REVIEW ARTICLE

Visualizing a Task Performer's Gaze to Foster Observers' Performance and Learning—a Systematic Literature Review on Eye Movement Modeling Examples

Selina N. Emhardt[1] · **Ellen Kok**[1,2] · **Tamara van Gog**[2] ·
Saskia Brandt-Gruwel[1,3] · **Tim van Marlen**[2] · **Halszka Jarodzka**[1]

Abstract

Eye movement modeling examples (EMMEs) are instructional videos (e.g., tutorials) that visualize another person's gaze location while they demonstrate how to perform a task. This systematic literature review provides a detailed overview of studies on the effects of EMME to foster observers' performance and learning and highlights their differences in EMME designs. Through a broad, systematic search on four relevant databases, we identified 72 EMME studies (78 experiments). First, we created an overview of the different study backgrounds. Studies most often taught tasks from the domains of sports/physical education, medicine, aviation, and STEM areas and had different rationales for displaying EMME. Next, we outlined how studies differed in terms of participant characteristics, task types, and the design of the EMME materials, which makes it hard to infer how these differences affect performance and learning. Third, we concluded that the vast majority of the experiments showed at least some positive effects of EMME during learning, on tests directly after learning, and tests after a delay. Finally, our results provide a first indication of which EMME characteristics may positively influence learning. Future research should start to more systematically examine the effects of specific EMME design choices for specific participant populations and task types.

Keywords Eye movement modeling examples · Systematic review · Performance · Learning · Attention guidance · Instructional design

---

Selina N. Emhardt and Ellen Kok contributed equally to this work.

✉ Ellen Kok
  e.m.kok@uu.nl

1 Department of Online Learning and Instruction, Open University of the Netherlands, Heerlen, The Netherlands

2 Department of Education, Utrecht University, P.O. Box 80140, 3508 CS Utrecht, The Netherlands

3 Zuyd Hogeschool, University of Applied Sciences, Heerlen, The Netherlands

Introduction

Following eye movements is considered vital for communication and learning gaze (Csibra & Gergely, 2009; Frischen et al., 2007). From early on, humans are highly sensitive to the direction of other people's gaze and when parents interact with their infants, they naturally look at the objects that they are verbally referring to, which fosters their infants' understanding (Baldwin, 1995; Bloom, 2002; Scaife & Bruner, 1975). Indeed, the direction of gaze contains valuable information about what objects in the world a person is looking at and, hence, presumably processes and thinks about (Just & Carpenter, 1980) and facilitates communication by guiding the attention of communication partners to the information that the other person is referring to (Frischen et al., 2007).

Nowadays, we can also harness this power of natural eye gaze for (asynchronous) communication and learning in online environments. That is, with modern eye-tracking technology, it is possible to capture where a person is looking over the course of time (for more information on eye tracking see Holmqvist et al., 2011) and to visualize this as an overlay on the image that was observed or on a video of the scene that was observed.

The potential of such gaze displays for learning has been recognized in the literature, and a major application has been in the design of instructional videos in which an individual (e.g., expert, teacher, peer student) demonstrates (and often explains) to observers how to perform a task (Van Gog et al., 2009). In educational sciences, such videos are often referred to as eye movement modeling examples (EMMEs, Jarodzka et al., 2012). For example, the gaze displayed in an EMME can show a medical student exactly what a doctor looks at when diagnosing a patient (e.g., Jarodzka et al., 2012; Litchfield et al., 2010) or can demonstrate to students how to integrate verbal and pictorial information by visualizing a model's gaze when reading an illustrated text (e.g., Mason et al., 2015, 2016, 2017; Scheiter et al., 2018). An example EMME video as used in Emhardt et al. (2022) can be found on YouTube at https://www.youtube.com/watch?v=iqU_BxtKP80.

Gaze visualizations can take different forms, such as scan paths in which dots (or circles) indicate individual fixations (moments when the eye is relatively still and takes in information) or attention maps that use shades of color ("heatmaps" or "spotlights") or blurring to show which objects or areas were fixated (and for how long), and which areas were not looked at. Gaze visualizations can be either static (e.g., an image in which the full scan path is visible with all dots at once) or dynamic (a video in which the scan path/attention map unfolds in real time as it was recorded, for instance by showing how the dots appear and disappear on different locations on the screen over time). Figure 1 shows static screenshot examples of an attention map (spotlight) and scan path visualization.

Including a visualization of the task performer's eye movements (hereafter referred to as "gaze visualization") in an instructional video (i.e., an EMME) has been found to foster observers' *performance* (e.g., score) on the task in the presence of the gaze visualization (e.g., Gegenfurtner et al., 2017; Litchfield et al., 2010; Nalanagula et al., 2006), and *learning* (i.e., later performance of
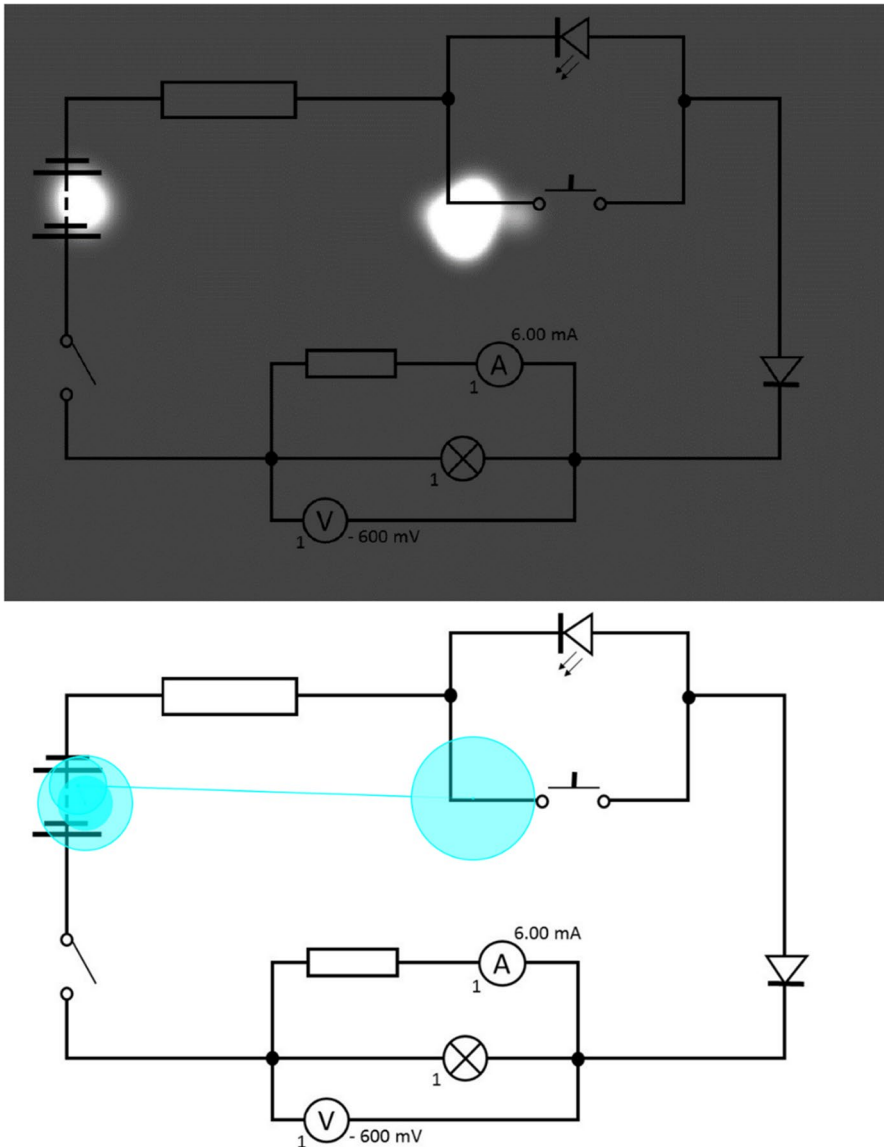
Fig. 1 Example screenshots of an attention map (top) and scan path visualization (bottom). Note: Visualizations based on materials of Chisari et al. (2020). The spotlight visualization darkens areas that the performer has not looked at. The scan path visualization overlays the performer's fixations (moments when the eye is relatively still and takes in information) as blue circles (larger circles indicate longer fixations) and uses connection lines to indicate saccades (jumps between fixations)

that same or an isomorphic task, in the absence of the instructional video either immediately after studying the EMME or after a delay; e.g., Bednarik et al., 2018; Jarodzka et al., 2013; Krebs et al., 2021; Mason et al., 2015).

A recent meta-analysis by Xie et al. (2021) analyzed 25 studies and concluded that EMME significantly affected observers' attention and learning and fostered observers' performance. However, this analysis merely included studies that used the specific term EMME, which restricts the findings to studies with a similar conceptual background (i.e., educational science). There is a long tradition of using gaze displays to foster task performance outside the field of educational science (e.g., movement sciences, medical image perception, ergonomics, computer sciences, see Harle & Vickers, 2001; Litchfield et al., 2010; Nalanagula et al., 2006; Stein & Brennan, 2004). While those studies do not use the term EMME, they are conceptually similar to EMME in that gaze is displayed on top of to-be-learned task material with the goal of improving the observer's performance or learning. While conceptually similar, those different research fields are relatively unconnected and cross-references between individual studies with the same goal are often lacking. Furthermore, a plethora of different terminologies is used, such as EMME, gaze(-based) training, gaze-augmented think-aloud, scanpath-based feedforward training, or visual attention training. Studies also differ in terms of the rationale for displaying gaze visualizations, study characteristics, the design of the visualizations, and ultimately, the study outcomes. Importantly, while they differ on those dimensions, what they share is that gaze is displayed on top of to-be-learned task material with the goal of fostering an observer's performance of learning. Thus, the field would benefit from focusing on the conceptual similarities and mapping the variation within the field to understand how this variation affects EMME effectiveness. To date, there is no detailed and comprehensive overview of studies available on the effects of visualizing a performer's gaze to foster observers' performance or learning of tasks from different domains, and the systematic literature review that we present here aimed to fill this gap. Ultimately, connecting studies from different research fields will help to provide a more complete overview of the current knowledge and to set a research agenda for the near future. Note that in the remainder of the paper, we will adopt the term EMME (used in educational sciences) to refer to all such gaze visualizations (i.e., even if the original authors used another term).

In the following sections, we introduce dimensions in which EMME studies are expected to vary: the rationales for using EMME, study characteristics, and outcome variables, before introducing our research questions.

Rationales for Using EMME to Foster Performance or Learning

Krebs et al., (2019, 2021) claimed that there are three mechanisms that underlie the effectiveness of EMME to foster observers' learning or performance: guiding attention, illustrating advanced perceptual strategies, and inducing a stronger social learning situation (Krebs et al., 2019, 2021). In this systematic literature review, we first identify which of these mechanisms that could underlie the effectiveness of EMME were used as rationales in EMME studies of different domains. Such study backgrounds should be reported to put the different studies in perspective.

First, the visualization of the task performer's eye movements may serve as visual cue to guide the observer's attention, which can facilitate selection of relevant

information. Previous studies in the field of educational science have repeatedly found that adding cues to multimedia materials (e.g., instructional videos that contain verbal and visual information) can foster learning (Richter et al., 2018; Van Gog, 2014, 2021). According to the Cognitive Theory of Multimedia Learning (Mayer, 2005), learners must select, organize, and integrate relevant information effectively to learn from multimedia materials successfully. Because information in instructional videos is often transient, if learners are unable to select the right information at the right time, they will not be able to organize and integrate it with their prior knowledge, and learning is hampered. Visual cues can resolve this problem by guiding learners' attention to the right information at the right time.

The gaze visualization presented in EMME can be considered a specific type of visual cue that can help observers to attend and select the relevant information at the right time (Van Gog, 2014). Although verbal explanations could also be used in videos to direct learners' attention, such explanations often remain insufficient and ambiguous for two reasons (Van Marlen et al., 2018). First, an instructional video might contain competing visual information, such as multiple objects that the teacher could refer to with a single term (e.g., when the instructor talks about a resistor, and there are several in the drawing of the electrical circuit). Second, the teacher's explanations possibly lack specificity when clear location indications are missing (Louwerse & Bangerter, 2010). Gaze visualizations can help to make an expert's verbal explanations and references more understandable and easier to follow by guiding the observers' attention to the referenced objects. This helps to select the relevant information at each moment in time and, hence, foster understanding (Betrancourt, 2005; de Koning & Jarodzka, 2017; Van Gog, 2014).

Another rationale for displaying EMME is that they can reveal and teach a performer's perceptual strategies that would otherwise not be accessible to observers. For example, EMME could show an expert learner's reading strategies when working with multimedia materials (e.g., Mason et al., 2015), an engineering expert's systematic strategy for a executing a visual inspection task (e.g., Nalanagula et al., 2006), or a golf expert's approach of making a long fixation on the target just before initiating a movement (e.g., Moore et al., 2014). People have limited insight into their own viewing behavior, and they have trouble reporting where they have looked (Clarke et al., 2017; Foulsham & Kingstone, 2013; Kok et al., 2017; Marti et al., 2015; Van Wermeskerken et al., 2018; Võ et al., 2016). Experts might even report executing a different strategy from the strategy they actually used (e.g., Aizenman et al., 2017). Thus, an expert's gaze visualization provides continuous information about all steps of the expert's visual search strategy without necessarily requiring conscious control by the expert (Gallagher-Mitchell et al., 2018).

Finally, inducing a social learning situation is also often reported as a rationale for displaying EMME to foster learning and performance. While other visual cues could also serve the purpose of guiding attention to foster learning, gaze visualizations are likely to evoke a greater sense of social presence than more abstract cues such as an animated arrows (Krebs et al., 2019). Adding social cues to multimedia materials can evoke a feeling of social interaction, which can in turn stimulate deeper processing of the materials and learning (social agency theory, Mayer et al., 2003; Moreno et al., 2001).

Study Characteristics That Might Affect Observers' Learning or Performance

Different variables may affect the effectiveness of EMME for observers' performance or learning (i.e., participant characteristics, task types, and EMME displays), which we took into account in this systematic review. First of all, and in line with Xie et al. (2021) and Krebs et al. (2021), we expect to find that the effects of EMME are different depending on observers' prior knowledge. Furthermore, the task type might play a role. The meta-analysis of Xie et al. (2021) distinguished between procedural and non-procedural task types. While this provides valuable insights, it may be worthwhile to identify task categories at a finer level than the dichotomous classification of procedural vs. nonprocedural tasks (as introduced in Van Marlen et al., 2016), because non-procedural tasks differ widely in their requirements and, therefore, the gaze visualization likely contains different kinds of information (e.g., how to perform a visual search or classification task or how experts use different reading strategies).Our systematic review aims to categorize the used task materials in a bottom-up manner on a more fine-grained level to better understand for which task types EMMEs are effective.

Third, studies differ in the way they visualize performers' eye movements in EMME. Two common types of data visualization techniques are scan paths and attention maps. Scan paths display the order of performers' fixations overlaid onto the stimulus, for instance by connecting fixation visualizations (e.g., circles or dots) through lines that indicate saccades. In contrast, attention maps show the spatial distribution of aggregated eye movement data (aggregation of fixations over time and/or participants), for instance as heat map (Blascheck et al., 2014; Holmqvist et al., 2011) (see Fig. 1 for examples). Gaze visualizations can be superimposed onto the original screen and thus add additional information, but can also remove information (e.g., when a spotlight shows where a performer looked, and other information is blurred out). Furthermore, they differ in whether the gaze visualization is a static picture or a dynamic video. Finally, gaze visualizations could be based on raw data (i.e., location of gaze over time, without filtering or detection of fixations and saccades) or processed data (e.g., fixations and saccades are detected in the data, and thresholds to only display fixations of a certain duration might be applied) which affects the amount of detail and noise in the visualizations. Aside from different gaze visualization options, EMME materials may also differ in terms of displayed performer behavior which may affect observers' learning and performance. Task performers may show their natural approach to performing the task or might (have been instructed to) adopt a didactic teaching approach. To date, it is unknown how this affects learning and performance. Thus, acquiring an overview of how EMMEs are created in studies across different domains can contribute to a better overview of the research field and can help identify gaps in our understanding of the effectiveness of EMME literature and thus outline directions for future research.

Exploring the Potential of EMME for Different Outcome Variables

Overall, EMMEs have often been found to positively influence learning (Xie et al., 2021). However, these beneficial effects of EMME could be reported on

a more fine-grained level, for instance by making a distinction between effects on learning and performance outcomes. We use the term "performance" when the outcome measure concerns the participant's immediate performance on the task on which the gaze is displayed; that is, it refers to observers' accuracy score, time on task (i.e., the time participants took to execute the task), and gaze behavior on the task on which the EMME is presented. "Learning" refers to performance on future tasks in the absence of the gaze display; that is, accuracy scores, time on task, and gaze behavior on the same or an isomorphic task presented *after* studying the EMME (cf. Soderstrom & Bjork, 2015). We argue that effects on performance measures indicate that EMMEs alter observers' attention and performance while processing the task on which the EMME is displayed. However, such effects do not necessarily imply a longer-lasting impact on participants after the EMME was presented, as effects on learning measures would imply (i.e., this would show that the altered attention while observing EMME affected knowledge acquisition, as evidenced by better performance on similar tasks later on in the absence of the EMME). Moreover, EMME studies can differ in when learning is measured, by having learners perform similar tasks immediately after observing the EMME or at a delay (i.e., measured at least one day after the EMME was presented), with the latter evidencing longer-lasting effects. Using these distinctions in our systematic literature review can provide an overview of what effects existing EMME studies have mostly investigated to date and reveal the potential of EMME to foster performance and (immediate and delayed) learning. While also important, we do not further distinguish different types of tasks used to assess learning outcomes (e.g., recall or transfer task) as there is too much variation in the tasks between studies to make meaningful comparisons between them.

Overview of the Present Study and the Research Questions

This systematic literature review provides a detailed overview on the effects of displaying a performer's gaze visualizations to foster observers' performance and learning. By using a wide range of search terms, we aimed to not only cover literature from traditional EMME research, but also to identify studies that use different terms to investigate comparable topics within different research fields. In this manner, we aim to connect results from different research domains and provide an overview of the characteristics of the existing studies. More specifically, we aim to answer the following research questions (RQs):

> RQ1. What is the background of the identified studies, their used terminologies to refer to EMME, and their rationale for using EMME?
> RQ2. To what extent do the studies vary in terms of participant characteristics, task type, and design of EMME materials? In the context of this question, we also aim to provide a first indication of which study characteristics may influence the effects of EMME on learning.

RQ3. Which different types of effects did the identified EMME studies explore? We distinguish between effects on observers' task performance and immediate and delayed learning, as well as effects on time on task and gaze.

Methods

Search and Selection

We conducted the systematic literature search in the online databases of EbscoHost (i.e., APA PsycInfo, ERIC, Psychology and Behavioral Sciences Collection, APA PsycArticles), PubMed, the ACM Guide to Computing Literature, and Web of Science. We chose the repositories of EbscoHost because of their focus on the literature from psychology and educational sciences, with ERIC being the largest repository in education. Pubmed and the ACM Digital Library were selected because of their specific focus on literature from the medical field and computer sciences. In the field of computer science, studies are often published in (peer-reviewed) conference proceedings. Therefore, we decided to include conference proceedings (full articles, no short summaries or abstracts) in our searches. Finally, we included Web of Science as a database because it covers a broader range of literature from other domains. We initially performed the systematic search in May 2020 and updated all searches on 12.01.2022.

We searched for the following terms in English language in the abstracts and titles:

Topic 1. Displaying: (display* OR visuali* OR augment* OR highlight* OR superimpos* OR
replay* OR "eye movement model*" OR "model's eye movement*") AND
Topic 2. Eye movements: ("eye movement*" OR gaze OR "visual attention" OR "scan path*"
OR "point of regard") AND
Topic 3. Learning: (learn* OR teach* OR instruct* OR train* OR guid* OR perform*).

We confirmed that we could find known key literature on EMME with these search terms. We restricted our search to literature published after January 2000. From the year 2000, commercial eye trackers became increasingly accessible for researchers and, consequently, an established tool in research fields such as educational sciences (cf. review of Lai et al., 2013). Our search furthermore focused on empirical, peer-reviewed studies from journals or conference proceedings to ensure a certain standard of quality for the included studies. The subsequent description of the search and selection process is visualized as a PRISMA flow diagram (Moher et al., 2009) in Fig. 2.

The searches in the four databases yielded 8434 publications in total: 1899 hits on EbscoHost (1544 initial results in May 2020 and 355 results from the updated search in January 2022), 4143 hits in Web of Science (3255 and 888 results), 1648 hits
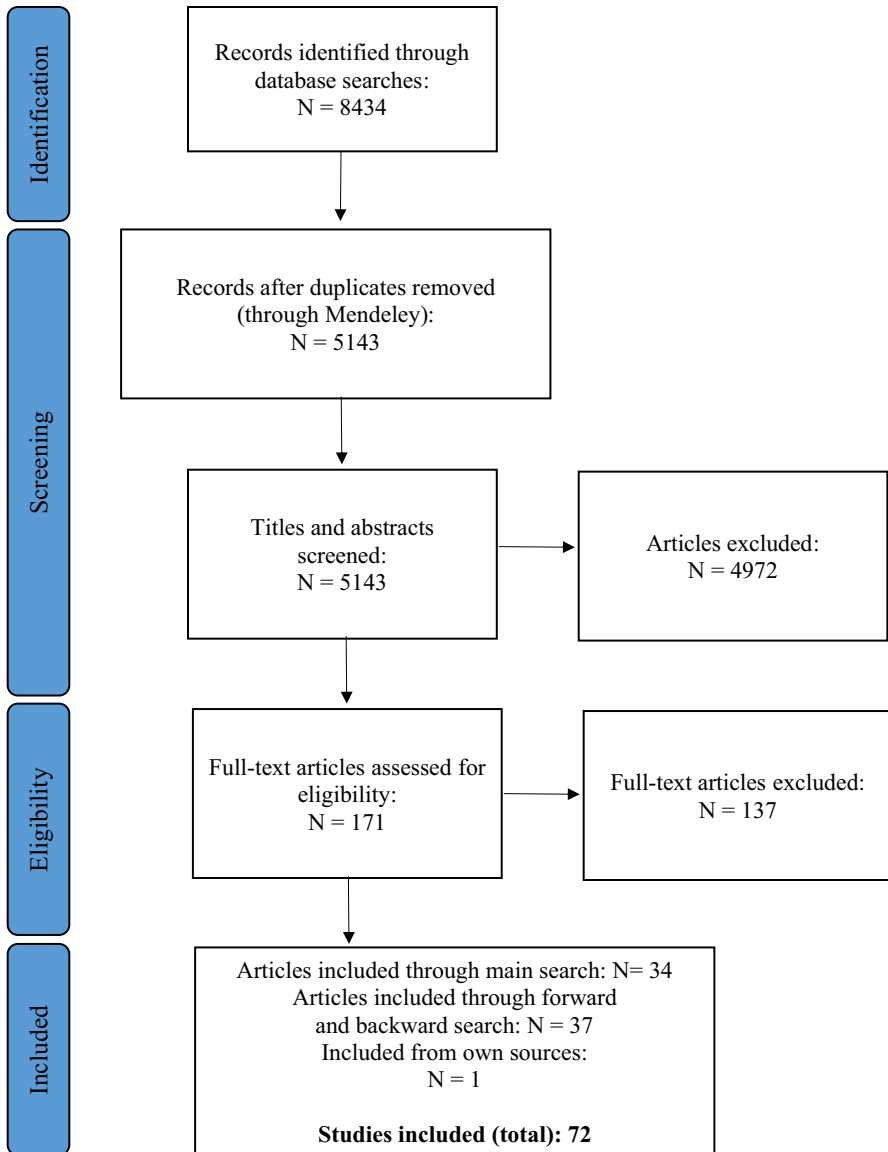
Fig. 2 Flow diagram of the systematic literature search

in PubMed (1257 and 391), and 744 hits in the ACM Digital Library (650 and 94 results). We removed duplicates within and across these databases with Mendeley (https://www.mendeley.com), which resulted in 5143 results. For the literature selection process, we uploaded all unique results into Rayyan (https://rayyan.qcri. org), which is an online platform to collaborate on the inclusion and exclusion process when performing a systematic review. On this platform, we screened the

titles and abstracts of all unique results manually. During this screening, we aimed to include studies investigating the effects of EMME to foster observers' performance and learning. We excluded studies that did not use any gaze visualizations. We calibrated this selection process based on abstracts and titles with Rayyan using the first 600 hits of the first search and three independent reviewers. Two researchers screened each of these records independently: The first author screened every record, the second and third author each screened half of the records and decided for each article whether it should be included, excluded, or discussed ("maybe"). The first author agreed on all cases with the other raters on definite inclusions and in 5 out of 8 of the cases on literature that would possibly be suitable for inclusion ("maybe"). The overwhelming majority of the articles could be excluded, because they did not visualize the gaze of another person to observers. Disagreements during the first screening round were resolved by discussion.

After this calibration process, the first author continued the selection process based on abstracts and titles independently. In case of doubt, the first author took a note for further discussions with the other authors. This first screening based on titles and abstracts resulted in 171 (112 and 59 articles) possibly relevant articles. For these articles, we read the details of the studies to decide whether they should be included or excluded. In case of doubt, the decision was discussed among the first three authors and we created a more specific guideline on inclusion and exclusion criteria (see Table 1).

Based on these exclusion criteria, we included 34 publications from the initial search (28 articles from the initial search and 6 articles from the updated search) and 1 from our own sources. Through examining reference lists and the literature that cited these articles, we identified 37 additional articles (out of which 19 articles belonged to the research area of Quiet Eye Training (Vickers, 2007) that was not originally included as search term and is to date not usually connected to the EMME literature). In total, our search resulted in 72 included articles. In all included articles, a gaze visualization was used with the aim of fostering observers' performance or learning.

Data Extraction and Analysis

To provide some general background information about the identified studies, we first gathered information about the publication type (i.e., conference proceeding or journal article) and the study design of the experiments. Regarding the study design, we reported the sample sizes after the main exclusions, whether the identified studies used a within or a between-subjects design, the average sample size per (between-subject) condition. Furthermore, we coded whether the EMME was the only difference between the experimental conditions (i.e., the main manipulation) or not. We did not code the EMME as the main intervention in cases where the EMME was part of a larger intervention or when the control group differed in more than the absence of the gaze visualization from the experimental condition. This was the case if the control group did not receive the same materials just without EMME, or if they received additional instructions. In addition, we assessed the quality of the

Table 1 Overview of the Specific Reasons for Including and Excluding Articles

Inclusion criteria	Description and explanation
Use of gaze visualizations	A visualization of another person's gaze is presented to human observers
Target outcome	The aim of the study is to foster learning or performance (as indicated by test scores)

Exclusion criteria	Description and explanation
Use of gaze visualizations	Main exclusion criterion. The study did not mention the use of gaze visualizations of another person that is displayed to human observers
Target outcome	Main exclusion criterion. The study did not focus on visualizing gaze to foster learning or performance (as indicated by test scores)
Data type	The study did not provide new, empirical data that was analyzed through significance tests. In this context, we also excluded methodological papers on eye movement analysis or technical overviews
Task	Participants interacted during the study while performing a task that was not performed in an instructional context (e.g., tasks with the main goals of cooperation, collaboration, communication, or competition tasks). In such cases, the EMME display likely change based on what the participants are doing, making a distinction between the effects of interaction and gaze visualizations difficult
Design	The study design did not include a control condition without any gaze visualizations
Publication type	The study was not published in a peer-reviewed conference proceeding or peer-reviewed scientific journal. We excluded (grey) literature sources such as theses, conference posters and abstracts, dissertations, pre-prints, book chapters, patents, and research reports

included studies using an adapted version of the Medical Education Research Study Quality Instrument (Reed et al., 2007). This instrument was originally developed to measures the quality of educational studies in the field of medical education. The major categories of the MERSQI focus on the quality of study design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes. Each study could receive at most 3 points per category, which resulted in a maximum score of 18 points. We adjusted the content of these subscales for our purposes (see last section of the table in the supplementary materials for detailed information on how the categories were rated). To answer the three main research questions, we extracted the following data.

Data Extraction for Research Question 1

In the context of RQ 1, we coded information on the taught domain and the rational for using EMME in each study. To report on the domain, we grouped the content of the task materials into categories with at least five studies (cutoff criterion). In addition, we indicated which terminologies the studies used in the title or abstract to describe the EMME intervention, using the same cutoff criterion of at least 5 studies using the same term to define a category. We furthermore coded which rationales for EMME the studies used in their introduction and distinguished between the three

main lines of reasoning that were described in detail in our introduction: First, gaze visualizations could serve as visual cues to guide the observers' attention to the most relevant areas (de Koning & Jarodzka, 2017). Second, EMME could reveal and teach a performer's perceptual and cognitive strategies during task performance, which would otherwise be difficult to observe (e.g., Mason et al., 2015). We consider strategies to be deliberate (perceptual) actions to effectively reach the task goal. In this context, we noticed that several studies referred to the aim of training the particular gaze strategy of a "quiet eye" (QE). A QE is an (expert) gaze strategy that is characterized by a longer final fixation towards a relevant object before executing a movement (e.g., hitting a golf ball) and has been described as part of optimal visual attentional control (Vickers, 2007). Due to the unexpectedly large amount of literature on Quiet Eye Training (QET) interventions, we decided to specifically report whether the EMME was displayed with the aim of demonstrating this gaze strategy and discuss QET in more detail in the discussion. Third, EMME could foster observers' performance and learning though social mechanisms such as evoking a more social learning situation by increasing social presence (e.g., Mayer, 2014; Mayer et al., 2003) or establishing the social state of joint attention (e.g., Butterworth, 1995; Shvarts, 2018; Tomasello & Farrar, 1986).

Data Extraction for Research Question 2

In the context of RQ2, we coded the characteristics of the identified studies that were related to the participants, type of task, the EMME materials (i.e., gaze visualization type and performer guidance). In the next section, we explain how all variables were coded in detail.

Regarding participant characteristics, we first categorized study participants based on their prior knowledge (i.e., no or low prior knowledge, some prior knowledge, or mixed). For this categorization, we took over the authors' description wherever possible. If the authors did not provide explicit information on participants' prior knowledge, we compared the background of the participants with the domain of the task materials to decide whether participants had no or some prior knowledge. For instance, medicine students learning a medical task were assumed to have at least some prior knowledge, while students from all other domains were not assumed to have any prior knowledge in medical tasks. Furthermore, we categorized participants' education level in primary and secondary education, tertiary education, professional, and other.

Regarding the task type, we categorized the literature into studies that dealt with problem-solving tasks, visual inspection or classification tasks, visuo-motor tasks, tasks that teach strategies of text processing or comprehension, and other tasks. This categorization was based on the identified literature of current systematic review and considerations of Van Marlen et al. (2016).

Regarding the gaze visualization, we first reported whether the EMME used a scan path or attention map visualization, coded whether the visualization added or removed information from the screen, and whether the EMME was static or dynamic. For instance, a moving spotlight can be classified as attention map that removes information by blurring the surrounding that was not fixated by the task

performer, while heatmaps add highlights to the task materials to show which information was fixated. Other EMMEs use scan path visualizations like moving dots and circles to add information to the screen (see also Jarodzka et al., 2013). We also indicated how the raw eye-tracking data were processed to create the EMME. We distinguished between EMME that display the raw eye movement data (i.e., gaze locations plotted over time), EMME of event-detected eye movement data (e.g., only fixations of a specific minimum duration are displayed), and EMME of higher-order processes (e.g., gaze visualizations that show aggregated eye movement data). If the data transformation was not described in detail, we coded it as not specified.

Regarding the type of guidance in the EMME materials, we indicated what type of performer was displayed (i.e., novice or peer, teacher or expert, a researcher, or not indicated), and which instructions they received to create the materials (i.e., natural, didactic, or not indicated). We furthermore indicated whether additional, verbal guidance was available, for instance as voice-over in videos. Finally, we calculated the relative frequency (%) with which studies using these different design decisions (i.e., participant characteristics, task type, gaze visualization, and type of instruction) reported at least one positive effect on immediate learning. We chose the outcome measures of immediate learning because such measures are highly relevant to educational science and because this measure was more frequently used than measures of delayed learning.

Data Extraction for Research Question 3

In the context of RQ 3, we distinguish between effects on observers' performance *while* observing the EMME or on observers' learning *after* studying the EMME. In terms of learning measures, we furthermore distinguished between measures of immediate learning which was assessed immediately after the EMME and delayed learning after at least one day. While learning and performance scores were our main measures of interest, we also reported on additional effects on participants' time on task (i.e., the time participants took to execute the task) and gaze measures while processing the EMME and while performing subsequent test tasks. After this classification, we coded whether the identified effects were positive, negative, or not significant. Since we were interested in exploring the potential of EMME, we used a liberal coding approach and scored mixed effects on various outcome variables from one category that also contained positive effects as positive. If no positive effect was available, we reported on negative effects (which, thus, outweighed non-significant effects). We used a significance level of 0.05. For the sake of completeness, we finally also coded which other types of dependent variables the included experiments used.

The specific coding instructions ("code book") can be found in the supplementary materials and the table that includes all coded data is available on https://doi.org/10.17026/dans-z8j-f97h (Open Universiteit Nederland, 2022). Furthermore, four tables that include all reported data (one for the general study information, and one for each research question), can be found in the supplementary materials. This data was analyzed and summarized to create the main results of this study by calculating absolute and relative frequencies for each variable and its categories, as well as

relevant contingency tables. The analyses were performed in JASP (2020 version 0.14.1.0). In the discussion, we provide a narrative synthesis of the main results.

Results

General Study Information

In total, the 72 identified studies included 78 experiments (See Appendix 2). All of these studies are included in the reference list and are highlighted with a * symbol. A total of 63 studies (87.50%) were published in peer-reviewed journals and 9 (12.5%) were published in conference proceedings. The experiments had an average sample size of $M_{sample} = 45.45$ ($SD_{sample} = 35.71$) participants and most (92.31%) used a between-subjects design with an average sample size of $M_{sample\ per\ condition} = 19.07$ ($SD_{sample\ per\ condition} = 14.43$) per condition. In 48 experiments (61.54%), the control group differed in more than the presence or absence of the gaze visualization (e.g., when the gaze visualization was part of a larger intervention and, thus, not the only manipulation or when the control group differed in more than the absence of the gaze visualization).

The quality assessment of the included studies yielded an average MERSQI score of $M_{MERSQI} = 13.37$ ($SD_{MERSQI} = 1.71$ with a range from 9 to 17). Overall, strengths of the studies were the objectivity of the measurements, and the frequent use of (pseudo)randomized controlled trials as study design. Furthermore, we concluded that the analysis methods were often appropriate and went beyond a merely descriptive analysis: 79.5% of all studies reported descriptive statistics and appropriate analyses, as well as fitting conclusions. Twenty-three studies also analyzed effects on long-term learning (after at least one day). A major limitation of various studies was the (reported) validity of the measurement instruments. In terms of sampling method, only few studies based their sample sizes on a power analysis or previous effect sizes.

Background and Rationales of the Included Articles (RQ1)

The following results are based on data from all 72 published studies (see Appendix 3). Table 2 summarizes the results by displaying the absolute and relative frequencies of different domains of the task materials, the mentioned rationales for using EMME, and the terminology used in the publication. The included studies used materials from the domains sports, medicine, and STEM fields (i.e., programming, mathematics, physics, biology) about equally often. Some studies also used task materials from the field of aviation and other fields. The majority of the studies justify the use of EMME by arguing that such displays provide information about the task performer's visual or cognitive strategies (such as an expert's Quiet Eye) and half of the studies mentioned attention guidance. Additionally, several studies mentioned social mechanisms as rationale (i.e., potentially underlying mechanism) for the use of EMME to foster learning or performance. Commonly used terms were

Table 2 Absolute and Relative Frequencies of the Variables in the Context of RQ 1 (i.e., Domains of the Task Materials, the Mentioned Rationales for Using EMME, and the Terminology Used)

Variable	Category	Absolute frequency	Relative frequency
Domain	Sports/PE	19	26.39%
	Medicine	17	23.61%
	Stem	17	23.61%
	Aviation	6	8.33%
	Other	13	18.06%
Rationale[a]	Attention guidance	39	54.17%
	Social mechanisms	15	20.83%
	Strategy	61	84.72%
	QE[b]	23	31.94%
Terminology	EMME	23	31.94%
	QE(T)	19	26.39%
	Gaze(-based) training	7	9.72%
	Other	23	31.94%

EMME, Quiet Eye Training, and gaze-(based) training. However, about a third of all studies also used other terms in their titles and abstracts (e.g., feedforward gaze training, point-of-regard display, eye-gaze cursor, gaze-augmented think aloud, gaze path cueing, visual attention training).

The contingency table (Table 3) furthermore provides information on the connection between the domain of task materials and the rationales for displaying EMME. Studies that focus on sports/physical education mostly mention that EMME can reveal and teach a quiet eye strategy. Some medical studies also focus on revealing a performer's strategy, others (additionally) mention attention guidance as rationale. Finally, studies that deal with tasks from STEM domains make use of a variety of rationales for using EMME and refer more frequently to social mechanisms than studies from other domains.

Studies that focus on sports/physical education mostly mention that EMME can reveal and teach a quiet eye strategy. In the field of aviation, the studies also often mention the goal of teaching a strategy, but without referring to the quiet

Table 3 Absolute Frequencies of how often Studies using Materials from Different Domains Report the Identified Rationales for Using EMME

Domain	Attention guidance	Social mechanisms	Strategy	QE
Sports/PE	1	0	19	18
Medicine	12	2	14	3
STEM	15	9	11	0
Aviation	2	1	6	0
Other	9	3	11	2
Total	39	15	61	23

eye. Medical studies focus on revealing a performer's strategy as well as attention guidance as rationale. Finally, studies that deal with tasks from STEM domains make use of a variety of rationales for using EMME and refer more frequently to social mechanisms than studies from other domains. Most tasks that fell into the other categories argued that EMME can be used to teach and reveal a performer's strategy.

Study Characteristics (RQ2)

The following results are based on data from all 78 individual experiments of the identified studies (see Appendix 4). Table 4 summarizes the results in the context of RQ 2 by displaying the absolute and relative frequencies of the analyzed variables. In terms of participant characteristics, most experiments examine the effects of EMME on participants from tertiary education with no or low reported prior knowledge. The task types were relatively balanced across experiments, but studies on problem-solving were relatively less frequent. Regarding the EMME display characteristics, most of the experiments added information to the display with the EMME, used a dynamic gaze visualization, and showed a performers' scan path. Over half of the experiments did not report how they processed the eye movement data to create the EMME. In terms of observer guidance, the performer was in most cases a teacher or expert that was more often instructed to behave didactically or in a scripted manner than naturally. However, in about half of the experiments, the observer instruction was not explicitly specified. Over half of the identified studies used additional auditory guidance to support observers' understanding. The percentage of studies that found at least one positive effect of EMME on immediate learning outcomes are reported in the last column of Table 4 and will be discussed in detail in the discussion section.

Outcome Variables and Results (RQ3)

The following results are based on data from the 78 individual experiments of the identified studies (see Appendix 5). Table 5 provides an overview of the effects on displaying EMME on gaze, learning, and performance outcomes. Figure 3 visualizes these findings by providing an overview of the absolute number of experiments that found a positive, a negative, or no effects on the different categories of the main dependent variables. The results show that most studies investigated effects on immediate post-test scores and participants' gaze after the EMME was presented (immediate learning outcomes). Fewer experiments used performance measures or measures to investigate delayed learning. In terms of outcome variables, scores were most frequently used, followed by gaze measures. Only few studies investigated effects on time on task. For over half of the selected outcome measures (i.e., performance score, gaze during the EMV, gaze during the post-test, delayed learning-test score, and gaze during a delayed-test), the studies did not report negative effects. The three gaze outcome measures showed the highest relative rates of positive EMME effects.

Table 4 Summary of Absolute and Relative Fequencies of the Variables in the Context of RQ 2

Variable	Category	Absolute frequency	Relative frequency	Relative frequency of a positive effect on immediate learning[1] (score)
Participant characteristics: prior knowledge	No or low prior knowledge	43	55.13%	71.86%
	Some prior knowledge	22	28.21%	62.50%
	Mixed	8	10.26%	83.33%
	Not specified	5	6.41%	75.00%
Participant characteristics:	Primary and secondary education	12	15.38%	66.67%
Education level of participants	Tertiary education	46	58.97%	67.57%
	Professionals	8	10.26%	75.00%
	More than 1 group	6	7.69%	100.00%
	Others	6	7.69%	80.00%
Type of task	Visual/classification task	26	33.33%	73.68%
	Visuo-motor task	25	32.05%	82.36%
	Problem-solving	7	8.97%	16.67%
	Text processing/comprehension	13	16.67%	70.00%
	Other	7	8.97%	83.33%
Gaze visualization	Adding information	66	84.61%	68.00%
	Removing information	4	5.13%	66.67%
	Adding and removing information	3	3.85%	100.00%
	Not reported whether information was added or removed	5	6.41%	100.00%
	Static visualization	5	6.41%	0.00%
	Dynamic visualization	70	89.74%	72.73%
	Static and dynamic visualization	3	3.85%	50.00%
	Scan path	48	61.54%	63.16%
	Attention map	11	14.10%	75.00%
	Scan path and attention map	3	3.85%	100.00%
	Other	1	1.28%	0.00%
	Visualization type not reported	15	19.23%	90.00%

Table 4 (continued)

Variable	Category	Absolute frequency	Relative frequency	Relative frequency of a positive effect on immediate learning[1] (score)
	Visualization based on raw eye movement data	2	2.56%	0.00%
	Visualization based on event-detected eye movement data	18	23.08%	64.29%
	Visualization based on higher-order data	8	10.26%	100.00%
	Visualization based on different types of data	2	2.56%	100.00%
	Data type for visualization not reported	48	61.54%	72.97%
Guidance in materials: Performer type	Novice or peer	6	7.69%	0.00%
	Teacher or expert	48	61.54%	77.14%
	Researcher	14	17.95%	76.92%
	More than one EMV	3	3.85%	0.00%
	Not specified	7	8.97%	57.14%
Guidance in materials: Performer instruction	Natural	13	16.67%	85.71%
	Didactic or scripted	24	30.77%	56.52%
	Not specified	41	52.56%	78.57%
Guidance in materials: Additional guidance tools	Auditory guidance available	37	47.44%	63.33%
	Auditory guidance not available	41	52.56%	78.57%
	Other guidance available	12	15.38%	66.67%
	Other guidance not available	66	84.61%	71.43%

[1]Only studies that used measures of immediate learning were included in this analysis

Table 5 Summary of the Main Findings of the Experiments in the Context of RQ 3

	Positive effect	Negative effect	No effect	Not investigated
Performance score	8 (10.26%)	0 (0.00%)	3 (3.85%)	67 (85.90%)
Performance time on task	1 (1.28%)	1 (1.28%)	4 (5.13%)	72 (92.31%)
Performance gaze	12 (15.38%)	0 (0.00%)	2 (2.56%)	64 (82.05%)
Immediate learning score	41 (52.56%)	4 (5.13%)	13 (16.67%)	20 (25.64%)
Immediate learning time on task	8 (10.26%)	3 (3.85%)	5 (6.41%)	62 (79.49%)
Immediate learning gaze	40 (51.28%)	0 (0.00%)	2 (2.56%)	36 (46.15%)
Delayed learning score	13 (16.67%)	0 (0.00%)	8 (10.26%)	57 (73.08%)
Delayed learning time on task	2 (2.56%)	1 (1.28%)	1 (2.38%)	74 (94.87%)
Delayed learning gaze	14 (17.95%)	0 (0.00%)	3 (3.85%)	61 (78.21%)
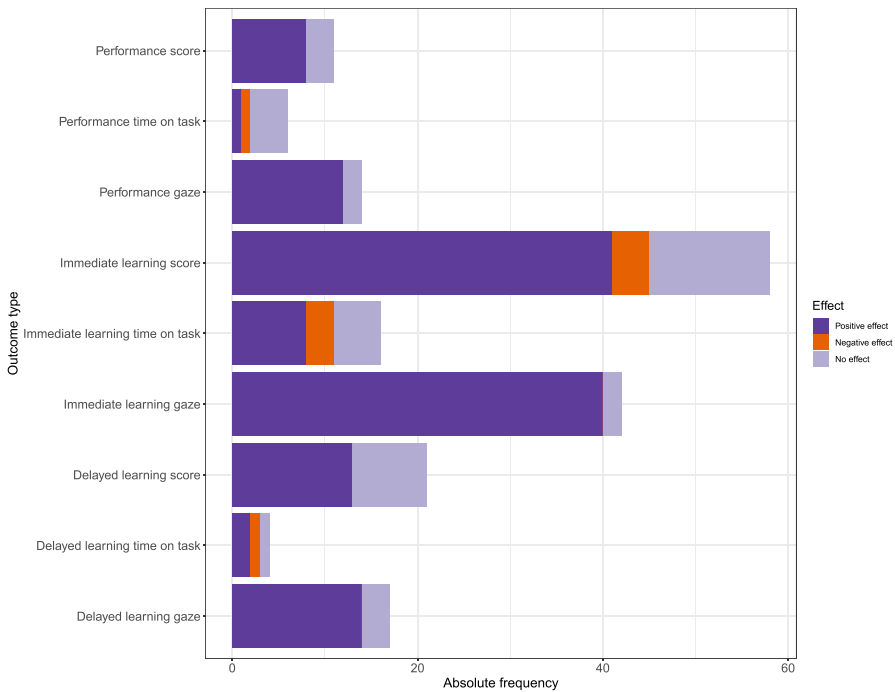
Fig. 3 Absolute frequency of studies that found a positive, negative, or no effect split by outcome type

Twenty-three experiments (29.49%) additionally reported on subjective ratings, 9 experiments (11.54%) on participants' qualitative responses, and 26 experiments (33.33%) investigated effects on other measures (e.g., EEG measures, movements, or mouse clicks).

Discussion

This systematic literature review provides a detailed overview of studies that examined the effects of visualizing a performer's gaze to foster observers' learning and performance. We use the term EMME here to refer to such gaze visualizations, but we used a variety of search terms to identify and include related literature from other fields by authors who used other terms.

In total, we identified 72 articles (which included 78 studies) on the effects of EMME. By using broader search terms, not limited to the use of the term EMME, our systematic review connects EMME research from the field of educational sciences with other relevant literature on the use of gaze displays to foster learning and performance (i.e., it is substantially more extensive than the meta-analysis by Xie et al. (2021), which included 25 articles using the term EMME). In the following sections, we summarize and discuss the results of our three main research questions and their implications for future research.

What is the Background of the Identified Studies?

Our first research question (RQ1) concerned the background of the included studies, such as the general domain of the tasks taught, the rationales for using EMME, and the term to refer to the EMME method. The most commonly used terms, besides the term of EMME, were Quiet Eye Training and gaze-(based) training. However, we found that about a third of all studies used other terms (e.g., feedforward gaze training, gaze-augmented think aloud, gaze path cueing, visual attention training).

Since the QET is such a large part of the literature identified, we discuss it in somewhat more detail. QET originates from research in sports such as basketball (Vickers, 1996). Elite sportsmen have showed to exhibit the quiet eye, which is defined as "a final fixation or tracking gaze that is located on a specific location or object in the visuomotor workspace within 3° of visual angle for a minimum of 100 ms" (Vickers, 2007, p. 11). The quiet eye occurs prior to the final movement of the task, in sports like golf, basketball, volleyball, and darts, but also in medical tasks such as surgical knot tying. The underlying assumption is that QE is a critical period when sensory information is synthesized to both plan and control a motor response (Miles et al., 2017). The assumption is that this quiet eye strategy can be taught to less-skilled participants in Quiet Eye Training with gaze displays, and such training has been found to alter novices' gaze and foster their performance on subsequent visuomotor tasks (Wilson et al., 2015). QET is often contrasted with a technical training in which no instructions regarding gaze are given. It usually includes a set of instructions (derived from expertise research for the specific task) combined with practice trials and feedback (Vickers, 2007; Wilson et al., 2015). In literature on QET, the EMME video is typically an aspect of the full training (i.e., not the main manipulation) aimed at modeling the expected gaze behavior to the novice and contrasting the quiet eye behavior of the novice and the expert (Wilson et al., 2015).

We found that the EMMEs are primarily applied for teaching learners tasks in the domains of sports/physical education, medicine, aviation, and STEM education (i.e., programming, mathematics, physics, biology). Studies in different domains seemed to have different rationales for using EMME. Studies that focused on teaching physical tasks from the domain of sports/physical education often aimed to reveal and teach the performer's (usually an expert's) perceptual strategies, such as a quiet eye (Vickers, 2007). Studies that taught tasks in the domain of medicine also mostly reported the goal to teach expert strategies, but additionally mentioned the use of EMME to guide observers' attention. Studies in STEM education had a broad variety of rationales and mentioned social mechanisms of EMME more frequently than studies of medicine or sports/physical education. Studies that taught tasks from other domains mostly referred to the goal of teaching and revealing the task performers' strategies to observers. To conclude, we identified studies that used different rationales for EMME, terminologies, and applied EMME in different contexts (e.g., sports, medicine, aviation, STEM).

The large variety of study backgrounds (e.g., in terms of terminologies and rationales used) indicates that there is a wide range of different approaches and research traditions on the topic of EMME (in a broader sense) that have not always been well connected to date. Our systematic literature review is a first step towards a better connection of the different research fields that could ultimately contribute to a broader knowledge of the effectiveness of EMME in different situations and, thus, fundamentally expand the knowledge of EMME research. For example, the cluster of studies on Quiet Eye Training (Vickers, 2007) is relatively unconnected to EMME research from the field of educational sciences but they are similar in their use of gaze visualizations to foster performance and learning. By including studies that use a variety of terms in our overview, we provide a more complete understanding of applications and boundary conditions for the use of EMME (in a broader sense). To unify research fields and traditions further, it is important that future studies use the same terminology.

To What Extent Do the Identified Experiments Differ in Terms of Study-Design Characteristics?

RQ2 asked to what extent the studies differed in terms of participant characteristics, task type, and EMME materials. In contrast to previous work (Krebs et al., 2019; Xie et al., 2021), our goal was not to draw conclusions on the effects of EMME in general or significant moderator variables based on statistical effect-size analysis. The heterogeneity of the identified studies is too large for such an analysis. Instead, we provided a comprehensive overview of the diversity of the published EMME literature by providing a list of categories in which the identified studies differ. In addition, we reported the relative frequency of studies that found a positive EMME effect on immediate learning outcomes (see last column of Table 4) for studies with different design choices. This can serve as a first basis for future studies to decide on their EMME design and possibly inspire series of systematic experiments or further meta-analyses on a selection of moderating variables.

Participant Characteristics

Although studies differed in terms of participants' prior knowledge, the identified articles mostly recruited participants from tertiary education and participants with no or low prior knowledge about the topic that was taught in the EMME. About 72% of the included studies with learners with no prior knowledge found positive effects of EMME on immediate learning, compared to 60% of studies using EMME with learners who had higher levels of prior knowledge. This overall finding could be linked to literature on the expertise reversal effect, which states that design principles that are effective for novice learners might not be effective for more experienced learners (Kalyuga, 2009, 2014) and the finding that signaling particularly supports low prior knowledge learners (Richter et al., 2016, 2018).

Research on the expertise-reversal effect has shown that instructional methods that enhance learning for beginners may become inefficient for more experienced learners (Kalyuga, 2009). For instance, learners with no or low prior knowledge seem to benefit more from cues in multimedia materials (e.g., color highlights) to emphasize relevant information than learners with higher prior knowledge (Richter et al., 2016). This is probably because cues can help novice learners to select the right information from the task materials in time. However, learners with higher prior knowledge know what to focus on, so additional visual cues to guide attention are not necessary. In fact, such cues might even hamper the processing of relevant task elements for advanced learners, since the unnecessary visualization of the eye movements might be hard to ignore and hence distract them.

However, support for a more positive impact of EMME on learners with low versus high prior knowledge is not consistent in the literature. Krebs et al. (2019), Litchfield et al. (2010), and Van Marlen et al. (2018) found evidence that EMMEs are more beneficial to foster learning for observers with low prior knowledge than high prior knowledge. Other studies did not find this effect (Chisari et al., 2020; Krebs et al., 2021; Wang et al., 2020), or even more beneficial effects for observers with higher prior knowledge (Gegenfurtner et al., 2017; Scheiter et al., 2018). Indeed, the moderating effects of prior knowledge on the effectiveness on EMME failed to reach statistical significance in the meta-analyses of Xie et al. (2021) and Krebs et al. (2021). In agreement with Krebs et al. (2021), we assume that this inconsistency might be caused by different definitions of prior knowledge and further research is necessary to investigate which learners benefit most from EMME.

Task Type

The different task types in the included studies could be categorized as visual/classifying tasks, problem-solving tasks, visuo-motor tasks, text processing/comprehension tasks, and other tasks. The relative frequency of finding a positive EMME effect on immediate learning outcomes seemed to vary across these task types. Seventeen percent of studies that taught a problem-solving task found positive effects of EMME on learning. In contrast, the relative frequency of finding a positive EMME effect ranged between 70 and 82% for visual/classification tasks, visuo-motor tasks, or text processing/comprehension tasks (e.g., text-picture integration).

For tasks of the latter categories, the processing of visual information is likely more relevant than for procedural problem-solving tasks, because the steps to solve such problem-solving tasks are often also visible to the observer in other ways, e.g., through the actions of the performer on the screen (as argued by Van Gog et al., 2009; Van Marlen et al., 2016). In this context, it should be noted, however, that prior knowledge might affect how observers learn procedural-problem solving tasks with EMME, which should be investigated in future research (Van Marlen et al., 2018). In conclusion, learners working on tasks with a higher visual component are likely to benefit more from EMME, and that EMMEs are in general less suited for teaching (procedural) problem solving tasks.

EMME Materials

Gaze Visualizations Although all EMME display a performer's eye movements, there are large differences in how studies visualized this gaze. To date, the impact of visualization choice has not been widely studied. In fact, several of the identified studies did not even provide detailed information about how they visualized the gaze. For example, about a fifth of all study did not report which visualization type they used. Furthermore, gaze visualizations could be based on raw or processed data (e.g., fixation based) with affects the amount of detail and noise in the visualizations. Over half of the experiments did not report how they processed the eye movement data to create the EMME.

The majority of the experiments that indicated how they visualized the EMME reported that they added information to the display through the EMME, used a dynamic rather than static gaze visualization, and showed a performers' scan paths rather than attention maps that were based on event-detected data. It is likely that these decisions are attributable to the default settings of standard eye-tracking software, as evidence-based guidelines for gaze visualizations are often lacking to date. Since the vast majority of gaze visualizations are similar on the dimensions that we report on, it is difficult to draw conclusions on the effects of gaze visualization techniques. Systematically investigating gaze visualization techniques is important, as different gaze visualization techniques have different affordances (Blascheck et al., 2014) and this could impact how observers interpret them (Bahle et al., 2017; Kurzhals et al., 2015).

Only three studies (Brams et al., 2021; Jarodzka et al., 2012, 2013) systematically investigated the effects of using a scan path (superimposed circle) visualization or attention map (blurred background) visualization for teaching a perceptual strategy. Jarodzka et al. concluded that the attention map visualization was more effective, presumably, because it reduced the amount of (redundant) visual information on the screen, which could help learners to select relevant information with less effort (Kalyuga & Sweller, 2014; Mayer & Moreno, 2003). However, Brams et al. (2021) were unable to replicate the beneficial effects of the attention map visualization and observed subtle benefits of the scan path visualization using static stimuli materials. A possible, first explanation could be that peripheral vision might be more relevant when observing moving stimuli (as argued by Brams et al., 2021). This would mean that attention map visualization might be more beneficial for dynamic learning task material (e.g., video, animation) than for static and vice versa. These studies form

an interesting starting point for further research to investigate whether different gaze visualizations affect observers' interpretation of the displayed performance (e.g., Van Wermeskerken et al., 2018), and whether they have different effects on observers' performance, learning, and gaze measures. However, it is important to keep in mind that different types of visualization might be more or less effective depending on the specific characteristics of the learning materials and task requirements (e.g., Kok et al., 2023).

Performer Type and Behavior, and Additional Elements of Attention Guidance Differences were also found between studies in the performers who created the EMME materials (i.e., performer expertise) and how they were instructed (i.e., instruction to behave naturally vs. didactically). More specifically, few EMME studies used peers as performers, and the studies that did, did not report positive EMME effects, whereas studies that used a more experienced performer more often found beneficial effects of EMME. One explanation could be that showing expert gaze guides attention more effectively, because experts tend to look faster and more at relevant information then novices (e.g., Emhardt et al., 2020; Reingold & Sheridan, 2011; Sheridan & Reingold, 2017).

Another observation regarding the EMME performers was that they were mostly teachers or experts who received the instruction to behave didactically (rather than naturally). Such instructions may play a role in the effectiveness of the EMME. When it comes to natural performance, the task performer's expertise may matter for the observers' learning, as expert performers show substantially different, often automatized task-solving behavior than novice performers (Boshuizen & Schmidt, 2008; Ericsson et al., 2018), which might be more difficult for novices to follow. On the other hand, it feels sensible to instruct performers when creating EMME to perform a task in a didactic way that is understandable to an inexperienced audience (e.g., Jarodzka et al., 2012, 2013), rather than in a natural manner (e.g., Litchfield et al., 2010; Nalanagula et al., 2006; Seppänen & Gegenfurtner, 2012; Stein & Brennan, 2004).

Emhardt et al. (2020) recently showed that the visual characteristics of EMME displays (i.e., the performer's eye movements and mouse clicks) change with such instructions. To date, it is not known how these different types of performer behaviors affect learning outcomes. However, the results of our systematic review could serve as a basis for future research on this topic. We observed that only 57% of the studies with didactic actor behavior, and 86% of the studies with natural actor behavior found at least some positive effects of EMME on immediate learning. One first, possible explanation for this could be that natural behavior demonstrates authentic (expert) behavior that can be learned and subsequently adopted though observation (Bandura, 1977). Note that seems particularly useful if EMMEs aim to convey expert strategies and potentially less so if EMMEs aim to guide attention.

A final point of interest regarding instructional design choices is the impact of adding other elements of attention guidance to EMME materials. In over half of the identified studies, the EMME materials included additional auditory explanations, which also provide guidance, for instance though verbal references to locations (D'Angelo & Begel, 2017). Studies reported positive effects of EMME on learning

in 63% of studies with additional auditory instruction, but in 78% of studies without auditory guidance. A similar trend was identified in the meta-analysis of Xie et al. (2021). A smaller effect might be present for other guidance elements (like the mouse-cursor visible). In this context, we found positive effects of EMME on learning in 67% of the studies with other guidance elements and 71% of studies without other guidance. One explanation for finding positive effects of EMME on learning somewhat less often in the presence of other attention guidance elements could be that EMMEs have more beneficial effects when the information conveyed in the gaze visualization is not redundant to the information conveyed by other video elements (e.g., a mouse cursor, Sweller, 2005; Van Gog et al., 2009). Note, furthermore, that if the EMME is part of a larger intervention that also includes other elements of attention guidance such as added auditory explanations or a mouse cursor, it might be difficult to distinguish effects of gaze visualizations from that of other forms of guidance, in particular if they were not systematically manipulated. Given these considerations, it is important that future studies systematically examine the benefits of EMME and compare the effects of different EMME designs on learning outcomes. The findings of the preceding sections show initial insights into effects of possible design choices and could serve as a stimulus and initial guide for future studies. Ultimately, these investigations will help researchers and educators make more informed decisions about the costs and benefits of creating EMME in different manners.

Which Types of Effects Did the Identified Studies Explore?

RQ3 asked which types of effects the identified studies explored. As with RQ2, we found a great deal of heterogeneity in the outcome measures. We distinguished between measures of performance and immediate and delayed learning and the type of outcome variable (i.e., scores, time on task, or gaze variables). We observed that most experiments investigated effects on observers' immediate learning outcomes. The effects on delayed learning and observers' performance have been studied less frequently, although delayed learning outcomes have a high practical relevance, and more studies should include this type of outcome measures. Although gaze guidance is a commonly used rationale for the use of EMME, it is not often investigated whether EMMEs actually direct observers' attention (gaze) while they are studying the EMME. Overall, Fig. 3 shows that the overwhelming majority of studies report at least some positive effects of EMME on the outcome variables, which shows the large potential of EMME and is in line with the conclusion of the EMME meta-analysis of Xie et al. (2021) that EMMEs overall foster learning and performance. In our results, it is furthermore striking that no study reported negative effects of EMME on gaze measures while and after studying the EMME. We cannot exclude the possibility that these interpretations were made post hoc, which means that the observers' changes in eye movement patterns might have been reinterpreted as being positive after observing the changes, without any (prior) theoretical foundation. Furthermore, it cannot be ruled out that such predominantly positive findings are

also partly due to a publication bias in EMME literature. Desirable effects of EMME on gaze patterns should be defined based on the characteristics of the task taught (cf. Jarodzka & Boshuizen, 2017) and should be defined in advance for the specific task materials. Expertise studies could help to define desired effects in advance (as was the case, for example, in the study of Kok et al., 2016).

Several studies found negative effects of EMME, for example causing participants to take more time to execute the task (Litchfield et al., 2010) or take longer during a posttest (Mehta et al., 2005; Salmerón & Llorens, 2018; Van Marlen et al., 2016, experiment 2; Van Marlen et al. 2018, experiment 1). However, it is an open question whether these are actually negative effects; For instance, it can also be the case that participants took longer to execute a task because they attempted to deliberately apply the strategy taught in the EMME (regardless of whether they succeeded at doing so and this had a beneficial effect on performance or not). For example, Salmerón and Llorens (2018) argue that the EMME showed the importance of deep and slow reading of texts that are relevant for the learners' goal, and that this caused longer reading times. Indeed, participants' posttest performance benefitted from the EMME. Thus, effects on time on task should be interpreted in the context of effects on learning and performance.

Additionally, some studies found negative effects of EMME on learning (Carroll et al., 2013; Eder et al., 2020; Scheiter et al., 2018; Van Gog et al., 2009). In those studies, the authors argued that the effects of the EMME did not align with the needs of (all) learners (cf. the redundancy effect, Kalyuga & Sweller, 2014). For example, Van Gog et al. (2009) argue that the EMME turned out to be redundant or even caused split attention effects, and thus hampered learning. Likewise, Scheiter et al. (2018) argued that the EMME hampered simple memorization, in particular for weaker students because the information was redundant for this goal. It is thus important to consider the (presumed) added value of EMME for tasks and for the learner.

In general, future researchers should be aware that effects on learning and performance measures have different practical implications. For instance, effects of EMME on performance measures imply changes in how observers process the instructional material. Litchfield et al. (2010) for example showed that learners take longer to execute the task in the presence versus absence of the EMME, and Winter et al. (2021) found that the EMME helped observers to execute the task. Conversely, effects of EMME on learning measures imply that the altered processing of the instructional material has longer-lasting consequences for the memory and task performance, showing later on in the absence of the EMME (see also Soderstrom & Bjork, 2015). We argue that researchers should align the type of measures with the objective of the study and its rationale. For example, measuring learners' gaze while studying the EMME is useful if the EMME aims to direct attention, as it provides information about whether this goal is being achieved (as, e.g., in Chisari et al., 2020; Jarodzka et al., 2013; Van Marlen et al., 2016, 2018). If the EMME aims to convey expert strategies, measuring gaze during a posttest makes sense as well, because it clarifies whether learners have learned to adopt expert's strategies (e.g., Brams et al., 2021; Carroll et al., 2013; Eder et al., 2020; García et al., 2021; Gegenfurtner et al., 2017; Litchfield et al., 2010; Mason et al., 2015, 2017).

Limitations of the Present Literature Review

This systematic literature review also had some limitations. First, it included studies in which the gaze visualization was not the only difference between the EMME and the control condition. This was mainly the case in studies on QET, which frequently embedded an EMME into a larger intervention program, and used different types of training interventions in the control groups (e.g., Causer et al., 2014a, b; Harle & Vickers, 2001; Hosseini et al., 2021; Jacobson et al., 2021; Miles et al., 2017; Norouzi et al., 2019; Vine & Wilson, 2010). While such conditions may represent authentic training interventions, they do not allow for direct conclusions about the added value of EMME over regular modeling examples without gaze visualizations.

When interpreting our results, it should furthermore be kept in mind that we always reported a positive effect of EMME on each type of outcome variable if at least one of the outcome measures per category showed a significant positive effect of EMME.

Finally, this review did not include studies that investigate the effects of displaying a communication partner's eye movements to foster (live) communication and cooperation or competition. We excluded these studies because the EMME representation in such situations is likely to change depending on what the participants are doing. This makes it difficult to distinguish between the effects of interaction and the effects of gaze visualizations. However, in our search, we came across a relatively large number of studies on sharing gaze in real time to foster communication and interaction/competition (e.g., Bai et al., 2020; Brennan et al., 2008; D'Angelo & Begel, 2017; Gupta et al., 2016; Müller et al., 2013; Niehorster et al., 2019; Piumsomboon et al., 2019). In this context, an avenue for future research is to combine the approach of live gaze sharing with teaching by projecting a teacher's gaze onto a classroom wall to foster their students' learning (as, e.g., suggested by Špakov et al., 2016), which could be an interesting extension to traditional EMME research.

General Conclusion

This systematic literature review identified 72 studies (78 experiments) that investigated the effectiveness of EMME in an educational context. The identified studies used a variety of terms to refer to EMME interventions (e.g., EMME, QET, gaze(-based) training), were performed in different domains (e.g., aviation, medicine, sports, and STEM education), and used a variety of rationales to refer to the EMME. The aim of this systematic review was to connect different lines of research and provide a general overview on the existing EMME studies. We concluded that most studies found at least some positive effects, which underlines the potential benefit of adding EMME to instructional materials. However, we argued that different study design choices, such as the type of task, participant sample, or the decision to display a naturally behaving performer (vs. the behavior of a scripted performer) might affect the effectiveness of EMME. Based on our findings, EMME seem to be particularly promising for teaching tasks with a strong visual component (and less for problem-solving tasks) with instructional materials with few other elements of

didactic guidance and they might be more effective for learners with limited prior knowledge. Several guidelines for development of EMME appear to be supported by the data, but a systematic comparison of the effect of each of those choices should be undertaken before those guidelines can be used in practice. It should furthermore be noted that design choices should be aligned with the specific task and learning goals. That being said, it seems that EMME that used more experienced performers are more effective than EMME with peer performers, and EMMEs in which the model behaved naturally rather than didactically were more effective for immediate learning. There is still limited evidence on the effects of different gaze visualizations, but there is some evidence that the type of display may affect different aspects of learning differently, so the type of visualization used should be carefully considered. The overview provided by this systematic review, however, provides a first basis for future research on the effects of such design choices, by highlighting areas where more systematic investigations of influencing variables are needed.

References

*Adolphe, R. M., Vickers, J. N., & Laplante, G. (1997). The effects of training visual attention on gaze behaviour and accuracy: A pilot study. *International Journal of Sports Vision, 4*(1), 28–33.

*Ahmadi, N., Romoser, M., & Salmon, C. (2022). Improving the tactical scanning of student pilots: A gaze-based training intervention for transition from visual flight into instrument meteorological conditions. *Applied Ergonomics, 100*, 103642. https://doi.org/10.1016/j.apergo.2021.103642

*Ahrens, M., Schneider, K., & Busch, M. (2019). Attention in software maintenance: An eye tracking study. *Proceedings of the 6th International Workshop on Eye Movements in Programming (EMIP)*, 2–9. https://doi.org/10.1109/EMIP.2019.00009

*Ahrens, M., & Schneider, K. (2021). Improving requirements specification use by transferring attention with eye tracking data. *Information and Software Technology, 131*, 106483. https://doi.org/10.1016/j.infsof.2020.106483

Aizenman, A., Drew, T., Ehinger, K. A., Georgian-Smith, D., & Wolfe, J. M. (2017). Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: An eye tracking study. *Journal of Medical Imaging, 4*(4), 045501. https://doi.org/10.1117/1.JMI.4.4.045501

Bahle, B., Mills, M., & Dodd, M. D. (2017). Human classifier: Observers can deduce task solely from eye movements. *Attention, Perception, & Psychophysics, 79*(5), 1415–1425. https://doi.org/10.3758/s13414-017-1324-7

Bai, H., Sasikumar, P., Yang, J., & Billinghurst, M. (2020). A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376550

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158).

Bandura, A. (1977). *Social learning theory*. Prentice-Hall.

*Bednarik, R., Schulte, C., Budde, L., Heinemann, B., & Vrzakova, H. (2018). Eye-movement modeling examples in source code comprehension: A classroom study. *Proceedings of the 18th Koli Calling International Conference on Computing Education Research*, 22–25. https://doi.org/10.1145/3279720.3279722

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.019

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014). *State-of-the-art of visualization for eye tracking data* Eurographics Conference on Visualization (EuroVis) Swansea, Wales. https://doi.org/10.2312/eurovisstar.20141173

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind & Language, 17*(1–2), 37–54. https://doi.org/10.1111/1468-0017.00188

Boshuizen, H. P., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–121). Butterworth-Heinemann.

*Brams, S., Ziv, G., Hooge, I. T. C., Levin, O., Verschakelen, J., Williams, A. M., Wagemans, J., & Helsen, W. F. (2021). Training focal lung pathology detection using an eye movement modeling example. *Journal of Medical Imaging, 8*(2), Article 025501. https://doi.org/10.1117/1.JMI.8.2.025501

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*(3), 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

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

*Carroll, M., Kokini, C., & Moss, J. (2013). Training effectiveness of eye tracking-based feedback at improving visual search skills. *International Journal of Learning Technology, 8*(2), 147–168. https://doi.org/10.1504/IJLT.2013.055671

*Castner, N., Geßler, L., Geisler, D., Hüttig, F., & Kasneci, E. (2020). Towards expert gaze modeling and recognition of a user's attention in realtime. *Procedia Computer Science, 176*, 2020–2029. https://doi.org/10.1016/j.procs.2020.09.238

*Causer, J., Harvey, A., Snelgrove, R., Arsenault, G., & Vickers, J. N. (2014a). Quiet eye training improves surgical knot tying more than traditional technical training: A randomized controlled study. *The American Journal of Surgery, 208*(2), 171–177. https://doi.org/10.1016/j.amjsurg.2013.12.042

*Causer, J., Vickers, J. N., Snelgrove, R., Arsenault, G., & Harvey, A. (2014b). Performing under pressure: Quiet eye training improves surgical knot-tying performance. *Surgery, 156*(5), 1089–1096. https://doi.org/10.1016/j.surg.2014.05.004

*Cheng, S., Sun, Z., Sun, L., Yee, K., & Dey, A. K. (2015). Gaze-based annotations for reading comprehension. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1569–1572. https://doi.org/10.1145/2702123.2702271

*Chisari, L. B., Mockeviciute, A., Ruitenburg, S. K., van Vemde, L., Kok, E. M., & van Gog, T. (2020). Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning, 36*(4), 569–579. https://doi.org/10.1111/jcal.12428

Clarke, A. D., Mahon, A., Irvine, A., & Hunt, A. R. (2017). People are unable to recognize or report on their own eye movements. *The Quarterly Journal of Experimental Psychology, 70*(11), 2251–2270. https://doi.org/10.1080/17470218.2016.1231208

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6245–6290. https://doi.org/10.1145/3025453.3025573

de Koning, B. B., & Jarodzka, H. (2017). Attention guidance strategies for supporting learning from dynamic visualizations. In R. Lowe & R. Ploetzner (Eds.), *Learning from dynamic visualizations: Innovations in research and practice* (pp. 255–278). Springer. https://doi.org/10.1007/978-3-319-56204-9

*Eder, T. F., Richter, J., Scheiter, K., Keutel, C., Castner, N., Kasneci, E., & Huettig, F. (2020). How to support dental students in reading radiographs: Effects of a gaze-based compare-and-contrast intervention. *Advances in Health Sciences Education, 26*, 159–181. https://doi.org/10.1007/s10459-020-09975-w

*Eder, T. F., Scheiter, K., Richter, J., Keutel, C., & Huttig, F. (2021). I see something you do not: Eye movement modelling examples do not improve anomaly detection in interpreting medical images. *Journal of Computer Assisted Learning*, 1–13. https://doi.org/10.1111/jcal.12619

Emhardt, S. N., Kok, E. M., Jarodzka, H., Brand-Gruwel, S., Drumm, C., & van Gog, T. (2020). How experts adapt their gaze behavior when modeling a task to novices. *Cognitive Science, 44*(9), e12893. https://doi.org/10.1111/cogs.12893

Emhardt, S. N., Jarodzka, H., Brand-Gruwel, S., Drumm, C., Niehorster, D. C., & van Gog, T. (2022) What is my teacher talking about? Effects of displaying the teacher's gaze and mouse cursor cues in video lectures on students' learning. *Journal of Cognitive Psychology, 34*(7), 846–864. https://doi.org/10.1080/20445911.2022.2080831

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.https://doi.org/10.1017/9781316480748

Foulsham, T., & Kingstone, A. (2013). Where have eye been? Observers can recognise their own fixations. *Perception, 42*(10), 1085–1089. https://doi.org/10.1068/p7562

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

*Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2018). Learning from where 'eye' remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology, 71*(7), 1526–1534. https://doi.org/10.1080/17470218.2017.1335335

*García, V., Amadieu, F., & Salmerón, L. (2021). Integrating digital documents by means of concept maps: Testing an intervention program with eye-movements modelling examples. *Heliyon, 7*(12), e08607. https://doi.org/10.1016/j.heliyon.2021.e08607

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

Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE Transactions on Visualization and Computer Graphics, 22*(11), 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778

*Harle, S. K., & Vickers, J. N. (2001). Training quiet eye improves accuracy in the basketball free throw. *The Sport Psychologist, 15*(3), 289–305. https://doi.org/10.1123/tsp.15.3.289

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 University Press.

*Hosseini, R. N. S., Norouzi, E., & Soleymani, M. (2021). Effects of Quiet Eye Training on performance of bimanual coordination in children with DCD. *Iranian Journal of Child Neurology, 15*(4), 43–54. https://doi.org/10.22037/ijcn.v15i4.18926

*Jacobson, N., Berleman-Paul, Q., Mangalam, M., Kelty-Stephen, D. G., & Ralston, C. (2021). Multifractality in postural sway supports quiet eye training in aiming tasks: A study of golf putting. *Human Movement Science, 76*, 102752. https://doi.org/10.1016/j.humov.2020.102752

*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*(5), 813–827. https://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://doi.org/10.1016/j.learninstruc.2012.11.004

Jarodzka, H., & Boshuizen, H. P. (2017). Unboxing the black box of visual expertise in medicine. *Frontline Learning Research, 5*(3), 167–183. https://doi.org/10.14786/flr.v5i3.332

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.013

Kalyuga, S. (2009). The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58–80). Information Science Reference. https://doi.org/10.4018/978-1-60566-048-6.ch003

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.028

*Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors, 56*(6), 1150–1162. https://doi.org/10.1177/0018720814523066

*Klostermann, A., Vater, C., Kredel, R., & Hossner, E.-J. (2015). Perceptual training in beach volleyball defence: Different effects of gaze-path cueing on gaze and decision-making. *Frontiers in Psychology, 6*, 1834. https://doi.org/10.3389/fpsyg.2015.01834

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*(1), 189–205. https://doi.org/10.1007/s10459-015-9624-y

Kok, E.M., Jarodzka, H., Sibbald, M. and van Gog, T. (2023). Did You Get That? Predicting Learners' Comprehension of a Video Lecture from Visualizations of Their Gaze Data. *Cognitive Science, 47*, e13247. https://doi.org/10.1111/cogs.13247

Kok, E. M., Aizenman, A. M., Võ, M.L.-H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision, 17*(12), 1–11. https://doi.org/10.1167/17.12.2

*Koury, H. F., Leonard, C. J., Carry, P. M., & Lee, L. M. (2019). An expert derived feedforward histology module improves pattern recognition efficiency in novice students. *Anatomical Sciences Education, 12*(6), 645–654. https://doi.org/10.1002/ase.1854

*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

*Krebs, M.-C., Schüler, A., & Scheiter, K. (2021). Do prior knowledge, model-observer similarity and social comparison influence the effectiveness of eye movement modeling examples for supporting multimedia learning? *Instructional Science, 49*(5), 607–635. https://doi.org/10.1007/s11251-021-09552-7

Kurzhals, K., Burch, M., Blascheck, T., Andrienko, G., Andrienko, N., & Weiskopf, D. (2015). A task-based view on the visual analysis of eye-tracking data. In M. Burch, L. Chuang, B. Fisher, A. Schmidt, & D. Weiskopf (Eds.), *Eye Tracking and Visualization. ETVIS 2015. Mathematics and Visualization.* (pp. 3–22). Springer. https://doi.org/10.1007/978-3-319-47024-5_1

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://doi.org/10.1016/j.edurev.2013.10.001

Lefrançois, O., Matton, N., & Causse, M. (2021). Improving airline pilots' visual scanning and manual flight performance through training on skilled eye gaze strategies. *Safety, 7*(4), 70–100, Article 70. https://doi.org/10.3390/safety7040070

*Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*(3), 251–262. https://doi.org/10.1037/a0020082

Louwerse, M. M., & Bangerter, A. (2010). Effects of ambiguous gestures and language on the time course of reference resolution. *Cognitive Science, 34*(8), 1517–1529. https://doi.org/10.1111/j.1551-6709.2010.01135.x

Marti, S., Bayet, L., & Dehaene, S. (2015). Subjective report of eye fixations during serial search. *Consciousness and Cognition, 33*, 1–15. https://doi.org/10.1016/j.concog.2014.11.007

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

*Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology, 47*(6), 1083–1095. https://doi.org/10.1111/bjet.12271

*Mason, L., Scheiter, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text–picture processing for multimedia comprehension. *Journal of Computer Assisted Learning, 33*(5), 443–460. https://doi.org/10.1111/jcal.12191

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*(2), 419–425. https://doi.org/10.1037/0022-0663.95.2.419

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge Univeristy Press. https://doi.org/10.1017/CBO9780511816819.004

Mayer, R. E. (2014). 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 edition)* (Vol. 16, pp. 345–370). Cambridge Univeristy Press. https://doi.org/10.1017/CBO9781139547369

*Mehta, P., Sadasivan, S., Greenstein, J., Gramopadhye, A., & Duchowski, A. (2005). Evaluating different display techniques for communicating search strategy training in a collaborative virtual aircraft inspection environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(26), 2244–2248. https://doi.org/10.1177/154193120504902606

*Melnyk, R., Campbell, T., Holler, T., Cameron, K., Saba, P., Witthaus, M. W., Joseph, J., & Ghazi, A. (2021). See like an expert: Gaze-augmented training enhances skill acquisition in a virtual reality robotic suturing task. *Journal of Endourology, 35*(3), 376–382. https://doi.org/10.1089/end.2020.0445

*Miles, C. A. L., Wood, G., Vine, S. J., Vickers, J., & Wilson, M. R. (2017). Quiet eye training aids the long-term learning of throwing and catching in children: Preliminary evidence for a predictive control strategy. *European Journal of Sport Science, 17*(1), 100–108. https://doi.org/10.1080/17461391.2015.1122093

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Moore, L. J., Vine, S. J., Smith, A. N., Smith, S. J., & Wilson, M. R. (2014). Quiet eye training improves small arms maritime marksmanship. *Military Psychology, 26*(5–6), 355–365. https://doi.org/10.1037/mil0000039

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*(2), 177–213. https://doi.org/10.1207/S1532690XCI1902_02

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology, 66*(7), 1302–1316. https://doi.org/10.1080/17470218.2012.737813

*Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*(4), 289–300. https://doi.org/10.1016/j.ergon.2005.11.008

Niehorster, D. C., Cornelissen, T., Holmqvist, K., & Hooge, I. (2019). Searching with and against each other: Spatiotemporal coordination of visual search behavior in collaborative and competitive settings. *Attention, Perception, & Psychophysics, 81*(3), 666–683. https://doi.org/10.3758/s13414-018-01640-0

*Norouzi, E., Hosseini, F. S., Vaezmosavi, M., Gerber, M., Pühse, U., & Brand, S. (2019). Effect of Quiet Eye and Quiet Mind Training on motor learning among novice dart players. *Motor Control, 24*(2), 204–221. https://doi.org/10.1123/mc.2018-0116

Open Universiteit Nederland. (2022). Data From a Systematic Literature Review on Eye Movement Modeling Examples, *DANS*, https://doi.org/10.17026/dans-z8j-f97h

Piumsomboon, T., Dey, A., Ens, B., Lee, G., & Billinghurst, M. (2019). The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 5–6. https://doi.org/10.3389/frobt.2019.00005

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA, 298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S.P. Liversedge, I.D. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 767–786). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0029

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

Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text–picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology, 110*(4), 544–560. https://doi.org/10.1037/edu0000220

*Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*, 343–359. https://doi.org/10.1177/0735633117751605

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*(5489), 265–266. https://doi.org/10.1038/253265a0

*Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning from illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology, 88*(1), 80–94. https://doi.org/10.1111/bjep.12175

*Seppänen, M., & Gegenfurtner, A. (2012). Seeing through a teacher's eyes improves students' imaging interpretation. *Medical Education, 46*(11), 1113–1114. https://doi.org/10.1111/medu.12041

Sheridan, H., & Reingold, E. M. (2017). The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology, 8*, 1620. https://doi.org/10.3389/fpsyg.2017.01620

Shvarts, A. (2018). Joint attention in resolving the ambiguity of different presentations: A dual eye-tracking study of the teaching-learning process. In N. Presmeg, L. Radford, M. R. Wolff, & G. Kadunz (Eds.), *Signs of signification* (pp. 73–102). Springer. https://doi.org/10.1007/978-3-319-70287-2_5

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199.

Špakov, O., Siirtola, H., Istance, H., & Räihä, K.-J. (2016). GazeLaser: A hands-free highlighting technique for presentations. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2648–2654. https://doi.org/10.1145/2851581.2892504

*Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. *Proceedings of the 6th International Conference on Multimodal Interfaces - ICMI '04*, 9–15. https://doi.org/10.1145/1027933.1027936

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–167). Cambridge University Press.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454–1463. https://doi.org/10.2307/1130423

Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge Univeristy Press.

*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*(3), 785–791. https://doi.org/10.1016/j.chb.2009.02.007

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

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

Van Wermeskerken, M., Litchfield, D., & van Gog, T. (2018). What am I looking at? Interpreting dynamic and static gaze displays. *Cognitive Science, 42*(1), 220–252. https://doi.org/10.1111/cogs.12484

Van Gog, T. (2021). The signaling (or cueing) principle in multimedia learning. In R. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology)* (pp. 221–230). Cambridge University Press. https://doi.org/10.1017/9781108894333.022

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of experimental psychology. Human Perception and Performance, 22*(2), 342–354. https://doi.org/10.1037//0096-1523.22.2.342

Vickers, J. N. (2007). *Perception, cognition, and decision training: The quiet eye in action*. Human Kinetics.

*Vine, S. J., & Wilson, M. R. (2010). Quiet eye training: Effects on learning and performance under pressure. *Journal of Applied Sport Psychology, 22*(4), 361–376. https://doi.org/10.1080/10413200.2010.495106

Võ, M.L.-H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance, 42*(10), 1477–1481. https://doi.org/10.1037/xhp000026

*Wang, F., Zhao, T., Mayer, R. E., & Wang, Y. (2020). Guiding the learner's cognitive processing of a narrated animation. *Learning and Instruction, 69*, Article 101357. https://doi.org/10.1016/j.learninstruc.2020.101357

Wilson, M. R., Causer, J., & Vickers, J. N. (2015). Aiming for excellence: The quiet eye as a characteristic of expertise. In J. Baker, & D. Farrow, (Eds.), *Routledge handbook of sport expertise* (pp. 22–37). Routledge.

*Winter, M., Pryss, R., Probst, T., & Reichert, M. (2021). Applying eye movement modeling examples to guide novices' attention in the comprehension of process models. *Brain Sciences, 11*(1), 72. https://doi.org/10.3390/brainsci11010072

Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning, 37*(4), 1194–1206. https://doi.org/10.1111/jcal.12568