Guest editorial

Eye tracking as a tool to study and enhance multimedia learning

Abstract

This special issue comprises a set of six papers, in which studies are presented that use eye tracking to analyse multimedia learning processes in detail. Most of the papers focus on the effects on visual attention of animations with different design features such as spoken vs. written text, different kinds of cues, or different presentation speeds. Two contributions concern effects of learner characteristics (prior knowledge) on visual attention when learning with video and complex graphics. In addition, in some papers eye tracking is not only used as a process measure in itself, but also as input for verbal reports (i.e., cued retrospective reporting). In the two commentaries, the contributions are discussed from a multimedia learning perspective and an eye tracking perspective, by prominent researchers in those fields. Together, the contributions to this issue give an overview of the various possibilities eye tracking opens up for research on multimedia learning and instruction.

Keywords: Eye tracking; Multimedia learning; Animations; Expertise

1. Introduction

The present special issue comprises a set of six papers, in which studies are presented that use eye tracking to analyse multimedia learning processes in detail. In the field of learning and instruction, eye tracking used to be applied primarily in reading research (Hyönä & Niemi, 1990; Just & Carpenter, 1980; for a review, see Rayner, 1998), with only a few exceptions in other areas such as text and picture comprehension and problem solving (Hannus & Hyönä, 1999; Hegarty & Just, 1993; Verschaffel, De Corte, & Pauwels, 1992). However, this has changed over the last years; eye tracking is starting to be applied more often, especially in studies on multimedia learning. Because eye tracking provides insights in the allocation of visual attention, it is very suited to study differences in attentional processes evoked by different types of multimedia and multi-representational learning materials (usually, but not necessarily computer-based; see, e.g., Holsanova, Holmberg, & Holmqvist, 2009, for a study using multimedia materials printed in a regular newspaper format).

Mayer (2005) defines multimedia learning as building a mental model from materials that involve both verbal and pictorial representations. The verbal representations can consist of either spoken or written text, and pictorial representations can comprise both static and dynamic visualizations such as drawings, photos, graphs, animations, or videos. Many studies on the effectiveness of different multimedia learning materials have been conducted, often inspired by Mayer’s Cognitive Theory of Multimedia Learning (see Mayer, 2005) and/or Sweller’s Cognitive Load Theory (Sweller, van Merriënboer, & Paas, 1998). However, most of these studies have mainly drawn conclusions about the cognitive effects of different types of multimedia learning materials based on (transfer test) performance measures, sometimes combined with measures of cognitive load and/or time-on-task. Direct investigations of the cognitive and perceptual processes underlying these effects are relatively rare. For research on multimedia multi-representational learning materials eye tracking can provide unique information concerning what medium or representations are visually attended to, in what order, and for how long. This information can be used in at least three different ways.

First, such eye-tracking data can provide a more detailed account of how exactly well-known effects (e.g., split-attention effect, modality effect, redundancy effect, goal-specificity effect) come about and hence may help in generating or corroborating explanations for these effects. For example, the split-attention effect refers to the consistent finding that separately presented but mutually referring text and diagrams hamper learning, whereas integrated formats foster learning (Chandler & Sweller, 1991; for a review see Ayres & Sweller, 2005). However, there are many different ways in which learners can process for example text and diagrams, and what exactly causes the negative effects on learning in a split format...
2.1. Effects of written vs. spoken text design features

(i.e., cued retrospective reporting). gaining an additional process measure, namely, verbal reports measure in itself, but use eye movement records as a cue for two contributions do not only use eye tracking as a process when learning with video and complex graphics. In addition, Another topic addressed by two contributions concerns effects different kinds of cues, or different presentation speeds.

with different design features such as spoken vs. written text, or picture presented with spoken rather than written text leads to better learning outcomes, is not so much due to the use of multiple working memory channels as a result of multimodal presentation, as it is due to the prevention of split attention by multimodal presentation (see also Low & Sweller, 2005). That is, when presenting an animation with written text, split attention is likely to occur as one cannot read and attend to the animation at the same time. Not surprisingly, Schmidt-Weigand et al. find that learners in spoken text condition spend more time viewing the animation. What is more interesting is that in written text conditions learners consistently start reading before alternating between text and visualization, and that they spend more time reading the text than inspecting the visualizations. With slower presentation speed (i.e., system-paced), additional time is used to view the animation, but under self-paced conditions, additional time taken is mostly used to read the text. With respect to the modality effect that is typically obtained under (fast) system-control conditions (cf. Ginns, 2005), the data of Schmidt-Weigand et al. suggest that even though learners with written text study the animations, the time available for inspecting them under fast system-control conditions may not be sufficient to build a coherent mental representation and may hence explain the inferior learning outcomes compared to conditions with spoken text.

2. Overview of contributions to the special issue

In the present special issue, most of the contributions focus on visual processes that occur when learning with animations with different design features such as spoken vs. written text, different kinds of cues, or different presentation speeds. Another topic addressed by two contributions concerns effects of learner characteristics (prior knowledge) on visual attention when learning with video and complex graphics. In addition, two contributions do not only use eye tracking as a process measure in itself, but use eye movement records as a cue for gaining an additional process measure, namely, verbal reports (i.e., cued retrospective reporting).

2.1. Learning from animations with different design features

2.1.1. Effects of written vs. spoken text

Schmidt-Weigand, Kohnert, and Glowalla (2010) propose that the modality effect, that is, the finding that an animation or picture presented with spoken rather than written text leads to better learning outcomes, is not so much due to the use of multiple working memory channels as a result of multimodal presentation, as it is due to the prevention of split attention by multimodal presentation (see also Low & Sweller, 2005). That is, when presenting an animation with written text, split attention is likely to occur as one cannot read and attend to the animation at the same time. Not surprisingly, Schmidt-Weigand et al. find that learners in spoken text condition spend more time viewing the animation. What is more interesting is that in written text conditions learners consistently start reading before alternating between text and visualization, and that they spend more time reading the text than inspecting the visualizations. With slower presentation speed (i.e., system-paced), additional time is used to view the animation, but under self-paced conditions, additional time taken is mostly used to read the text. With respect to the modality effect that is typically obtained under (fast) system-control conditions (cf. Ginns, 2005), the data of Schmidt-Weigand et al. suggest that even though learners with written text study the animations, the time available for inspecting them under fast system-control conditions may not be sufficient to build a coherent mental representation and may hence explain the inferior learning outcomes compared to conditions with spoken text.

2.1.2. Cueing

In studying an animation, it may be very hard for learners to determine which parts are relevant and need to be attended to and which are irrelevant and need to be ignored (see also Section 2.2 below). The most perceptually salient parts of an animation tend to attract attention, but these are not always the most relevant ones. Cueing might help guide learners’ attention to relevant parts of the animation. Different cueing options are studied in this special issue. De Koning, Tabbers, Rikers, and Paas (2010) investigate the effects of a spotlight cue on visual attention and learning from a heart animation. In such a spotlight cue, all non-cued elements are shaded, and as a consequence, the cued part becomes more perceptually salient. De Koning et al. (2010) show that when only one part of the heart animation is cued, this part gains more visual attention initially, but the effect disappears over time. When multiple parts are cued one after another, the cue is also successful in guiding attention to the cued parts.

Boucheix and Lowe (2010) study the effects of arrows and spreading colour cues in learning from an animation on the functioning of a piano. Spreading colour cues consist of ‘ribbons’ of colour overlaid on the animation, that spread through the animation, thereby focusing the learners’ attention on causal chains in a relatively non-intrusive way, as they are spatially and temporally aligned with the animation. Their results show that whereas arrow cues do not improve comprehension compared to no cues, spreading-colour cues do, and the eye-tracking data suggest that this effect results from redirecting attention to the more relevant parts of the animation. The eye movement and performance data obtained in a second experiment show that this effect applies only when the spreading-colour cues are appropriately synchronized.
(i.e., spatially and temporally aligned with important events in the animation).

Taken together, the studies by De Koning et al. (2010) and Boucheix and Lowe (2010) provide valuable insights into the attentional effects evoked by different forms of cueing, which can be used to enhance multimedia materials by informing the design of cueing devices.

2.1.3. Presentation speed

Another problem that tends to occur in learning with animations is that some important processes depicted may be difficult to perceive due to the speed of the animation’s presentation, which may either be too slow or too fast for relevant changes to be detected (cf. Fischer, Lowe, & Schwan, 2008). Meyer, Rasch, and Schnitz (2010) propose that changing the speed of replay of an animation can emphasize and guide attention to different parts of the animation (cf. cueing). For an animation depicting the working of a four-stroke engine, they hypothesize that slower presentation speeds emphasize micro-events and higher speeds macro-events (higher up in a hierarchy of events), and that different speeds will therefore draw attention to different parts of the animation and result in different learning outcomes. They also study whether going (system-controlled) from a faster to slower speed (“zooming-in”) or from slower to faster speed (“zooming-out”) affects processing and understanding of the animation. Findings from the first study suggest that the higher the chosen presentation speed by the learner, the better the acquisition of macro-events. However, in the second study, using system-controlled animations, it seems that it is not so much the speed, as the sequence of speeds (“zooming-in” or “zooming-out”) that influences visual attention allocation. Although this second study shows no differences in learning outcomes (which, as the authors indicate, is difficult to interpret as they have collected learning outcome data only after viewing the whole animation, not after the first half after which the speed was changed) it does suggest that using different speeds in succession may be an effective way of directing attention.

2.2. Effects of expertise

Eye-tracking research has shown that attention allocation is also often influenced by expertise: With increasing knowledge of a task, individuals tend to fixate faster and proportionally more on task-relevant information. This difference has been found between experts and novices (Charness, Reingold, Pomplun, & Stampe, 2001), but there are indications that it also occurs between individuals with smaller differences in expertise (Van Gog, Paas, & Van Merriënboer, 2005), and it has been found within individuals over time as a result of practice (Haider & Frensch, 1999).

Jarodzka, Scheiter, Gerjets, and Van Gog (2010) study the visual processes of experts and novices while they observe videos of fish with the instruction to classify the fish’s locomotion pattern. They find that—in line with the findings mentioned above—experts attend more to relevant aspects of the stimulus, use more heterogeneous task approaches, and apply knowledge-based shortcuts. As Jarodzka et al. point out, these results can help establish guidelines for cueing novices’ attention and could be used to make modelling examples in this domain more effective (see also Section 3 below).

Canham and Hegarty (2010) investigate whether the findings by Haider and Frensch (1999) that with increasing expertise, individuals learn to ignore task-redundant information, extends to the comprehension of weather maps, and to situations in which expertise is not gradually developed, but in which explicit instruction is given (in this case, on meteorological principles). Simultaneously, they study how different designs of weather maps influence map processing and performance. Their results show that after instruction, more attention is paid to relevant and less to irrelevant information, thereby again corroborating the findings by Haider and Frensch (1999). In addition, inclusion of irrelevant features in the maps seems to impair performance.

2.3. Cued retrospective reporting based on records of eye movements

Useful as they may be, eye movement data require a substantial degree of inferences about underlying cognitive processes, as they do not explain why a participant was looking at certain representations for a certain amount of time and in a certain order. To reduce the amount of inferences required by the researcher, eye-movement data can also be used to complement concurrent or retrospective verbal protocols to obtain a more comprehensive picture of the learning or task performance processes (cf. Van Gog, Paas, & Van Merriënboer, 2005). Moreover, because eye-tracking equipment nowadays allows not only for recording, but also for replaying the records of eye movements (i.e., gaze replays), new possibilities are opened up. For example, records of eye movements can be used to cue retrospective verbal reports of a learning or task performance process (Van Gog, Paas, Van Merriënboer, & Witte, 2005; see also Hansen, 1991; Russo, Johnson, & Stephens, 1989). It has been suggested that this technique of cued retrospective reporting might provide a valuable alternative to concurrent reporting. Especially for research with novice participants, who often experience a high cognitive load and as a result may stop verbalizing their thoughts in concurrent reporting, or with instructional materials that make concurrent reporting impossible, such as animations or videos that contain spoken text, cued retrospective reporting may provide an alternative. This technique has previously been used with problem-solving or information-search tasks in which mouse and keyboard operations were also recorded. That is, the cue consisted of a screen recording including a visualization of the participants’ eye movements (Brand-Gruwel, Van Meeuwen, & Van Gog, 2008; Schonwone et al., 2009; Van Gog, Paas, Van Merriënboer, & Witte, 2005). So in this case, both types of information (mouse/keyboard actions and eye movements) may serve as cues for the retrospective verbal report. Interestingly, in the two contributions to this special issue that use cued retrospective reporting, the eye movements constitute the sole source...
of information in the recording (De Koning et al., 2010; Jarodzka et al., 2010). Moreover, in both studies, eye tracking is used in a double function, namely, as a process measure in itself, and as a cue for another process measure (verbal reports). This illustrates how eye-movement data can be used in multiple ways in research on learning and instruction.

2.4. Commentaries

The present special issue concludes with two commentaries on the contributions, by Richard Mayer and Jukka Hyönä. Mayer (2010) has adopted a multimedia learning perspective, discussing how eye tracking in general, and the contributions to the special issue in particular, can advance research on multimedia learning. Hyönä (2010) has taken an eye-tracking perspective in discussing the contributions, and provides suggestions for other potential approaches to data analysis in multimedia research that may further advance this upcoming area of eye-tracking research.

3. Conclusion and outlook

We believe that the studies presented in the present special issue can contribute relevant insights to multimedia research by using eye tracking to investigate how different design interventions (e.g., spoken vs. written text, cues, presentation speed) and (developing) expertise affect processing of animations, videos, and complex visual displays.

Concerning the role eye tracking might play in future research developments on learning and instruction, one should note that the possibility to replay records of eye movements can not only be used to study multimedia learning processes, and thereby (indirectly) contribute to better design of instructional materials or procedures. It may also be applied more directly in (multimedia) learning materials or procedures. For example, Kostons, Van Gog, and Paas (2009) used replays of participants’ own eye movement records as a tool to help them self-assess their task performance. However, reviewing records of others may also be useful for learners. The findings that experts and novices allocate their attention differently (see Canham & Hegarty, 2010; Jarodzka et al., 2010) might also imply that when an expert demonstrates a problem-solving procedure to a novice (as in a modelling example), the expert model and the novice learner may not necessarily be attending to the same things at the same time. It has been suggested that such a discrepancy in attention allocation may be reduced by showing the learner not only the experts’ actions in a modelling example, but also his/her eye movements, which might lead to better processing of the example and as a result enhanced learning outcomes (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). Findings by Velichkovsky (1995) suggest that it is possible to guide novices’ attention based on expert eye-movements, at least in real-time cooperative problem-solving. And similar results in favour of synchronizing the visual attention distribution between two persons have also been obtained in discourse scenarios, where it has been shown that a better comprehension of a dialogue is achieved if a listener attends to the same objects at the same time as a speaker (Richardson & Dale, 2005).

In conclusion, the contributions to this special issue provide a timely overview of what eye tracking has to offer research on learning and instruction, in particular research on learning with multimedia. With technological developments making eye tracking more accessible (i.e., not only making equipment more affordable, but also making it easier to gather and analyse data), we believe that different applications of eye tracking will continue to increase in our field to study learning processes and improve the design of instruction.

Acknowledgements

The guest editors would like to thank the reviewers of the special issue.

References


Tamara van Gog*

Centre for Learning Sciences and Technologies, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands

*Tel.: +31 45 576 2276; fax: +31 45 576 2907.

E-mail address: tamara.vangog@ou.nl

Katharina Scheiter

*Katharina Scheiter

Applied Cognitive Psychology and Media Psychology, University of Tuebingen, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany