Learning to see: Guiding students' attention via a Model's eye movements fosters learning

Halszka Jarodzka [a,*], Tamara van Gog [b], Michael Dorr [c], Katharina Scheiter [d], Peter Gerjets [d]

[a] Centre for Learning Sciences and Technologies, Open University of The Netherlands, The Netherlands
[b] Institute of Psychology, Erasmus University Rotterdam, The Netherlands
[c] Schepens Eye Research Institute, Dept. of Ophthalmology, Harvard Medical School, Boston, United States
[d] Knowledge Media Research Center, Tuebingen, Germany

ARTICLE INFO

ABSTRACT

This study investigated how to teach perceptual tasks, that is, classifying fish locomotion, through eye movement modeling examples (EMME). EMME consisted of a replay of eye movements of a didactically behaving domain expert (model), which had been recorded while he executed the task, superimposed onto the video stimulus. Seventy-five students were randomly assigned to one of three conditions: In two experimental conditions (EMME) the model's eye movements were superimposed onto the video either as a dot or as a spotlight, whereas the control group studied only the videos without the model's eye movements. In all conditions, students listened to the expert's verbal explanations. Results showed that both types of EMME guided students' attention during example study. Subsequent to learning, students performed a classification task for novel test stimuli without any support. EMME improved visual search and enhanced interpretation of relevant information for those novel stimuli compared to the control group; these effects were further moderated by the specific display. Thus, EMME during training can foster learning and improve performance on novel perceptual stimuli.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In many domains such as medicine, navigation and control of transport, or biology, people examine complex, dynamic, and information-dense visual material such as CT scans, cockpit displays, or film recordings of animals. Tasks within these domains rely heavily on visually searching and interpreting perceptual information; hence, we refer to them as *perceptual tasks* (cf. Chi, 2006). An important question for educators in these domains is how to effectively teach accomplishment of perceptual tasks. A promising avenue is suggested by studies that show that guiding people's attention via the eye movements of a successful model enhances their performance on the perceptual task for which guidance was provided (Grant & Spivey, 2003; Litchfield, Ball, Donovan, Manning, & Crawford, 2010). However, it is crucial to take this research one step further and investigate whether such attention guidance cannot just enhance performance on the task at hand, but can also foster *learning*, that is, enhance performance on

future, novel tasks when attention guidance is no longer present. The present study addresses this question against the backdrop of instructional design theories that analyze learning from a cognitive information processing perspective.

1.1. Learning as cognitive processing of perceptual input

Instructional design theories, such as Mayer's (2005) cognitive theory of multimedia learning (CTML) and Sweller's cognitive load theory (CLT; Sweller, Van Merriënboer, & Paas, 1998) describe how the human mind handles information from a perceptual input (e.g., visual or auditory information) at different stages of processing in order to construct a long-lasting mental representation of this input in long-term memory, that is, in order to learn. Both theories rely on models with a long-standing tradition in cognitive psychology for describing human information processing, such as the multi-store model of memory by Atkinson and Shiffrin (1968) or the working memory (WM) model by Baddeley (e.g., Baddeley, 1992; for a more recent version cf. Baddeley, 2012). According to the CTML and CLT, attention and working memory are fundamental during learning, because all information needs to be attended and processed in working memory before it can be integrated with prior knowledge and stored in long-term memory. Because

* Corresponding author. Centre for Learning Sciences and Technologies, Open University of The Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands. Tel.: +31 45 576 2410; fax: +31 45 576 2800.
E-mail address: Halszka.Jarodzka@ou.nl (H. Jarodzka).

attentional as well as working memory resources are limited with regard to the amount of information that can be processed in parallel (Fougnies & Marois, 2006; Miller, 1956; Peterson & Peterson, 1959), according to CTML and CLT instruction needs to be designed in a way that makes optimal use of these limited processing resources. Accordingly, various instructional design principles have been postulated that prescribe how instructional materials should be designed so that learners can select relevant information from the perceptual input, organize it in working memory and integrate it with prior knowledge. Only if such an active processing of the instructional materials can be achieved, (deeper) learning will occur. We will refer to two of these principles, namely, (1) the worked examples principle, that advocates the use of examples for novices' skill acquisition (e.g., Atkinson, Renkl, Derry, & Wortham, 2000; Van Gog & Rummel, 2010) and (2) the cueing or signaling principle, that is, the use of cueing techniques to highlight relevant information (e.g., De Koning, Tabbers, Rikers, & Paas, 2009), in a later stage of this paper.

The assumptions made by the CTML and CLT are assumed to hold true for any kind of task. In the following, we will address the specifics of learning to accomplish perceptual tasks, which are at the focus of the present paper, by first describing the processes necessary for task accomplishment and the way they develop with experience and then highlighting the characteristics of highly perceptual tasks that make them particularly difficult to accomplish.

1.1.1. Active processing in perceptual tasks

In the case of visuo-perceptual tasks, information is selected by means of *visual* attention (i.e., the conscious devotion of attention to visual information required to absorb it and transfer it to working memory). Visual attention is closely linked to where a person is looking (Deubel & Schneider, 1996; Just & Carpenter, 1980), which in turn can be captured by means of eye tracking (Holmqvist et al., 2011). Eye tracking research has shown that experience with a perceptual task influences visual attention, for instance in static material — -such as artwork — or dynamic material — such as biological motion or traffic (Huestegge, Skottke, Anders, Müsseler, & Debus, 2010; Jarodzka, Scheiter, Gerjets, & Van Gog, 2010; Vogt & Magnussen, 2007). In particular, less experienced individuals often attend to salient information (i.e., bottom-up or stimulus-driven attention allocation) that may not necessarily be relevant for task performance (Jarodzka et al., 2010; Lowe, 1999). Individuals with higher expertise, on the other hand, know which information is important for task performance (i.e., top-down or observer-driven attention allocation), which enables them to efficiently attend to it. It has been repeatedly shown that individuals with higher expertise focus faster and/or proportionally longer on relevant information than individuals with lower expertise, while ignoring potentially salient, but irrelevant information (e.g., Haider & Frensch, 1999; Huestegge et al., 2010; Jarodzka et al., 2010; Van Gog, Paas, & Van Merriënboer, 2005). Hence, individuals with little or no experience in a perceptual task may have difficulties to *select* task-relevant information by adequately controlling their visual attention. This may not be overly problematic with static stimuli, where it will only take more time to identify the relevant information; however, it might pose more severe problems with dynamic stimuli, which often require that information is attended to at a certain time, since, otherwise, this information will be missed. Consequently, novices need support in selecting information from a complex perceptual input.

Attending to relevant information is necessary, but not sufficient for successfully performing or learning to perform a perceptual task, however. This information also needs to be organized and interpreted, which requires it to be *integrated* with other information from the environment and with prior knowledge. Research has shown that even when less experienced individuals attend to thematically relevant information, they do not necessarily know how to interpret it. For instance, Cook, Wiebe, and Carter (2011) showed that when learning from complex visualizations of DNA strings, students of diverse expertise levels all attended to the thematically relevant features. However, less experienced students were not able to interpret the ongoing processes. Therefore, when teaching perceptual tasks, it is important to keep in mind that less experienced individuals will not only have difficulties with *selecting*, but also with *interpreting* information.

1.1.2. Specific challenges in active processing of realistic and dynamic stimuli

As mentioned above, visually searching for the right information and correctly interpreting that information is challenging, the more so when materials are *realistic* and *dynamic*.

Realistic materials, such as photographs and videos, reproduce the real world and keep many features of real objects intact, such as color, shape, etc. (cf. Rieber, 1994). Such materials typically contain much more information than is relevant to the task at hand. Hence, relevant information has to be searched for among many irrelevant elements, which, however, may be salient and could therefore easily distract visual attention away from the thematically relevant information (Dwyer, 1969). Visual saliency has a large influence on visual attention, in the sense that visually salient areas are looked at faster and longer and are also better remembered; this latter effect, however, can be overridden by the thematic relevance of specific areas — given the observer knows what is thematically relevant (Kaakinen, Hyönä, & Viljanen, 2011). This however, is not the case for novices in a domain; hence, we must assume that they will be mainly driven by visual saliency. Often the relation between thematic relevance and visual salience is not straightforward, and as a consequence the most attention attracting elements may be entirely irrelevant to the task, whereas the crucial element may not be very visually salient at all (Schnotz & Lowe, 2008). Research has shown that students have more difficulties to learn from realistic than from schematic material (e.g., Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009). Hence, for realistic material the *visual search* for relevant information is particularly challenging. In an educational context this means that when studying realistic material it may be beneficial to support the learner's visual search of the relevant information.

Dynamic material represents changes over time. The challenging aspect is that information is transient, that is, it appears only in specific moments in time (Hegarty, 1992). Thus, this information not only has to be attended to at the right moment in time (i.e., selected), but it also has to be kept active in working memory, while new information enters the perceptual system, which imposes high working memory load (e.g., Ayres & Paas, 2007; Spanjers, Van Gog, & Van Merriënboer, 2010). Moreover, several important elements may be present at the same time, which makes it difficult to attend to all of them (Lowe, 2003). That is, the attention of the observer would need to be divided among multiple elements, thereby causing split attention, which has been shown to hamper information processing (cf. split-attention effect: Chandler & Sweller, 1991). Hence, it is not only challenging to select dynamic information but also to keep it active in working memory so that it can be *interpreted*. For educational purposes this means that when studying dynamic material it may be beneficial to support the learner's visual search *and* interpretation of the relevant information.

In the following section, we will introduce an instructional format for perceptual tasks that fulfills the aforementioned premises and that relies heavily on the use of examples.

1.2. Designing example-based Instruction for perceptual tasks

According to CTML (Mayer, 2005) and CLT (Sweller et al., 1998), instructional materials should be designed to optimally make use of limited attentional and cognitive resources by using these resources for effective rather than ineffective processes. Example-based instruction is such an instructional format that has shown to be effective for novice learners (for reviews see Atkinson, Derry, Renkl, & Wortham, 2000; Sweller et al., 1998). Examples show students how they should perform a task, either by providing them with a written, worked-out problem solution to study (i.e., worked examples) or by allowing them to observe an experienced model demonstrating how to perform the task live or on video (i.e., modeling examples; for a review see Van Gog & Rummel, 2010). By doing so, learners do not 'waste' working memory resources on searching for the correct solution by means of trial and error, but can focus on learning the correct solution procedure instead (cf. worked example effect, Atkinson et al., 2000; Sweller et al., 1998).

For motor tasks, the central task steps are directly observable, but for cognitive tasks this is not the case. Instead, the model has to be asked to unravel these covert processes by verbalizing them while performing the task (Collins, Brown, & Newman, 1989). To be able to understand the model's verbal explanations in a perceptual task, the students should also attend to the same information as the model, since the verbal explanations may not be self-explanatory without attending to the elements that they reference. The eye tracking research described above suggests that is unlikely that a student without experience in a perceptual task and an experienced model will spontaneously attend to the same information simultaneously, since novices' visual attention is likely to be guided more strongly by visual salience than thematic relevance. In line with this reasoning, Boucheix and Lowe (2010) state that students' attention should be guided in dynamic complex material in the appropriate order and speed for them to learn, for instance, by means of cueing, that is, by making relevant information more visually salient (for a review: De Koning et al., 2009).

Research often finds ambiguous effects of cueing on learning (e.g., De Koning, Tabbers, Rikers, & Paas, 2010; Moreno, 2007). One reason might be that designers place cues where they think students should look, instead of where a successful task performer would actually look. Accordingly, using eye movements of experienced and successful task performers might improve instruction by cueing. In line with this reasoning, Grant and Spivey (2003) developed a cue based on a comparison of eye movements of individuals who either were or were not successful in solving an insight problem (i.e., Duncker's radiation problem). In a follow-up experiment, individuals received the same problem, either without any cue, with the information cued that successful problem solvers had attended to more in the previous study (i.e., relevant), or with other (i.e., irrelevant) information cued. Participants who saw the relevant information cued solved the problem significantly more often than participants in the other conditions. Thus, Grant and Spivey showed that designing a cue based on successful problem solvers' eye movements could enhance performance. However, the material used in this study was not visually complex, as it consisted only of a simplistic, static visualization composed of one solid black dot surrounded by one black circle, so the elements competing for the observers' attention were few and clearly distinguishable.

Perceptual tasks, as addressed in this article, often rely on realistic and dynamic material. In such cases, it is far more difficult to distinguish irrelevant from relevant information; moreover, often several elements need to be inspected and in a specific order. Here, it may be beneficial to cue the entire process of visual search by displaying eye movements of a successful performer directly to

guide other people's attention. This kind of cue has shown to be effective, for instance, Litchfield et al. (2010) showed that a direct use of eye movements as cues enhanced pulmonary node detection in information-rich, but static chest X-rays. Moreover, research by Velichkovsky (1995) has shown that directly displaying eye movements on dynamic material facilitates communicative situations (i.e., collaboratively completing a puzzle). More importantly, these studies showed that cues based on or direct displays of eye movements improved immediate performance on the task for which guidance was provided. The question remains, however, whether such an eye movement based instruction could foster learning. Therefore, the present study investigated whether this kind of attention guidance would also enhance learning, that is, performance on future, novel perceptual tasks when the guidance is no longer present.

1.2.1. Eye movement modeling examples (EMME)

To sum up, studying examples of experienced models solving a task is an effective teaching method. For perceptual and cognitive tasks the model has to externalize covert thought processes by verbalizing them; additionally, the otherwise inaccessible process of attention allocation can be made visible by displaying the model's eye movements, which might synchronize students' attention with that of the model. This can be expected to aid students' selection of relevant information during example study, but it might also foster their own future search behavior on novel tasks. In addition, synchronized attention might aid students' interpretation of what the model is explaining during example study, which should enhance their learning from the example, which would be evident in their ability to interpret novel tasks by themselves. Based on these considerations we developed eye movement modeling examples (EMME; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). EMME consist of a display of a model's eye movements superimposed onto the video stimulus that the model is interpreting (i.e., a gaze replay) along with the model's verbal explanations.[1]

In an earlier study, we compared the effects of EMME and traditional modeling examples (i.e., without gaze replay) on learning a procedural problem-solving task (Van Gog et al., 2009). The challenge in this task was to come up with one specific order of moving several objects according to certain rules (cf. Tower of Hanoi). In hindsight, however, the verbal explanation and the eye movements of the model were probably redundant. This was the case because the students could easily infer from the verbalizations alone where to look. That is, students always had to choose between one of two moves, which could be easily be verbally referred to by the appearance of the objects and their position on the screen. As a consequence, EMME with verbal explanations were not effective for learning. When verbal explanations are sufficient to guide visual attention, displaying the eye movements would be redundant, and presenting redundant information is known to hamper learning from examples (for a review, see Sweller, 2005).

The present study focuses on a visually complex perceptual task (i.e., realistic and dynamic) in which the guidance provided by the eye movements and the verbal explanations is unlikely to be redundant: distinguishing fish locomotion patterns based on realistic and dynamic videos (for more detailed information on the relevance of the verbalizations and the eye movements see

---

[1] These videos are not plain recordings of someone's performance. Instead, they are carefully didactically prepared in that (1) the model is chosen not only for domain, but also for teaching experience, (2) the model is carefully instructed to behave in a didactic manner, and (3) the model evaluates each recording for its usefulness in an educational setting. For details on this procedure, see Section 2.2.2.

description of EMME used in this study in Section 2.2.2). Previous research has shown that students have difficulties not just with interpreting, but already with *visually searching* for the relevant information in this task (Jarodzka et al., 2010). Thus, EMME are likely to be beneficial in this task.

1.2.2. Displaying eye movements in EMME

One question that arises regarding the design of EMME is how to best depict the model's eye movements. Standard eye tracking software displays often show fixations as solid dots, which is comparable to visual cues that increase the contrast of the to-be-cued element by highlighting it with a salient color. However, using the manufacturers' display options would add another element to the realistic stimuli that already contain a lot of perceptual information. Accordingly, the display itself might either obscure relevant information or attract attention to it, thereby drawing attention temporarily away from the information the model looked at. Another option might be to reduce the visual saliency of the other elements on the screen, only keeping the fixated elements clearly visible, as in a spotlight (so-called anti-cueing; Lowe & Boucheix, 2011). To manipulate visual saliency, realistic and dynamic videos can be manipulated in their contrast in space and over time (Dorr, Jarodzka, & Barth, 2010). Research in Computer Science has shown that this manipulation of videos guides visual attention (Vig, Dorr, & Barth, 2009); it attracts attention to the cued areas by *reducing* visual information elsewhere (for an example with gaze-contingent manipulation of naturalistic videos, see Barth, Dorr, Böhme, Gegenfurtner, & Martinetz, 2006). The two ways of representing the model's eye movements, namely via dots and via spotlights, were implemented in the present study.

1.3. Hypotheses

We hypothesized that attention guidance via the model's eye movements during example study would be successful, that is, participants who studied EMME would follow the model's eye movements with their eyes instead of looking at other areas of the video. This would result in more similar scanpaths across students in the EMME conditions, whereas students in the control group might look anywhere so that their scanpaths were expected to be more diverse (Hypothesis 1).

Moreover, it was expected that due to this guidance, the EMME groups would learn which information to attend to and how to interpret it, which would be evident in more efficient search behavior on novel test videos (Hypothesis 2) and better interpretation performance of the fish locomotion patterns on novel test videos (Hypothesis 3) than the control group. Students' ability to *visually search for relevant information* on novel test videos was assessed by means of eye tracking. To test the ability to *integrate relevant information with prior knowledge* students were asked to actively interpret information from novel test videos.

Based on the current state of knowledge in the field, it is unclear whether we may expect any differences between the two EMME conditions on learning outcomes in terms of either visual search or interpretation performance. One may argue that representing eye movements in a way that the contrast of unattended information is reduced (spotlight display) rather than in a way that adds information (dot display) would lower the amount of information that has to be processed and thus reduce working memory load, which might have a positive effect on learning. On the other hand, the test videos no longer contain such guidance, and have to be processed in their full complexity, which the dot display rather than the spotlight group might already be used to.

2. Method

2.1. Participants and design

Seventy-five students at a German university were recruited by on-campus flyers. Consequently, these students were from different majors (mostly Psychology). Biology students were excluded from participation and participating students had no prior knowledge on the domain of fish locomotion as determined by a prior knowledge questionnaire. They had normal or corrected-to-normal vision and were randomly assigned to one of three conditions ($n = 25$ each): (1) control condition without attention guidance, (2) attention guidance with the original manufacturer-provided display of the model's eye movements (dot display), (3) attention guidance by blurring out non-attended areas and leaving fixated areas displayed at a high resolution (spotlight display). Three participants from the spotlight display condition had to be excluded due to poor eye tracking data quality, resulting in 72 participants in the final sample (age: $M = 22.83$ years, $SD = 4.04$; 50 female).

2.2. Apparatus and materials

2.2.1. Eye tracking equipment

Participants' eye movements while studying the examples and while completing the visual search test tasks were recorded with a Tobii 1750 remote eye tracking system (50 Hz) using ClearView software. This equipment was also used to create the examples.

2.2.2. Examples

The examples for the control group consisted of four digital videos, sized 720 * 576 pixel, depicting single fish, each swimming according to a different locomotion pattern. The videos were looped to last the length of a spoken description of the locomotion pattern by the model. Hence, their duration varied (mean duration = 81.00 s, $SD = 27.24$). The model was a professor of marine zoology, with extensive experience in teaching the topic of fish locomotion patterns. Hence, he was not only an expert in this domain, but also in teaching it. The expert was instructed to behave didactically, that is, to explain to novice students what the relevant aspects of the locomotion pattern shown in each video were. Each recording was replayed to the expert so that he could self-evaluate the replay data whether they would be suitable to be used for instructional purposes based on a number of statements (e.g., for a novice student, the locomotion pattern is explained in enough detail, in comprehensible terms, etc.; cf. Jucks, Schulte-Löbbert, & Bromme, 2007), and if necessary, he could re-record it. To ensure a close relation between the verbal explanations and the eye movements, the model could pre-inspect each video carefully to prepare what he wanted to say during the actual recording. Such a procedure has been shown to result in a tight gaze-voice coupling (Richardson & Dale, 2005). Table 1 shows the transcription and coding of one of the modeling videos. From this table it becomes clear that the model looks at the areas that he is talking about. For instance, in timeslot 00:00 to 00:06 he refers to the dorsal and anal fin and at the same time his eye movements are on these fins. Moreover, in this same timeslot we clearly see the additional benefit of the eye movement display: a novice would not know which the 'dorsal' or the 'anal' fin is and would lose time looking for it, while the video continues. When seeing the eye movement display, however, the novice immediately sees which fin the model refers to. Furthermore, we can see the additional value of the verbalizations from this table. For example, in timeslot 00:12 to 00:18 only the verbal utterance provides the information on how to interpret the motion (i.e., in a wavelike manner).

Table 1
Coding of one eye movement modeling video (Balistiform) across time for verbalizations of the model in relation to his gaze location.

Time (in s)	Verbalizations[a]	Location of eye movements	Task step
00:00	–	• Center of fish body	–
00:00–00:06	*"This fish swims using its dorsal and its anal fin."* [Clarification that these are the fins used for propulsion and not for other purposes, such as navigation]	• Dorsal fin • Anal fin [Indication of the relevant areas on the video, i.e., disambiguation of technical terms for fins]	Detection of relevant areas
00:06–00:12	*"Those are the two fins that are in opposition to each other, quite at the end of the fish's body."*	• Dorsal fin & posterior part of body	Detection of relevant areas
00:12–00:18	*"And it is striking that a wave movement is running across both fins."* [Interpretation of motion]	• Dorsal fin [Disambiguation, which body part's motion is interpreted]	Interpretation of their motion
00:18–00:24	*"This is not only a simple flapping, but this is simply a wiggling on these two fins."* [Interpretation of motion]	• Dorsal fin • Brief look at pectoral fin • Posterior part of body - [Disambiguation, which body part's motion is interpreted] - [Preview to which area the following statement will relate to]	Interpretation of their motion
00:24–00:34	*"And in addition to these two fins the fish only moves, and this is clearly less striking, its pectoral fins."* [Explanation of irrelevant moving body part]	• Pectoral fin [Disambiguation, which body part is referred to]	
00:34–00:42	*"All other fins, the caudal fin as well, are not moving."*	• Looking quickly twice between pectoral and caudal fin • Caudal fin • Briefly pectoral fin - [Disambiguation of technical term] - [Location of other potentially relevant fins]	

In square brackets the additional value of either the verbalizations or the eye movement display is provided.

[a] Translated from German.

In the control condition, learners received only the videos of the fish that were accompanied by the model's verbal explanations. The verbal explanations were present in all three experimental conditions and were identical across conditions.

In the dot display condition, participants received the same examples as the control group, but those additionally included a dot display of the model's eye movements. These were created using the manufacturer rendered eye movement display with an I-DT filter (i.e., Dispersion Threshold Identification, cf. Salvucci & Goldberg, 2000), where fixations are defined as lasting a minimum of 100 ms with a 30 pixel dispersion. The eye movements were displayed as a solid yellow dot (no gaze trail) that changed size dynamically (smaller, larger) according to fixation duration (shorter, longer).

In the spotlight display condition, potentially distracting features in the unattended areas of the video were filtered out. That is, the area in the focus of the model's attention (with a radius of 32 pixels, i.e., approx. .80° visual angle) was visible in an unaltered way, whereas the areas surrounding it were 'blurred' by reducing local spatio-temporal spectral energy (i.e., contrast in space and time) and removing color saturation from non-attended areas in a similar fashion (for a detailed description of the underlying technical procedure see Dorr, Jarodzka, et al., 2010; Dorr, Martinetz, Gegenfurtner, & Barth, 2010). Fig. 1 shows a screen shot from each of the three conditions.

2.2.3. Dependent measures

Dependent measures were obtained during the learning and during the testing phase and will be discussed separately.

2.2.3.1. Learning measures. In the learning phase, two measures were obtained. First, we determined whether students had followed the model's gaze (if available) during studying the examples. In the dot display condition the exact spot of the model's gaze was

Fig. 1. Screenshots from the video-based modeling examples of all three conditions used in the study (from left to right): Video example of a swimming fish with only verbal explanations of the model on how to classify its locomotion pattern (control condition); video example with a visualization of the model's eye movements as a solid dot (dot display); video example with blurred out non-attended areas and fixated areas displayed at a high resolution (spotlight display).

occluded by the dot itself, thus students were guided to look just next to the dot. In contrast, in the spotlight condition students were guided to look at the exact spot of the model's gaze. Hence, calculating the distance between the model's and the student's gaze would have been an unfair comparison of both conditions. To avoid this issue, the *coherence of participants' eye movements on the instructional videos* was computed following Dorr, Martinetz, et al.'s (2010) scanpath similarity measure for videos. Gaze distribution maps were generated for each frame across the four videos. For each participant her or his similarity to the maps of the remaining participants from the same condition on that example was determined, and then the mean of all four examples was calculated per participant. If participants closely follow the model's gaze, we will expect high correspondence between scanpaths within conditions (as indicated by low values).[2]

Second, we obtained a subjective rating of invested mental effort (an indicator of experienced cognitive load), using the rating scale developed by Paas (1992). Participants were asked to rate mental effort after studying each video example ("How much effort did you invest to complete this task?") on a 9-point rating scale ranging from "very, very, very low effort" to "very, very, very high effort".

2.2.3.2. Testing measures.
In the testing phase two different aspects of learning were captured: the efficiency of visual search during inspection of novel videos of fish swimming according to one of the learned locomotion patterns, and the ability to interpret this fish locomotion. For both aspects, three measures were obtained as described in the following.

Visual search was assessed by recording eye movements while participants were shown 4 novel videos (one for each locomotion pattern) of 4 s duration featuring different fish than in the examples. While this may seem short, it should be kept in mind that in real-life diving scenarios fish can be inspected only very briefly, as they tend to swim away from divers. Because fish move very fast, these videos displayed three to six full movement cycles. All in all, the relatively short duration of the videos is ecologically valid for this task. Moreover, in an earlier study, we could show that this duration is enough to detect differences in search behavior between individuals at different expertise levels (Jarodzka et al., 2010). In particular, three eye movement measures were derived on the test videos to determine visual search: (1) the coherence between students' scanpaths within conditions was calculated according to the same method as described above, (2) the time elapsing until participants first looked at areas of the videos that were relevant to the identification of the fish's locomotion pattern for at least 100 ms (time until first fixation in ms), and (3) the total time spent on these relevant areas (total dwell time in ms). These areas were determined with the help of a domain expert and, dependent on the particular locomotion pattern deployed by a fish, covered a fish's fins as well as parts of its body. Since the fish shown in the videos moved around, the position of these relevant areas changed over time. Accordingly, dynamic areas of interest (AOI) were created for the four videos. Each video was divided into 100 ms segments. For each segment AOIs were defined manually. The length of the segments was determined based on the

maximum amount of time for which AOIs did not change within each segment. The data for each AOI were aggregated per video (i.e., across all 100 ms segments) and across all four videos to determine overall dwell time. Since multiple AOIs could be relevant to identifying the fish's locomotion pattern, the time that elapsed until one of these AOIs was fixated was used to determine the time until first fixation of a relevant area.

For capturing students' *interpretation ability* participants were shown another 22 videos (their eye movements were not recorded during this task), and after each video participants had to answer multiple-choice questions on the interpretation of that fish's locomotion pattern, by checking the correct answers on two questions relevant for describing a locomotion pattern (Lindsey, 1978): (1) *which* part of the body is used to produce propulsion (pectoral, pelvic, dorsal, anal, caudal fin, or the entire body), and (2) *how* does this part move (undulating/wavelike or oscillating/paddlelike). Students received 1 point for each question correctly answered (max. of 22 points for each question); however, the point for the second question (how does this part move?) was only assigned if the first (which part produces propulsion?) had been answered correctly. Afterwards, both scores were transformed into percentage correct for easier interpretation. Finally, (3) the same perceived mental effort scoring as during learning was obtained (Paas, 1992).

2.3. Procedure

The experiment was run in individual sessions of approximately 45 min. The eye tracking system was adjusted to the individual features of the participant based on a nine-point calibration. Participants were told that they would subsequently receive four examples in which a biology expert explains each to-be-learned locomotion pattern. Participants in the dot display and spotlight display conditions were additionally informed that they would see what the expert was attending to, and an example of that was shown with an off-topic video. While studying the examples, participants' eye movements were recorded. Then, participants completed the two tests. Before the visual search test, during which eye movements were also recorded, the system was recalibrated. After each learning example and each interpretation ability test, participants rated their perceived mental effort on a 9-point rating scale (cf. Paas, 1992).

3. Results

For all analyses reported here, a significance level of .05 is used.[3] Table 2 presents descriptive statistics of all variables.

3.1. Measures during example study

3.1.1. Attention guidance during example study
The comparison of the coherence of participants' scanpaths during example study by means of an ANOVA showed a main effect of condition, $F(2,69) = 22.16$, $p < .01$, $\eta_p^2 = .39$. Subsequent planned contrasts were conducted to test first whether the two EMME conditions differed from the control condition and second whether the two EMME conditions differed from each other. The first contrast revealed that EMME guided the students' attention more

---

[2] Calculating the distance between each students' and the model's gazes (Euclidean distance over time) led in practice to the exact same result pattern as the scanpath coherence measure (control condition: 136.4 pixel, spotlight condition: 88.6 pixel, and dot condition: 89.1 pixel distance to the model's gaze). The standard deviations in the dot condition are larger than for the spotlight condition, indicating that students indeed were guided somewhere around the dot making a direct comparison of the distances between the model's gazes and the students' gazes confounded with the different visualization types.

[3] Regarding assumptions for calculating ANOVAs it should be noted that "time to first fixation on relevant areas", "interpretation ability of which body parts move", and "interpretation ability of which body parts move" were not normally distributed. Analyses with logarithmized values revealed exactly the same patterns of results.

Table 2
Mean learning and testing outcomes in the control group and both experimental groups (Standard Deviations given in Brackets).

Measures	Condition			Total
	Control	Dot	Spotlight	
Learning				
Coherence of scanpaths	12.07 (1.69)	15.11 (2.08)	15.10 (1.71)	14.05 (2.33)
Perceived mental effort	2.52 (1.32)	3.26 (1.59)	2.99 (1.46)	2.92 (1.47)
Testing				
Visual search				
Time to first fixation on relevant AOIs	1632.45 (681.52)	1530.36 (509.55)	1236.53 (390.88)	1473.84 (564.89)
Total dwell time on relevant AOIs	467.24 (373.14)	701.00 (332.07)	709.80 (313.32)	617.97 (357.04)
Coherence of scanpaths	14.48 (3.29)	14.55 (3.69)	14.84 (2.60)	14.61 (3.21)
Interpretation				
Which body parts are moving?	79.86 (12.46)	88.64 (10.58)	81.64 (11.64)	83.20 (12.09)
How are these body parts moving?	53.79 (16.53)	58.55 (19.65)	56.72 (12.46)	56.21 (16.39)
Perceived mental effort	4.01 (1.35)	3.76 (1.50)	4.70 (1.48)	4.13 (1.48)

Note. Scanpath coherence is reported as a *z*-score. Both visual search variables are given in milliseconds. Both interpretation variables are given in percentage correct.

strongly in comparison to the control group, in that the coherence (or similarity) of scanpaths within the two EMME conditions was larger than in the control condition, $t(69) = 6.65$, $p < .01$, $r = .62$. The second contrast revealed that the concrete design of EMME as a spotlight or as a dot had no effect on students' scanpath coherence, $t(69) = .01$, $p = .99$, $r = .001$.

3.1.2. Mental effort during example study

An ANOVA on invested mental effort during learning showed no significant differences across conditions, $F(2,69) = 1.64$, $p = .20$, $\eta_p^2 = .05$.

3.2. Performance for novel stimuli during testing

3.2.1. Visual search

Two ANOVAs were conducted to analyze students' visual search for novel test stimuli. They revealed that the conditions significantly differed in the time participants took before looking at a relevant area for the first time, $F(2,69) = 4.19$, $p = .02$, $\eta_p^2 = .11$ as well as in the dwell time on the relevant areas, $F(2,69) = 3.74$, $p = .03$, $\eta_p^2 = .10$.

Planned contrasts revealed that in general studying EMME made students attend faster to relevant areas on novel videos, $t(69) = -2.15$, $p = .03$, $r = .25$, and for a longer time, $t(69) = 2.70$, $p = .01$, $r = .31$, in comparison to the control group. The planned contrasts for comparing the two EMME conditions to each other showed that studying spotlight EMME resulted in relevant areas being attended more quickly than studying dot EMME, $t(69) = 2.01$, $p = .048$, $r = .24$, whereas there were no differences between the two conditions with respect to the total time spent looking at relevant areas on novel videos, $t(69) = -.53$, $p = .60$, $r = .06$.

ANOVA on the coherence of participants' scanpaths during testing showed no significant main effect of condition, $F < 1$.

3.2.2. Interpretation ability and mental effort during testing

Two ANOVAs were calculated to investigate the effect of EMME on students' interpretation ability as indicated by their ability to identify relevant body parts and how they move. The first ANOVA showed a significant difference between conditions for indicating *which body part* produces propulsion, $F(2,69) = 4.57$, $p = .01$, $\eta_p^2 = .12$. A first planned contrast revealed that studying EMME led to a marginally better interpretation regarding which body parts produce propulsion in comparison to the control group, $t(69) = 1.81$, $p = .07$, $r = .21$, whereas the second contrast showed that studying dot EMME yielded better performance than studying spotlight EMME, $t(69) = 2.35$, $p = .02$, $r = .27$. The second ANOVA

showed no significant differences for interpreting *how* the propulsion producing body part moves, $F < 1$.

An ANOVA on perceived mental effort during testing showed marginal differences between conditions, $F(2,69) = 2.66$, $p = .08$, $\eta_p^2 = .07$. A first planned contrast revealed that studying EMME had no significant effect on perceived mental effort when compared to the control group, $t(69) = .64$, $p = .53$, $r = .08$. However, the second contrast indicated that when studying spotlight EMME students had to invest significantly more mental effort in comparison to the dot EMME, $t(69) = -2.24$, $p = .03$, $r = .26$.

4. Discussion

This study investigated whether eye movement modeling examples (EMME) improve learning to perform a perceptual task — in this case, classifying fish locomotion patterns. As this perceptual task requires the examination of realistic and dynamic material, it may be difficult for students to follow what the model is explaining when their attention is not synchronized with the model's attention. EMME provide a solution to this problem by enriching modeling examples with the eye movements of the model. We hypothesized that EMME would guide students' attention during example study, which would in turn lower their working memory load and enhance their learning. The eye movements were displayed both in a dot display (adding information) and in a spotlight display (reducing other information) format (or cue vs. anti-cue format in terms of Lowe & Boucheix, 2011).

Results showed that in line with our first hypothesis, visual attention can be successfully guided by cueing a successful person's eye movements in video-based modeling examples. Analyses of students' eye movements during example study showed that students' scanpaths in both EMME conditions were more coherent, that is, they looked at similar areas of the videos, whereas in the control condition students' scanpaths were diverse, that is, they looked at very different areas of the videos. This finding suggests that both display types of the model's eye movements were likely to be successful in guiding students' attention. Moreover, in line with Hypotheses 2 and 3 the results showed that these EMME fostered learning by improving students' visual search and their ability to identify relevant information in novel test stimuli. A further interpretation of the motion of these relevant elements could not be significantly improved by EMME and leaves room for fine-tuning of this instructional method. Nevertheless, given the rather short intervention in this study, the results are already promising: In terms of Mayer's (2005) CTML and the CLT (Sweller et al., 1998) we could show that EMME positively influence two central aspects of information processing, namely *visual selection* of

relevant information from a complex perceptual stimulus and its *organization and integration* of that information with prior knowledge and the verbal explanation provided by the model.

Other researchers have claimed that the specific manner of cueing – as color highlight or as an anti-cue spotlight – should not make much of a difference on learning, as long as the visuo-spatial contrast remains the same (Lowe & Boucheix, 2011). The results of this study indeed seem to suggest, that the beneficial effect of EMME is independent from the design (i.e., from the eye movement display), as both EMME conditions guided students' visual attention during learning, affected visual search on new test videos, and improved interpretation compared to the control group. However, we did also find differences between the two EMME conditions that may result from the design of the eye movement display: Whereas students within each condition acted very similar with respect to their visual exploration of the videos, only the spotlight condition increased visual search efficiency on novel stimuli in terms of looking faster at relevant areas. On the other hand, the dot condition compared to the spotlight condition led to better interpretation performance in terms of indicating which body parts were involved in the locomotion. That is, the two display types fostered different steps of the learning process; whereas learning from spotlight EMME fostered the *visual selection* of information from the environment, learning from dot EMME fostered the *organization and integration* of this information with prior knowledge (cf. Mayer, 2005; Sweller et al., 1998). Hence, none of these displays can be declared to be better suited for teaching this perceptual task overall; instead each seems to support a different learning goal. As a consequence it is unclear based on the data of this study whether there is indeed a deterministic relation between 'visual search' and 'interpretation ability'. For instance, some research on cueing in instructional animations has also shown that sometimes cues do guide visual attention to the relevant elements, but do not result in increased learning outcomes (e.g., Kriz & Hegarty, 2007).

A possible explanation for this differential effect is that the different displays might stimulate different processes. The spotlight display guides the students' attention directly on the model's eye movements and leaves no other information to look at (as it is blurred). This might have fostered a mainly perceptual aspect of learning that allows to re-enact perceptual processes (cf. Goldstone, 1998). The dot display, on the other hand, occluded the exact spot the model looked at, but at the same time did not 'force' students to follow the model's gaze and allowed for attending to other information. Thus, students learning with the dot display were guided "around" the relevant areas during learning instead of being guided directly to the spot the model looked at. Hence, they might have experienced problems to locate those areas as quickly during testing as the spotlight condition. However, this may also explain why the dot display had a positive effect on the ability to interpret the fish locomotion patterns, as these students could get a holistic view of the entire fish and its locomotion, which students in the spotlight condition could not. This holistic view may have allowed students in the dot display condition to perceive some other aspects than only those that the model looked at by peripheral vision, which might have given them a benefit. Research on visual short term memory has shown that when a global configuration of elements remains unaltered, it is *easier* for participants to remember specific features of one relevant element (e.g., color, position) than when some or all other elements are removed (Jiang, Olson, & Chun, 2000). As such, the dot display condition might have had an advantage with regard to building a mental model of the overall configuration of the locomotion pattern. Our findings on perceived mental effort lend tentative support to this assumption: whereas the groups did not differ in the amount of overall perceived mental effort, the spotlight group invested significantly more mental effort during testing than the dot group. That is, considering both the higher amount of invested mental effort and the lower interpretation performance, the conclusion seems warranted that it was indeed more *difficult* for the spotlight display group to remember specific aspects of the relevant elements and thus, to *integrate* them with what they just saw.

In our previous research, EMME in contrast to the present findings proved only somewhat helpful for learning when no verbal explanation was present, but hindered learning when combined with a verbal explanation (Van Gog et al., 2009). This difference with the present study can be explained both by the nature of the task as well as by the high redundancy between the verbal and the visual input in the previous study. In the present study, verbal and visual inputs were highly complementary. Most importantly, the verbal explanations were necessary to provide information on why the information attended by the expert was relevant at a given moment. For instance, the expert might have attended to a particular fin while saying that this fin was relevant either to propulsion or navigation – which is information that a novice could have not extracted from the visual input alone. Similarly, a verbal description of a fish swimming forward while moving multiple fins in parallel to do so, would be difficult if not impossible to comprehend in isolation, given the complex visuo-spatial nature of these movements.

Further research is required to investigate the explanations for the differences between the two modes of eye movement display as cues in EMME and investigate the practical significance of EMME beyond a laboratory context. Nevertheless, these findings are interesting for educators and trainers, because they show that guiding an individual's attention based on a more competent person's eye movements can not only guide that individual's thought, thereby affecting direct performance, but can go well beyond that to guiding learning. Therefore, the question arises now how to transfer the findings from this study into practice. Many real-world tasks require perceptual skills. When instructors are teaching novices to execute these kinds of tasks, they should keep in mind to explicitly teach perceptual components as well. EMME might be an appropriate way to do this for many tasks.

Acknowledgements

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214. http://dx.doi.org/10.2307/1170661.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

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

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559. http://dx.doi.org/10.1126/science.1736359.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29. http://dx.doi.org/10.1146/annurev-psych-120710-100422.

Barth, E., Dorr, M., Böhme, M., Gegenfurtner, K. R., & Martinetz, T. (2006). Guiding the mind's eye: improving communication and vision by external control of the

scanpath. In B. E. Rogowitz, T. N. Pappas, & S. J. Daly (Eds.), *Human vision and electronic imaging* (pp. 1–8). San Jose: SPIE Press. http://dx.doi.org/10.1117/12.674147.

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

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332. http://dx.doi.org/10.1207/s1532690xci0804_2.

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

Collins, A. F., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Cognition and instruction: Issues and agendas* (pp. 453–494). Mahwah, NJ: Erlbaum.

Cook, M., Wiebe, E., & Carter, G. (2011). Comparing visual representations of DNA in two multimedia presentations. *Journal of Educational Multimedia and Hypermedia, 20*, 21–42.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educational Psychology Review, 21*, 113–140. http://dx.doi.org/10.1007/s10648-009-9098-7.

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

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research, 36*, 1827–1837. http://dx.doi.org/10.1016/0042-6989(95)00294-4.

Dorr, M., Jarodzka, H., & Barth, E. (2010). Space-variant spatio-temporal filtering of video for gaze visualization and perceptual learning. In C. Morimoto, & H. Instance (Eds.), *Proceedings of the 2010 Symposium on eye tracking research & Applications ETRA '10* (pp. 307–314). New York: ACM. http://dx.doi.org/10.1145/1743666.1743737.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*, 1–17. http://dx.doi.org/10.1167/10.10.28.

Dwyer, F. M. (1969). The effect of varying the amount of realistic detail in visual illustrations designed to complement programmed instruction. *Programmed Learning, 6*, 147–153. http://dx.doi.org/10.1080/1355800690060301.

Fougnies, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory. *Psychological Science, 17*, 526–534. http://dx.doi.org/10.1111/j.1467-9280.2006.01739.x.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612. http://dx.doi.org/10.1146/annurev.psych.49.1.585.

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

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: more evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 172–190. http://dx.doi.org/10.1037//0278-7393.25.1.172.

Hegarty, M. (1992). Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084–1102. http://dx.doi.org/10.1037/0278-7393.18.5.1084.

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, UK: Oxford University Press.

Huestegge, L., Skottke, E.-M., Anders, S., Müsseler, J., & Debus, G. (2010). The development of hazard perception: dissociation of visual orientation and hazard processing. *Transportation Research Part F: Traffic Psychology and Behaviour, 13*(1), 1–8. http://dx.doi.org/10.1016/j.trf.2009.09.005.

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

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 683–702. http://dx.doi.org/10.1037/0278-7393.26.3.683.

Jucks, R., Schulte-Löbbert, P., & Bromme, R. (2007). Supporting experts' written knowledge communication through reflective prompts on the use of specialist concepts. *Journal of Psychology, 215*, 237–247. http://dx.doi.org/10.1027/0044-3409.215.4.237.

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

Kaakinen, J. K., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *Quarterly Journal of Experimental Psychology, 64*(7), 1372–1387. http://dx.doi.org/10.1080/17470218.2010.548872.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies, 65*, 911–930. http://dx.doi.org/10.1016/j.ijhcs.2007.06.005.

Lindsey, C. C. (1978). Form, function, and locomotory habits in fish. *Fish Physiology, 7*, 1–88.

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*, 251–262. http://dx.doi.org/10.1037/a0020082.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*, 225–244. http://dx.doi.org/10.1007/BF03172967.

Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*, 157–176. http://dx.doi.org/10.1016/S0959-4752(02)00018-X.

Lowe, R. K., & Boucheix, J.-M. (2011). Cueing complex animations: does direction of attention foster learning processes? *Learning and Instruction, 21*, 650–663. http://dx.doi.org/10.1016/j.learinstruc.2011.02.002.

Mayer, R. E. (2005). A cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 41–61). New York: Cambridge University Press.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81–97. http://dx.doi.org/10.1037/h0043158.

Moreno, R. (2007). Optimizing learning from animations by minimizing cognitive load: cognitive and affective consequences of signaling and segmenting methods. *Applied Cognitive Psychology, 21*, 765–782. http://dx.doi.org/10.1002/acp.1348.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive load approach. *Journal of Educational Psychology, 84*, 429–434. http://dx.doi.org/10.1037/0022-0663.84.4.429.

Peterson, L. R., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198. http://dx.doi.org/10.1037/h0049234.

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

Rieber, L. P. (1994). *Computers, graphics, and learning.* Madison, WI: Brown & Benchmark.

Salvucci, D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye tracking protocols. In A. Duchowski (Ed.), *Proceedings of the 2000 symposium on eye tracking research and applications ETRA '00* (pp. 71–78). New York: ACM. http://dx.doi.org/10.1145/355017.355028.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494. http://dx.doi.org/10.1016/j.learninstruc.2008.08.001.

Schnotz, W., & Lowe, R. K. (2008). A unified view of learning from animated and static graphics. In R. K. Lowe, & W. Schnotz (Eds.), *Learning with animation: Research and design implications* (pp. 304–356). New York: Cambridge University Press.

Spanjers, I. A. E., Van Gog, T., & Van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review, 22*(4), 411–423. http://dx.doi.org/10.1007/s10648-010-9135-6.

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.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–295. http://dx.doi.org/10.1023/A:1022193728205.

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

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

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

Velichkovsky, B. M. (1995). Communicating attention: gaze position transfer in cooperative problem solving. *Pragmatics and Cognition, 3*, 199–224. http://dx.doi.org/10.1075/pc.3.2.02vel.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision, 22*, 397–408. http://dx.doi.org/10.1163/156856809789476065.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: eye movement patterns and visual memory in artists and laymen. *Perception, 36*, 91–100. http://dx.doi.org/10.1068/p5262.