even though EMME are assumed to be grounded in the social cognitive learning theory (Bandura, 1986) and thus underlying social mechanisms for the effectiveness of EMME are automatically alleged (e.g., Jarodzka et al., 2013; Mason, Pluchino et al., 2015; Mason et al., 2016), up to date there is no empirical evidence that this assumption is justified. By showing that if EMME are framed as a social learning situation, social factors can influence their effectiveness, the present thesis offers first evidence that a social perspective on EMME might be valid. On the basis of the results in this thesis, however, no clear conclusion can be drawn on the concrete role of social mechanisms.

8.5.2 Practical implications

The findings from the present dissertation have not only theoretical, but also practical implications. First, the results of Experiments 2 and Experiment 3 indicated in line with the study by Scheiter et al. (2017) that EMME are not only an effective instructional tool for multimedia learning for younger learners, but also for adult learners. Hence, they might be used as instructional support not only in schools, but also in higher and further education settings.

Moreover, results from Experiment 2 indicated that contextualizing EMME through an introductory text that provides information about the model can have consequences for the effectiveness of EMME. In contrast to the presentation of a video depicting the model in person (e.g., in video modeling examples), the description of the model in the introductory text is a very indirect way of providing learners with information about the model. The fact that even this indirect communication of model characteristics can influence the effectiveness of EMME further underlines the importance of taking into account all types of contextualization of instructional measures that could provide indications about the model and especially his / her expert status.

Results from Experiments 2 and 3 further indicated that if EMME are framed as a social learning situation, a possible influence of social factors on their effectiveness should be considered. With view to the results from Experiments 2 and previous research by Scheiter et al. (2017), this might be especially important for learners with lower prior knowledge. In line with the study by Scheiter et al. (2017), the findings from both experiments suggest that learners with higher prior knowledge are less influenceable by social factors than learners with lower domain prior knowledge: Neither did they benefit from a competent model (as was assumed based on the model-observer similarity hypothesis), nor were they harmed by introducing the model as a peer learner. Against this background, it is therefore suggested to focus on the requirements of learners with lower domain prior knowledge and adapt the introduction of the EMME model accordingly. This approach would also be in line with previous research on aptitude-treatment interactions (ATIs; Kalyuga & Renkl, 2010). According Kalyuga and Renkl (2010), instructional principles usually do not apply to any type of learner. Furthermore, previous research indicated that learners' existing prior knowledge often moderates the effects of instructions. The effect that whereas instructional support is beneficial for learners with lower prior knowledge, the same instructional support can be ineffective or even detrimental for learners with higher prior knowledge, is called the '(partial) expertise reversal effect' (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Renkl, 2010). This effect has been replicated in many studies with various materials (see Kalyuga, 2007; Kalyuga et al. 2003 for an overview).

According to Kalyuga and Renkl (2010), a resulting practical implication of the expertise reversal effect is to adjust instructional methods and procedures to the individual requirements of the learners. Thus, with view to the use of EMME in learning situations, it might be necessary to contextualize the introductory texts to learners' cognitive prerequisites.

8.6 Conclusions

The present thesis aimed at investigating the role of social mechanisms in the effectiveness of EMME for multimedia learning. More specifically, the question of whether the consideration of social mechanisms contributes to explaining the effectiveness of EMME for multimedia learning beyond merely considering perceptual and / or cognitive explanations was addressed. Since results were inconsistent across the three experiments, it is not possible to draw final conclusions about the overall role of social mechanisms in the effectiveness of EMME. However, by showing that if EMME are framed as a social learning situation, social factors can influence their effectiveness for multimedia learning, the present thesis offers first empirical evidence that social mechanisms my contribute to the explanation of the effectiveness of EMME. Furthermore, results also support findings from previous research that EMME can be an effective instructional tool for multimedia learning not only for younger, but also for adult learners. Future research is required that further investigates the role of social mechanisms in the effectiveness of EMME, as well as replicates the findings of the presented research.

9 Summary

The preset dissertation aimed at investigating the role of social mechanisms and factors in the effectiveness of EMME (Eye Movement Modeling Examples) for multimedia learning. More specifically, the presented research addressed the question of whether social mechanisms and factors contribute to explaining the effectiveness of EMME for multimedia learning above and beyond perceptual and / or (meta-) cognitive mechanisms and factors.

Empirical research revealed that learning with multimedia material results in higher learning outcomes compared to learning with text alone (multimedia effect; Mayer, 2009). According to the CTML (Mayer, 2009), it is crucial that during multimedia learning learners actively perform cognitive processes of selection, organization and integration of information to successfully learn from multimedia. This assertation is underlined by empirical research showing that without instructional guidance learners often have difficulties in adequately applying these cognitive processes, and thus do not fully benefit from multimedia material (e.g., Hannus & Hyönä, 1999; Scheiter & Eitel, 2015; Schmidt-Weigand et al., 2010). As a learner-oriented instructional tool for multimedia learning, EMME illustrate how multimedia material should be processed in order to learn effectively. EMME consist of gaze replays that show the dynamic change in the gaze behavior of the model as it processes the task material. By watching the recorded videos with the model's eye movements overlaid onto the task material, other learners can observe the model's processing strategies. This is supposed to help them to perform a task that is either identical or similar to the task accomplished earlier on by the model. Although there is empirical evidence that EMME improve learners' multimedia processing behavior as well as their learning outcomes in multimedia environments (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017), the question of possible mechanisms underlying this effectiveness remained open. For this purpose, in the present dissertation it was differentiated between three perspectives with view to possible underlying mechanisms: The perceptual learning perspective, the (meta-) cognitive perspective, and the social perspective.

Based on the line of reasoning that EMME represent a special case of video-based modeling (Jarodzka et al., 2013), and thus are theoretically grounded in the social cognitive learning theory (Bandura, 1986), it was assumed that also social mechanisms contribute to explaining the effectiveness of EMME. For this reason, the preset dissertation focused at investigating the role of social mechanisms and factors in the effectiveness of EMME for multimedia learning.

Overall, three experiments were conducted in the present dissertation. In the experiments, participants with instructional support either received EMME-videos or abstract-cuing videos before the learning phase. Participants in the control condition received no instructional support before the learning phase. The EMME-videos used in the experiments were adopted from Scheiter et al. (2017) and showed a model's gaze replay of performing effective multimedia processing strategies on the first four pages of the later learning material. The abstract-cuing videos were created based on the EMME-videos and showed the same effective processing strategies by highlighting the processing sequence using an abstract attention-guiding cue. Importantly, no reference was made to the fact that the cueing was based on data obtained from a human model. All participants studied the same learning material and completed subsequently a knowledge test.

Two approaches were used to investigate the question of whether social mechanisms and social factors play a role in the effectiveness of EMME. In a first approach, it was investigated whether social mechanisms contribute to the effectiveness of EMME beyond the effect of perceptual and (meta-) cognitive mechanisms. Following the first approach, Experiment 1 compared EMME with a comparable attention guidance tool (i.e., abstract-cuing videos). It was assumed that if social mechanisms contribute to explaining the effectiveness of EMME, EMME would be more effective than an attention guidance tool without social cues. Results of Experiment 1 showed no beneficial effect on information processing and / or learning outcomes for either of the two kinds of instructional support tools, nor were there differences between EMME and abstract-cuing videos. Hence, the question of whether social mechanisms play role in the effectiveness of EMME remained open.

The second approach addressed the question of whether social factors can moderate the effectiveness of EMME. It was assumed that if social mechanisms underlie the effectiveness of EMME, social factors can influence the effectiveness of EMME. This hypothesis was investigated in all three experiments. In Experiments 2 and 3, (alleged) model competence and thus (alleged) model-observer similarity was manipulated by experimentally varying the description of the model in the introductory texts (competent vs. peer). In Experiments 1 and 3, learners' social comparison orientation was assessed as a potential moderator.

In Experiment 2, learners with lower prior knowledge benefitted from EMME, but only if the model was introduced as a 'peer' learner. This finding point to the possibility that (alleged) model-observer similarity might be one social factor that can influence the effectiveness of EMME. Experiment 3 aimed to replicate Experiment 2 as well as to extend it by addressing learners' social comparison orientation as other potential social factor. In contrast to Experiment 2, Experiment 3 did not only use learners' naturally occurring differences in prior knowledge to study its impact on the effectiveness of EMME (related to alleged model competence), but in addition also experimentally activated it. In contrast to Experiment 2, alleged model-observer similarity had no influence on the effectiveness of EMME, neither with regard to the existing prior knowledge, nor with regard to the activated prior knowledge. For social comparison orientation, results of Experiment 3 revealed an interaction between social comparison orientation and (alleged) model competence, indicating that learners with lower social comparison orientation benefitted only from EMME, if the model was introduced as a 'peer' learner. In contrast, Experiment 1 showed no effect of social comparison orientation on the effectiveness of EMME. To conclude, results from Experiments 2 and 3 point towards a relevance of social factors for the effectiveness of EMME, but the empirical evidence across the experiments is inconsistent.

In summary, results were inconsistent across the three experiments. Hence, it is not possible to draw final conclusions about the overall role of social mechanisms and social factors in the effectiveness of EMME. However, results supported findings from previous EMME-research, showing that EMME can be an effective instructional tool for multimedia learning (Mason, Pluchino et al., 2015, Mason et al., 2016, 2017; Scheiter et al., 2017). Moreover, results offered initial empirical evidence that social mechanisms may contribute to explaining the effectiveness of EMME. Thus, these results strengthened the assumption that the consideration of any kind of contextualization in instructional measures that provides information about the model's level of expertise might be important. Future research is required to gain further insights into the role of social mechanisms and the resulting consequences for the effectiveness of EMME for multimedia learning, as well as to replicate the findings of the presented research.

10 Zusammenfassung

Ziel der vorliegenden Dissertation war es, die Rolle sozialer Mechanismen und Faktoren bei der Wirksamkeit von EMME (Eye Movement Modeling Examples) für das multimediale Lernen zu untersuchen. Im besonderen Fokus stand hierbei die Frage, ob soziale Mechanismen und Faktoren dazu beitragen, die Wirksamkeit von EMME für multimediales Lernen über perzeptuelle und / oder (meta-) kognitive Mechanismen und Faktoren hinaus zu erklären.

Bisherige Forschung konnte zeigen, dass das Lernen mit multimedialem Material zu höheren Lernergebnissen führt als das Lernen mit Text allein (Multimedia-Effekt; Mayer, 2009). Um erfolgreich mit multimedialen Materialien zu lernen, ist es laut der CTML (Mayer, 2009) entscheidend, dass Lernende während des multimedialen Lernens kognitive Prozesse der Auswahl, Organisation und Integration von Informationen aktiv ausführen. Diese Annahme wird durch empirische Arbeiten gestützt, die zeigen, dass Lernende ohne instruktionale Unterstützung oft Schwierigkeiten haben, diese kognitiven Prozesse angemessen anzuwenden und somit nicht in vollem Umfang von multimedialem Material profitieren können (z. B. Hannus & Hyönä, 1999; Scheiter & Eitel, 2015; Schmidt-Weigand et al., 2010). Als lerner-orientierte instruktionale Unterstützung für multimediales Lernen veranschaulicht EMME, wie multimediales Material verarbeitet werden sollte, um effektiv zu lernen.

EMME bestehen aus Videos mit aufgezeichneten Blickbewegungen eines Modells, die die dynamische Veränderung des Blickverhaltens des Modells bei der Verarbeitung des Aufgabematerials zeigen. Durch das Betrachten der der Videos mit den auf das Aufgabenmaterial überlagerten Blickbewegungen des Modells, können andere Lernende die Verarbeitungsstrategien des Modells beobachten. Dies soll sie dabei unterstützen, eine Aufgabe auszuführen, die entweder identisch oder ähnlich der zuvor vom Modell durchgeführten Aufgabe ist. Obwohl bisherige Studienergebnisse zeigen, dass EMME sowohl die Nutzung multimedialer Verarbeitungsstrategien der Lernenden als auch ihre Lernergebnisse in multimedialen Umgebungen verbessern (Mason, Pluchino et al., 2015; Mason et al., 2016, 2017; Scheiter et al., 2017), blieb die Frage nach möglichen Mechanismen, die dieser Wirksamkeit zugrunde liegen, offen. Zu diesem Zweck wurde in der vorliegenden Dissertation mit Blick auf mögliche zugrunde liegende Mechanismen zwischen drei Perspektiven unterschieden: Die Perspektive des Wahrnehmungslernens, die (meta-) kognitive Perspektive und die soziale Perspektive.

Basierend auf der Annahme, dass EMME einen Sonderfall der videobasierten Modellierung darstellen (Jarodzka et al., 2013) und damit in der Theorie des sozialen kognitiven Lernens theoretisch begründet sind (Bandura, 1986), wurde vermutet, dass auch soziale

Mechanismen zur Erklärung der Wirksamkeit von EMME beitragen. Aus diesem Grund lag der Fokus der vorliegenden Dissertation auf der Erforschung der Rolle sozialer Mechanismen und Faktoren für die Effektivität von EMME für multimediales Lernen. Im Rahmen der vorliegenden Dissertation wurden insgesamt drei Experimente durchgeführt. In den Experimenten erhielten die Teilnehmenden mit instruktionaler Unterstützung entweder EMME-Videos oder Videos mit einem abstrakten Cue vor der Lernphase. Teilnehmende in der Kontrollbedingung erhielten keine instruktionale Unterstützung vor der Lernphase. Die in den Experimenten verwendeten EMME-Videos wurden von Scheiter et al. (2017) übernommen und zeigten auf den ersten vier Seiten des späteren Lernmaterials die Blickwiedergabe eines Modells währende der Durchführung effektiver Multimedia-Verarbeitungsstrategien. Die Videos mit einem abstrakten Cue wurden auf Basis der EMME-Videos designt und zeigten die gleichen effektiven Verarbeitungsstrategien, indem sie die Multimedia-Verarbeitungsstrategien mit einem abstrakten Cue hervorhoben. Hierbei ist wichtig, dass es keinen Hinweis für die Teilnehmenden gab, dass das Cueing auf Daten von einem menschlichen Modell basierte. Alle Teilnehmenden lernten mit dem gleichen Lernmaterial und absolvierten anschließend einen Wissenstest.

Mithilfe von zwei verschiedenen Ansätzen wurde die Frage untersucht, ob soziale Mechanismen und Faktoren eine Rolle für die Wirksamkeit von EMME spielen. In einem ersten Ansatz wurde untersucht, ob soziale Mechanismen über die Wirkung von perzeptuellen und (meta-) kognitiven Mechanismen hinaus zur Wirksamkeit von EMME beitragen. Diesem Ansatz folgend, verglich Experiment 1 EMME mit einer vergleichbaren instruktionalen Unterstützung zur Lenkung der Aufmerksamkeit (Videos mit einem abstrakten Cue). Es wurde angenommen, dass, wenn soziale Mechanismen zur Erklärung der Wirksamkeit von EMME beitragen, EMME effektiver sind als eine instruktionale Unterstützung ohne soziale Hinweise. Allerding zeigten die Ergebnisse von Experiment 1 weder einen positiven Einfluss von einer der beiden instruktionalen Maßnahmen auf die Informationsverarbeitung und / oder die Lernergebnisse, noch einen Unterschied zwischen EMME und den Videos mit einem abstrakten Cue. Damit blieb die Frage, ob soziale Mechanismen eine Rolle bei der Wirksamkeit der EMME spielen, ungeklärt.

Der zweite Ansatz zielte auf die Frage ab, ob soziale Faktoren die Wirksamkeit von EMME moderieren können. Es wurde angenommen, dass, wenn der Wirksamkeit von EMME soziale Mechanismen zugrunde liegen, soziale Faktoren die Wirksamkeit von EMME beeinflussen können. Diese Annahme wurde in allen drei Experimenten untersucht. In den Experimenten 2 und 3 wurde die (vermeintliche) Modellkompetenz und damit die (vermeintliche) Modell-Beobachterähnlichkeit manipuliert, indem die Beschreibung des Modells in den

Einführungstexten (kompetent vs. Peer) experimentell variiert wurde. In den Experimenten 1 und 3 wurde die soziale Vergleichsorientierung der Lernenden als potenzieller Moderator erhoben.

In Experiment 2 profitierten Lernende mit geringerem Vorwissen von EMME. Dies taten sie allerdings nur, wenn das Modell zuvor als "Peer " eingeführt worden war. Dieser Befund deutet auf die Möglichkeit hin, dass (vermeintliche) Modell-Beobachterähnlichkeit ein sozialer Faktor sein könnte, der die Wirksamkeit von EMME beeinflussen kann. Experiment 3 zielte darauf ab, Experiment 2 zu replizieren und zu erweitern, indem es zusätzlich die soziale Vergleichsorientierung der Lernenden als anderen potenziellen sozialen Faktor berücksichtigt. Im Gegensatz zu Experiment 2 nutzte Experiment 3 dabei nicht nur die natürlich auftretenden Unterschiede im Vorwissen der Lernenden, um dessen Auswirkungen auf die Effektivität von EMME (bezogen auf die angebliche Modellkompetenz) zu untersuchen, sondern aktivierte es zusätzlich auch experimentell. Im Gegensatz zu Experiment 2 hatte die (angebliche) Modell-Beobachterähnlichkeit keinen Einfluss auf die Wirksamkeit von EMME, weder im Hinblick auf das vorhandene Vorwissen noch im Hinblick auf das aktivierte Vorwissen. In Hinblick auf die soziale Vergleichsorientierung zeigten die Ergebnisse von Experiment 3 eine Interaktion zwischen sozialer Vergleichsorientierung und (vermeintlicher) Modellkompetenz, die darauf hindeutete, dass Lernende mit geringerer sozialer Vergleichsorientierung nur von EMME profitierten, wenn das Modell als "Peer" eingeführt wurde. Im Gegensatz dazu zeigten die Ergebnisse von Experiment 1 keinen Einfluss der sozialen Vergleichsorientierung auf die Wirksamkeit von EMME. Insgesamt deuten die Ergebnisse von Experiment 2 und 3 auf eine Relevanz sozialer Faktoren für die Wirksamkeit von EMME hin, allerdings sind jedoch die Ergebnisse über die Experimente hinweg eher inkonsistent.

Zusammenfassend ergeben die Ergebnisse aus den drei Experimenten insgesamt ein eher inkonsistentes Bild. Daher ist es nicht möglich, endgültige Schlussfolgerungen über die Rolle sozialer Mechanismen und Faktoren für die Wirksamkeit der EMME zu ziehen. Die Ergebnisse stützten jedoch die Ergebnisse früherer EMME-Forschungen und zeigten, dass EMME ein wirksame instruktionale Unterstützung für multimediales Lernen sein kann (Mason, Pluchino et al., 2015, Mason et al., 2016, 2017; Scheiter et al., 2017). Darüber hinaus lieferten die Ergebnisse erste empirische Hinweise darauf, dass soziale Mechanismen dazu beitragen können, die Wirksamkeit von EMME zu erklären. So bestärken diese Ergebnisse auch die Annahme, dass die Berücksichtigung jeglicher Art von Kontextualisierung in instruktionalen Unterstützungsmaßnahmen, die Auskunft über den Kenntnisstand des Modells gibt, wichtig ist. Zukünftige Forschung ist erforderlich, um weitere Erkenntnisse über die Rolle sozialer

Mechanismen und die daraus resultierenden Konsequenzen für die Wirksamkeit von EMME für das multimediale Lernen zu gewinnen und die Ergebnisse der vorgestellten Forschung zu replizieren.

11 References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions* (1st ed.). Newbury Park, CA: SAGE.

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education, 125,* 413–428. https://dx.doi.org/10.1016/j.compedu.2018.06.023

Amadieu, F., Mariné, C., & Laimay, C. (2011). The attention-guiding effect and cognitive load in the comprehension of animations. *Computers in Human Behavior, 27,* 36–40. https://doi.org/10.1016/j.chb.2010.05.009

Anglin, G., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. *Perception, 2,* 865–916.

Arndt, J., Schüler, A., & Scheiter, K. (2015). Text–Picture integration: How delayed testing moderates recognition of pictorial information in multimedia learning. *Applied Cognitive Psychology, 29,* 702–712. https://dx.doi.org/10.1002/acp.3154.

Baddeley, A. (1999). *Human memory.* Boston: Allyn & Bacon.

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

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (pp. 71–81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], *Encyclopedia of mental health*. San Diego: Academic Press, 1998).

Berger, S. M. (1977). Social comparison, modeling, and perseverance. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 209-234). Washington, DC: Hemisphere.

Bjerrum, A. S., Hilberg, O., Van Gog, T., Charles, P., & Eika, B. (2013). Effects of modelling examples in complex procedural skills training: a randomised study. *Medical education, 47*, 888–898. https://dx.doi.org/10.1111/medu.12199

Blanton, H., Gibbons, F. X., Buunk, B. P., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*, 420–430. https://dx.doi.org/10.1037/0022-3514.76.3.420

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*, 445–457. https://dx.doi.org/10.1016/S0883-

0355(99)00014-2

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

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction, 25,* 71–84. https://doi.org/10.1016/j.learninstruc.2012.11.005

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

Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22*, 1–36. https://dx.doi.org/10.1207/s1532690Xci2201_1

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. https://dx.doi.org/10.1016/j.cognition.2007.05.012

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174–205). Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139547369.010

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*, 3–21. https://dx.doi.org/10.1016/j.obhdp.2006.09.007

Buunk, B. P., Zurriaga, R., Gonzalez-Roma, V., & Subirats, M. (2003). Engaging in upward and downward comparisons as a determinant of relative deprivation at work: A longitudinal study. *Journal of Vocational Behavior, 62,* 370–388. https://dx.doi.org/10.1016/S0001-8791(02)00015-5

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

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.).(1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. Knowing, learning, and instruction:

Essays in honor of Robert Glaser, 18, 32–42.

Corcoran, K., Crusius, J., & Mussweiler, T. (2011). Social comparison: motives, standards, and mechanisms. *Theories in Social Psychology*, 119–139.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 9*, 109–142. https://dx.doi.org/10.1016/S0959-4752(98)00040-1

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

Desjarlais, M. (2017). The use of eye-gaze to understand multimedia learning. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 122–142). Hershey PA: IGI Global.

Dole, J. A., Nokes, J. D., & Drits, D. (2009). Cognitive Strategy Instruction. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 347–373). New York: Routledge.

Duchowski, A. T. (2007). *Eye Tracking Methodology* (2nd ed.). London: Springer London. https://dx.doi.org/10.1007/978-1-84628-609-4

Dzeng, R. J., Lin, C. T., & Fang, Y. C. (2016). Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification. *Safety science, 82,* 56–67. https://dx.doi.org/10.1016/j.ssci.2015.08.008

Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction, 41,* 70–84. https://dx.doi.org/10.1016/j.learninstruc.2015.10.003

Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology*, *27*, 451–461. https://dx.doi.org/10.1002/acp.2922

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction, 28,* 48–63. https://dx.doi.org/10.1016/j.learninstruc.2013.05.002

Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*, 117–140. https://dx.doi.org/10.1177/001872675400700202

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108,* 528–546. https://dx.doi.org/10.1037/edu0000065

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. https://dx.doi.org/10.1037/0033-2909.133.4.694

Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2017). Learning from where 'eye'remotely look or point: impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology, 71*, 1526 – 1534. https://dx.doi.org/10.1080/17470218.2017.1335335

Garcia-Rodicio, H. G. (2012). Learning from multimedia presentations: The effects of graphical realism and voice gender. *Electronic Journal of Research in Educational Psychology, 10,* 885 – 906.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research*, *60*, 517–529. https://dx.doi.org/10.3102/00346543060004517

Garton, A. F. (2008). Strategy Use and Learning in Problem Solving. In *Exploring Cognitive Development: The Child as Problem Solver* (pp. 44–68). Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470773574.ch3

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye tracking research in professional domains. *Educational Psychology Review, 23,* 523–552. https://dx.doi.org/10.1007/s10648-011-9174-7

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

Gegenfurtner, A., & van Merriënboer, J. J. (2017). Methodologies for studying visual expertise. *Frontline Learning Research, 5,* 1–13. https://dx.doi.org/10.14786/flr.v5i3.316

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*. https://dx.doi.org/10.1037/0022-3514.76.1.129

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,* 1–25. https://doi.org/10.1111/cogs.12529

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*, 434–441. https://dx.doi.org/10.1177/0963721417704888

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition, 33*, 1235-1247. https://dx.doi.org/10.3758/BF03193225

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

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

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013b). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, *83*, 3–28. https://dx.doi.org/10.1111/j.2044-8279.2011.02052.x

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, *24*, 95–123. https://dx.doi.org/10.1006/ceps.1998.0987

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9,* 188–194. https://dx.doi.org/10.1016/j.tics.2005.02.009

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717–742. https://dx.doi.org/10.1006/jmla.1993.1036

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York: Guilford Press.

Ho, H. N. J., Tsai, M. J., Wang, C. Y., & Tsai, C. C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International Journal of Science and Mathematics Education, 12*, 525–554. https://dx.doi.org/10.1007/s10763-013-9489-6

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Hoogerheide, V. (2016). *Effects of observing and creating video modeling examples on cognitive and motivational aspects of learning*. (Doctoral dissertation, Erasmus University, Rotterdam). Retrieved from https://repub.eur.nl/pub/93239/Hoogerheide-proefschrift-DEF.PDF

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://dx.doi.org/10.1016/j.learninstruc.2016.02.004

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (2017). Testing the model-observer similarity hypothesis with text-based worked examples. *Educational Psychology*, *37*, 112–127. https://dx.doi.org/10.1080/01443410.2015.1109609

Hoogerheide, V., Van Wermeskerken, M., Van Nassau, H., & Van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior*, 457 – 464. https://doi.org/10.1016/j.chb.2017.11.012

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20,* 172-176. https://dx.doi.org/10.1016/j.learninstruc.2009.02.013

Hyönä, J., Lorch, R. F. J., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*, 44–55. https://doi.org/10.1037/0022-0663.94.1.44

Jamet, E. (2014). An eye tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, *32*, 47–53. https://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. https://dx.doi.org/10.1007/s11251-012-9218-5

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

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*, 178–191. https://dx.doi.org/10.1037/a0026923

Johnson, C. S., & Lammers, J. (2012). The powerful disregard social comparison information. *Journal of Experimental Social Psychology, 48,* 329–334. https://dx.doi.org/10.1016/j.jesp.2011.10.010

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*, 329 – 354. http://dx.doi.org/10.1037/0033-295X.87.4.329

Kalyuga, S. (2013). Effects of learner prior knowledge and working memory limitations on multimedia learning. *Procedia - Social and Behavioral Sciences*, *83*, 25–29. https://dx.doi.org/10.1016/j.sbspro.2013.06.005

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19,* 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychlogist, 38,* 23–31. http://dx.doi.org/10.1207/s15326985ep3801_4

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38,* 209-215. http://dx.doi.org/10.1007/s11251-009-9102-0

Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education*, *17*, 371–388. Retrieved from http://digitalcommons.usu.edu/itls_facpub/62/

Kok, E. M., Jarodzka, H., De Bruin, A. B., BinAmir, H. A., Robben, S. G., & Van Merriënboer, J. J. (2016). Systematic viewing in radiology: seeing more, missing less?. *Advances in Health Sciences Education, 21*, 189–205. https://doi.org/10.1007/s10459-015-9624-y

Kok, E. M., & Jarodzka, H. (2017a). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical education, 51,* 114–122. https://doi.org/10.1111/medu.13066

Kok, E. M., & Jarodzka, H. (2017b). Beyond your very eyes: eye movements are necessary, not sufficient. *Medical education, 51,* 1190–1190. https://doi.org/10.1111/medu.13384

Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction*, *20*, 424–433. https://dx.doi.org/10.1016/j.learninstruc.2009.05.002

Krebs, M.-C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: The influence of model-observer similarity on Eye Movement Modeling Examples. *Learning and Instruction, 61*, 126–137. https://doi.org/10.1016/j.learninstruc.2018.10.005).

Krueger, J., & Stanke, D. (2001). The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin, 27*, 878–888 . http://dx.doi.org/10.1177/0146167201277010

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review, 10,* 90-115. https://dx.doi.org/10.1016/j.edurev.2013.10.001

Larson, C. O., Dansereau, D. F., Hythecker, V. I., O'Donnell, A., Young, M. D., Lambiotte, J. G., & Rocklin, T. R. (1986). Technical training: An application of a strategy for learning structural and functional information. *Contemporary Educational Psychology*, *11*, 217–228. https://dx.doi.org/10.1016/0361-476X(86)90018-4

Laugksch, R. C., & Spargo, P. E. (1996). Construction of a paper-and-pencil test of basic scientific literacy based on selected literacy goals recommended by the American association for the advancement of science. *Public Understanding of Science, 5,* 331–359. https://doi.org/10.1088/0963-6625/5/4/003

LeBel, M. E., Haverstock, J., Cristancho, S., van Eimeren, L., & Buckingham, G. (2018). Observational learning during simulation-based training in arthroscopy: Is it useful to novices? *Journal of Surgical Education*, *75*, 222–230. https://dx.doi.org/10.1016/j.jsurg.2017.06.005

Leopold, C., Doerner, M., Leutner, D., & Dutke, S. (2015). Effects of strategy instructions on learning from text and pictures. *Instructional Science*, *43*, 345–364. https://dx.doi.org/10.1007/s11251-014-9336-3

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations : A review of research. *Educational Technology Research and Development, 30,* 195–232.

Lin, Y. Y., Holmqvist, K., Miyoshi, K., & Ashida, H. (2017). Effects of detailed illustrations on science learning: an eye-tracking study. *Instructional Science*, *45*, 557–581. https://dx.doi.org/10.1007/s11251-017-9417-1

Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science, 38,* 503 – 521. https://doi.org/10.1007/s11251-009-9115-8.

Litchfield, D., & Ball, L. J. (2011). Rapid communication using another's gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology*, *64*, 649–656. https://dx.doi.org/10.1080/17470218.2011.558628

Litchfield, D., Ball, L. J., Donovan, T., Manning, D., & Crawford, T. J. (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. https://dx.doi.org/10.1037/a0020082

Liu, H. C., & Chuang, H. H. (2011). An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments, 19*, 503–517. https://dx.doi.org/10.1080/10494820903520123

Lorch, R. F., Lemarié, J., & Chen, H. T. (2013). Signaling topic structure via headings or preview sentences. *Psicologia Educativa*, *19*, 59–66. https://doi.org/10.1016/S1135-755X(13)70011-3

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*, 172–187. https://dx.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*. https://dx.doi.org/10.1111/bjet.12271

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

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers and Education*, *60*, 95–109. https://dx.doi.org/10.1016/j.compedu.2012.07.011

Mason, L., Tornatora, M. C., & Pluchino, P. (2015). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: an eye-movement study. *Reading and Writing*. https://dx.doi.org/10.1007/s11145-015-9552-5

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E. (2014a). 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. http:// doi.org/10.1017/CBO9781139547369.005

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 345–368). New York: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139547369.017

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*, 419–425. https://doi.org/10.1037/0022-0663.95.2.419

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*, 177–213. https://doi.org/10.1207/S1532690XCI1902

Neugebauer, J., Ray, D. G., & Sassenberg, K. (2016). When being worse helps: The influence of upward social comparisons and knowledge awareness on learner engagement and learning in peer-to-peer knowledge exchange. *Learning and Instruction*, *44*, 41–52. https://doi.org/10.1016/j.learninstruc.2016.02.007

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: An examination of fixation patterns in a science simulation. *Computers in Human Behavior*, *35*, 234–242. https://dx.doi.org/10.1016/j.chb.2014.02.040

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, *26*, 110–117. https://dx.doi.org/10.1016/j.chb.2009.09.001

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27 – 42). New York: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139547369.004

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16,* 87–91. http://dx.doi.org/10.1016/j.learninstruc.2006.02.004

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, M. (2001). Types of eye movements and their functions. In D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, & M.

Williams (Eds.), *Neuroscience* (2nd ed.). Sunderland (MA): Sinauer Associates. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK10991/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4–22. https://dx.doi.org/10.1016/j.cogsys.2005.07.002

Renkl, A. (2011). Instruction based on examples. In R. Mayer, & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 272–295). New York: Routledge.

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

Renkl, A., & Scheiter, K. (2015). Studying visual displays: How to instructionally support learning. *Educational Psychology Review*. https://dx.doi.org/10.1007/s10648-015-9340-4

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

Rodriguez Buritica, J. M., Eppinger, B., Schuck, N. W., Heekeren, H. R., & Li, S. C. (2016). Electrophysiological correlates of observational learning in children. *Developmental science, 19*, 699–709. https://dx.doi.org/10.1111/desc.12317.

Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction, 21*, 58–67. https://doi.org/10.1016/j.learninstruc.2009.11.001

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction, 32,* 37–50. https://doi.org/10.1016/j.learninstruc.2014.01.002

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*, 201-241. https://dx.doi.org/10.1207/s15327809jls1402_2

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. International *Journal of Computer-Supported Collaborative Learning, 4*, 69-92. https://dx.doi.org/10.1007/s11412-008-9054-4

Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on Eye-Movements Modeling Examples. *Journal of Educational Computing Research.*

https://dx.doi.org/10.1177/0735633117751605

Sanchez, R. P., Lorch, E. P., & Lorch, R. F. J. (2001). Effects of headings on test processing strategies. *Contemporary Educational Psychology*, *26*, 418–428. https://doi.org/10.1006/ceps.2000.1056

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, *36*, 11–26. https://dx.doi.org/10.1016/j.learninstruc.2014.11.002

Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In C. Was, F. Sansost, & B. Morris (Eds.), *Eye-tracking technology applications in educational research* (pp. 143–164). IGI Global.

Scheiter, K., Schubert, C., Gerjets, P., & Stalbovs, K. (2014). Does a strategy training foster students' ability to learn from multimedia? *The Journal of Experimental Education*, 1–24. https://dx.doi.org/10.1080/00220973.2013.876603

Scheiter, K., Schubert, C., & Schüler, A. (2017). Self-regulated learning from illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology*, 1–15. https://dx.doi.org/10.1111/bjep.12175

Schlag, S., Florax, M., & Ploetzner, R. (2007). The use of learning strategies in learning from text and pictures. In T. Hirashima, H. U. Hoppe, & S. S.-C. Young (Eds.), *Supporting learning flow through integrative technologies: Proceedings of the 15th International Conference of Computers in Education* (pp. 331–338). Amsterdam: IOS.

Schlag, S., & Ploetzner, R. (2011). Supporting learning from illustrated texts: Conceptualizing and evaluating a learning strategy. *Instructional Science*, *39*, 921–937. https://dx.doi.org/10.1007/s11251-010-9160-3

Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text?. *Learning and Instruction, 51,* 10–25. https://doi.org/10.1016/j.learninstruc.2017.02.002

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*, 100–110. https://dx.doi.org/10.1016/j.learninstruc.2009.02.011

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media*. Educational Research Review, 23,* 1-24. https://doi.org/10.1016/j.edurev.2017.11.001

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*, 375–397. https://doi.org/10.1007/s11412-013-9181-4

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.

Schnotz, W. (2014). Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 72–103). Cambridge: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139547369.006

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology, 106*, 974–989. https://dx.doi.org/10.1037/a0037054

Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction*, *49*, 218–231. https://dx.doi.org/10.1016/j.learninstruc.2017.03.001

Schüler, A., Scheiter, K., & Gerjets, P. (2013). Is spoken text always better? Investigating the modality and redundancy effect with longer text presentation. *Computers in Human Behavior*, *29*, 1590–1601. https://doi.org/10.1016/j.chb.2013.01.047

Schunk, D. H. (1987). Peer models and children’s behavioral change. *Review of Educational Research*, *57*, 149–174. https://dx.doi.org/10.3102/00346543057002149

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children’s self-efficacy and achievement. *Journal of Educational Psychology*, *77*, 313–322. https://dx.doi.org/10.1037/0022-0663.77.3.313

Seppänen, M., & Gegenfurtner, A. (2012). Seeing through a teacher’s eyes improves students’ imaging interpretation. *Medical Education*, *46*, 1113–1114. https://dx.doi.org/10.1111/medu.12041

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*, 227–237. https://dx.doi.org/10.1016/S0959-4752(02)00022-1

She, H. C., & Chen, Y. Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers and Education*, *53*, 1297–1307. https://dx.doi.org/10.1016/j.compedu.2009.06.012

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10,* 104–109.

https://doi.org/10.1111/j.1467-7687.2007.00571.x

Stalbovs, K., Scheiter, K., & Gerjets, P. (2015). Implementation intentions during multimedia learning: Using if-then plans to facilitate cognitive processing. *Learning and Instruction*, *35*, 1–15. https://dx.doi.org/10.1016/j.learninstruc.2014.09.002

Stanley, J., Gowen, E., & Miall, R. C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 915–926. https://dx.doi.org/10.1037/0096-1523.33.4.915

Stanley, J., Gowen, E., & Miall, R. C. (2010). How instructions modify perception: an fMRI study investigating brain areas involved in attributing human agency. *Neuroimage, 52*, 389-400. https://dx.doi.org/10.1016/j.neuroimage.2010.04.025

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. Cognitive Science, 12, 257–285. https://dx.doi.org/10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., Kalyuga, S. (2011). *Cognitive Load Theory.* New York: Springer-Verlag New York. https://dx.doi.org/10.1007/978-1-4419-8126-4

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296. https://doi.org/10.1023/A:1022193728205

Symons, L. A., Lee, K., Cedrone, C. C., & Nishimura, M. (2004). What are you looking at? Acuity for triadic eye gaze. *The Journal of General Psychology*, *131*, 451–69.

Tanaka, A., & Murayama, K. (2014). Within-person analyses of situational interest and boredom: Interactions between task-specific perceptions and achievement goals. *Journal of Educational Psychology, 106*, 1122–1134. http://dx.doi.org/10.1037/a0036659

Töpper, J., Glaser, M., & Schwan, S. (2014). Extending social cue based principles of multimedia learning beyond their immediate effects. *Learning and Instruction*, *29*, 10-20. https://doi.org/10.1016/j.learninstruc.2013.07.002

Van Gog (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: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139547369.014

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*, 785–791. https://dx.doi.org/10.1016/j.chb.2009.02.007

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review, 21,* 21–30. https://dx.doi.org/10.1007/s10648-008-9094-3

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205–221. https://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. https://dx.doi.org/10.1007/s10648-010-9134-7

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

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, *65*, 448–459. https://dx.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 Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J., & Van Merriënboer, J. J. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction, 32,* 10–21. https://doi.org/10.1016/j.learninstruc.2014.01.004

Van Wermeskerken, M., Litchfield, D., & Van Gog, T. (2018). What am I looking at? Interpreting dynamic and static gaze displays. *Cognitive science, 42*, 220–252. https://doi.org/10.1111/cogs.12484

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer, & P. Al-exander (Eds.), *Handbook of research on learning and instruction* (pp. 197–218). New York: Routledge.

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, *110*, 250–268. https://doi.org/10.1037/edu0000221

Wang, J. (2011). Pupil Dilation and Eye-tracking. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide* (pp. 185–204). New York: Psychology Press.

Wetzels, S. A. J., Kester, L., & Van Merriënboer, J. J. G. (2011). Adapting prior knowledge activation: Mobilisation, perspective taking, and learners' prior knowledge. *Computers in Human Behavior*, *27*, 16–21. https://doi.org/10.1016/j.chb.2010.05.004

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1,* 27–31. https://doi.org/10.1016/0022-1031(66)90062-X

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Testing a short scale of intrinsic motivation. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *15*, 31–45

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660–668. https://dx.doi.org/10.1037/0022-0663.94.4.660

12 List of Figures

13 List of Tables

14 Appendix

Table 8: Means and Standard Deviations (in parentheses) of additional measures in the conditions with and without instructional support in Experiments 1 and 3

	EMME (Overall)	**EMME (Competent-Model)**	**EMME (Peer-Model)**	**Abstract-cuing Video**	**Control**
Experiment 1 (*N* = 114)	*n* = 37	-	-	*n* = 41	*n* = 59
Judgement of Learning (in percent)	64.95 (18.66)	-	-	62.93 (20.33)	64.17 (19.85)
Judgement of Learning (Accuracy)	4.01 (17.64)	-	-	1.73 (18.21)	5.41 (15.75)
Motivation with regard to Instructional Support Tool (mean)	2.63 (0.74)	-	-	3.25 (0.62)	-
Experiment 3 (*N* = 174)	*n* = 115	*n* = 57	*n* = 58	-	*n* = 59
Judgement of Learning (in percent)	59.44 (22.64)	59.93 (22.51)	58.97 (22.95)	-	47.81 (21.48)
Judgement of Learning (accuracy)	-5.06 (18.61)	-4.77 (18.69)	-5.35 (18.69)	-	-9.40 (19.76)
Interest and Motivation (mean)	3.37 (0.74)	3.32 (0.75)	3.43 (0.73)	-	3.11 (0.76)
Motivation with regard to Instructional Support Tool (mean)	2.74 (0.69)	2.74 (0.69)	2.75 (0.70)	-	-