

Gaze Transitions when Learning with Multimedia

Krzysztof Krejtz

University of Social Sciences and Humanities,
and National Information Processing Institute,
Warsaw, Poland

Andrew T. Duchowski

Clemson University, Clemson, USA

Izabela Krejtz

University of Social Sciences and Humanities,
Warsaw, Poland

Agata Kopacz

National Information Processing Institute,
Warsaw, Poland

Piotr Chrzastowski - Wachtel

Math, Informatics, and Mechanics, Institute of Informatics,
University of Warsaw, Warsaw, Poland

Eye tracking methodology is used to examine the influence of interactive multimedia on the allocation of visual attention and its dynamics during learning. We hypothesized that an interactive simulation promotes more organized switching of attention between different elements of multimedia learning material, e.g., textual description and pictorial visualization. Participants studied a description of an algorithm accompanied either by an interactive simulation, self-paced animation, or static illustration. Using a novel framework for entropy-based comparison of gaze transition matrices, results showed that the interactive simulation elicited more careful visual investigation of the learning material as well as reading of the problem description through to its completion.

Keywords: Multimedia learning, transition matrix entropy, scanpath comparison, working memory

Introduction

Multimedia learning materials increasingly make use of animations and interactive simulations to supplement or replace static (*book-style*) illustrations (Moreno & Mayer, 2007). Although such multimedia learning environments would appear to offer more powerful pedagogical tools than those with only static illustrations (Paas et al., 2007), their impact on learning is not yet clear. For example, according to Paik and Schraw's *illusion of understanding* hypothesis (2013), when people are learning with multimedia presentations, animation

affects metacognitive monitoring such that they perceive the presentation to be easier to understand and develop more optimistic metacomprehension. Consequently, learners invest less cognitive effort when learning with animation.

Learning effects can be different depending on the type of animation, representational, or referential. Representational animation, typically used to portray behavior of dynamic systems over time, may have a negative effect on learning by creating an illusion of understanding. In contrast, referential animation, by including elements of cueing of the viewer's attention toward a particular region of an image (e.g., flashing,

We thank Ms. Hanna Stachera and Ms. Regina Lewkowicz from *Staszic* High School in Warsaw, Poland, for their help in organizing the present study. We also thank Mr. Marek Młodożeniec, Ms. Ewa Domaradzka, and Dr. Rafał Albiński, for their help in conducting the study, and Ms. Karolina Chmiel for the preparation of experimental materials.

change of colors), may have a positive effect on learning by highlighting relevant visual elements that help the learner to integrate aural and visual components of multimedia (Paik & Schraw, 2013).

We are just beginning to understand when animation or interaction can foster learning and when it can overburden the learner's cognitive system (Mayer, 2010; Boucheix & Lowe, 2010). Interactive media, in contrast to animation, which allows the user to manipulate graphical elements, e.g., with the use of a mouse, is generally thought to offer learning benefits primarily due to the highly specialized experience afforded by the interaction. Interaction affords a self-paced, customized presentation of the material that the learner constructs via direct manipulation of the interactive tools. Compared to mass-produced video intended for a large audience, affording customization mainly via the familiar "VCR control" metaphor (e.g., pause, play, rewind, etc.), interaction offers a highly individualized learning experience.

Positive learning effects of interactive multimedia, if any, may derive either from its potential for individualized customization or perhaps from its potential for directing visual attention (e.g., its directive effect similar to that of referential animation). However, research designed to evaluate learning tends to focus on the measurement of retention and comprehension and not necessarily on the measurement of attentional distribution (Hegarty, 2004). Supplementing established empirical metrics of retention and comprehension, eye tracking methodology offers direct evidence of the distribution of visual attention during learning. Previous eye tracking studies have effectively utilized traditional eye movement metrics such as fixation counts and fixation durations, supplemented by heatmap and scanpath visualizations. Analysis of gaze transitions can provide additional insight into how attention switches between visual elements of the learning environment.

The purpose of this paper is to introduce the use of high-level analyses of eye movements, in particular entropy-based statistical comparison of transition matrices, into the multimedia learning domain. Although transition matrices have been employed previously in this context (multimedia learning), their use has been limited. Following a review of past work, we introduce K. Krejtz et al.'s (2014, 2015) framework for computing transition matrices given a set of Areas Of Interest (AOIs) defined

atop the stimulus area (e.g., computer screen). We provide details on how these matrices are quantitatively compared using empirical entropy, which allows computation of statistical significance between two or more conditions.

We use transition matrix analysis to compare visual attention when learning to solve the Towers Of Hanoi (TOH) problem using different forms of visualizations—static illustration, self-paced animation, or interactive simulation, each accompanying related textual information. We consider differences of visual attention distribution in relation to learners' working memory capacity.

Background

In this survey of previous work on multimedia learning, we focus on efforts in which eye tracking was used, noting that a statistical comparison between gaze transition matrices has not previously been applied in this context. Assuming that humans attend to and process visual information under fixation (Just & Carpenter, 1980; Hyönä, 2010), eye tracking is used to delve deeper into the cognitive processing that occurs during integration of textual and pictorial content.

Reading and scene perception yield different eye movement patterns. During reading, fixations last on average about 200-300 ms, however, when exploring a scene, fixations can range from under 100 to over 500ms. in duration, averaging about 300ms. (Rayner, 1998). Fixation duration and fixation counts are treated as indices of cognitive effort in information processing. For example, longer fixation durations on stimuli are indicators of greater processing difficulty. More important elements of a scene receive more attention (more and longer fixations), than scene elements that are less relevant to the task (Christiansen, Loftus, Hoffman & Loftus, 1991). Similarly, a number of eye tracking studies on reading has demonstrated that not all words are fixated equally. Longer, less frequently occurring content words are more likely to be fixated (Rayner, Pollatsek, Ashby & Clifton, 2012) than, for example, function words e.g., articles such as "the", "and", etc. (Rayner & McConkie, 1976).

Differences in reading and scene viewing are even more pronounced in saccadic characteristics. Due to typically higher stimuli density, shorter saccades of about 2 degrees are typical for reading, whereas saccades twice as long are observed during scene perception (Rayner, 1998).

Scanpath Comparison

In the context of multimedia learning, previous eye tracking studies mainly focused on visual attention deployed to animations with different design features such as spoken or written text (Schmidt-Weigand, Kohnert & Glowalla, 2010), different forms of cues (Boucheix & Lowe, 2010), or different presentation speeds (Meyer et al., 2010), taking into account learner characteristics such as prior knowledge (Jarodzka et al., 2010). Thus far, few studies explicitly investigated the process of information acquisition from multimedia learning materials with scanpath analysis (temporal sequence of gaze fixations), which allows for careful investigation of attention allocation by the analysis of transitions between different parts of the learning material (Duchowski, 2002).

An interesting example of scanpath analysis in the context of multimedia learning comes from Yoon and Narayanan (2004), who tested whether mental imagery used as a strategy of solving a problem is reflected in eye movement patterns. They reported that scanpaths of participants engaged in mental imagery while looking at a blank display reflected their scanpaths recorded when they previously looked at the diagram. Their scanpaths differed from those of participants who did not engage in mental imagery.

Scanpath analysis has also been used to compare eye movement patterns of experts and novices (Bednarik & Tukiainen, 2008; Rosengrant, 2010). For example, Bednarik and Tukiainen (2008) observed how programmers visually attend to program code and its animation. Novices relied on animation to formulate their mental model of the program, whereas experts first formed a mental model of the code and then used the animation to verify their hypotheses of the program's functionality.

In a special issue on eye tracking in the context of learning, Hyönä (2010) reviewed traditional eye movement metrics, including analyses based on fixation counts, durations, as well as scanpath similarity.

Transition matrices were not highlighted in the review. Although a good deal is known regarding animations and learning, most of the knowledge centers around product-related measures, e.g., what has been gleaned from comprehension (de Koning et al., 2010). Much less is known about how learners visually attend to instructional animations, that is, indicating the real-time perceptual and cognitive processes involved. Evaluations of the proportion of the number of fixations in each of a number of AOIs along with the proportion of total time fixated on each AOI are typical, but transition matrices could be helpful by indicating attentional switching between the AOIs.

Lowe and Boucheix (2011) note the sophisticated nature of cognitive processing of animation and the need for a better understanding of the various perceptual and cognitive activities that it involves. When animation accompanies text, quantitative analysis of the frequency of gaze transitions is likely to be helpful. Lowe and Boucheix compared attention distribution in terms of fixation durations between static and dynamic segments of the stimulus but they did not evaluate transitions per se.

More recently, Eitel et al. (2013) examined the so-called scaffolding assumption with analysis of recorded eye movements. They used heatmaps and rose diagrams as well as quantitative metrics to gauge the effect of audio inclusion in pictorial diagrams of pulley systems. Transition diagrams were not employed.

Van Meeuwen et al. (2014) used transition matrices but only evaluated mean transition differences, i.e., differences in the mean number of transitions between one pair of AOIs differing from a transition between another pair of AOIs. This is similar to SchmidtWeigand et al.'s (2010) two-celled transition matrix analysis. Both approaches essentially considered pairwise AOI transitions piecemeal, limiting the analysis to a small number of AOIs and transitions between them. Our matrix-based approach allows comparison of transitions between any number of AOIs and, via computation of entropy, offers a holistic comparison between all transitions performed (e.g., by individual participants in a given experimental condition).

Our approach is perhaps most similar to that of Jian et al.'s (2014) use of transition diagrams, however, as with previous examinations of transitions, only pairwise

transition comparisons were made. In other words, while employing transition diagrams, Jian et al. effectively compared transitions between corresponding diagram edges. Our transition matrices contain this information implicitly (matrix elements represent diagram edges), but due to computation of entropy, we are able to compare between two transition matrices (complete diagrams) holistically instead of element by element, i.e., computing a single number per transition matrix.

Before demonstrating the use of our transition matrix analysis in the particular context of multimedia learning, we first review prior work in this area, with emphasis on why multimedia learning environments are thought to offer potential for *deep learning* (Marton & Säljö, 1976). We also review the concept of working memory capacity and consider how it is likely to impact learning from interactive multimedia materials. We then present results of our eye tracking study in which we measured what is attended to using traditional eye movement metrics, and then employ transition matrices to gain insight into: (a) how much attentional switching there is between different media components, (b) which components are linked together during attentional switching, (c) how readers choose entry points and reading paths, and (d) how they integrate text and media when making sense of novel content.

Interactive Multimedia Learning Materials

Multimedia learning materials consist of at least two modal contents, namely textual and pictorial (Mayer, 2002). This definition also includes static (*book-style*) illustration accompanying text. Combining verbal and non-verbal knowledge representations can enhance understanding of the material (Schnotz & Horz, 2010; Krejtz et al., 2012). On the one hand, the learner's understanding of presented material may improve from an increase in the perceptual processing of relevant portions of the illustration (Holsanova et al., 2009). On the other hand, dynamic animations may interfere with knowledge acquisition by overloading cognitive resources (Ayres & Paas, 2007).

Currently, interactive multimedia learning is often associated with interactive applications providing multi-directional communication between the learner and instruction (Moreno & Mayer, 2007). Interaction makes a substantial difference on how knowledge is acquired. Students construct their knowledge in a selfpaced style by

selecting, organizing, and integrating new information, e.g., by manipulating graphical elements on the screen. This may require substantially more cognitive resources, but it may also lead to better understanding of learned material. For example, Nusir et al. (2012) showed that teaching children basic mathematics skills with multimedia materials affected their attention especially when cartoon characters were used. Similarly, in the context of medical education, Holzinger et al. (2009) presented evidence that interactive multimedia are cognitively demanding but are beneficial when additional guidance is provided and when students have sufficient previous knowledge of the topic.

Traditionally, most of the work evaluating interactive multimedia learning materials has focused on its instructional design principles such as spatial and temporal contiguity (Moreno & Mayer, 1999), the type of delivery media used (Mayer, 2002), cognitive load (Moreno & Valdez, 2005), or sense modalities used to receive information (Moreno & Mayer, 2001). Although multimedia learning requires complex cognitive processing, the relationship between multimedia learning and working memory, the core concept of recent cognitive theories, is less clear (Dutke & Rinck, 2006; Unsworth & Engle, 2007).

Impact of Working Memory Capacity

Working memory capacity (WMC) is an individual's ability to simultaneously process a primary task, maintain new information, and retrieve relevant information regarding the current task goal (Unsworth & Engle, 2007). Working memory theories differentiate working memory into subsystems which are responsible for processing information from different modalities and executive functions which control processing and integration of newly acquired information (Baddeley et al., 1998). Baddeley's (2000) model of working memory describes a central executive system, associated with controlled processing and attention, which coordinates operations of three subsystems: the phonological loop for speech-based information, the visuospatial sketchpad for visuospatial-based information, the episodic buffer (responsible for integrating information). The subsystems have limited capacity for parallel information processing (Baddeley, 1999). Working memory is strongly associated with the effectiveness of learning, and the mental processes of text comprehension and reading (Baddeley, 1986). High working memory capacity is

favorable for performance of complex cognitive tasks including attentional control (Engle & Kane, 2004), and mathematical performance (Ashcraft & Kirk, 2001).

Controlling for working memory capacity may explain variability in learning outcomes (Andrade, 2001). When faced with complex material, WMC may influence the strategies used to learn (Schuler et al., 2011). Gyselinck and Meneghetti (2011) reviewed studies focused on the role of working memory in processing text containing illustrations and reported consistent finding of the involvement of the visuospatial working memory during processing of illustrations. They pointed to the value of working memory in examining mechanisms involved in complex cognitive tasks.

Gyselinck et al. (2000) compared high and low working memory individuals on text comprehension in two conditions: with and without accompanying static illustration. Only learners with high visuospatial working memory benefited from the illustration. Gyselinck et al. concluded that sufficient visuospatial working memory capacity is required for effective illustration processing.

In an eye tracking study, Sanchez and Wiley (2006) tested comprehension in three conditions: text-only, text with relevant pictures, and text with irrelevant pictures, among high and low WMC individuals. Comprehension of text was lower for low WMC individuals when irrelevant pictures were presented. Monitoring of eye movements showed that learners with high WM capacity spent less time looking at irrelevant pictures, which suggests better control of attention.

An integrated cognitive model of text and picture comprehension put forth by Schnotz and Bannert (2003) describes how knowledge from different modalities is simultaneously acquired and integrated into mental representations. The integration and interaction between different modalities (textual and pictorial) starts at the early stages of information processing. The side effect of the early integration of different modalities is the high requirement for working memory resources. Assuming sufficient cognitive resources, continuous integration fosters creation of a common mental model that is modality-unspecific.

Dutke and Rinck (2006) presented empirical evidence for the link between working memory capacity and processing information from two modalities. They showed that integration of elements from different

sources (verbal descriptions and pictorial depictions) posed more demands on working memory resources than integrating information from one modality. In another study, segmentation of multimedia material facilitated deep learning and allowed high working memory capacity individuals to outperform those with lower working memory capacity (Lusk et al., 2008).

In our study, we expected to observe differences in visual attention distribution and patterns of gaze dynamics when learning from text accompanied by different types of visualization (static illustration, self-paced animation, and interactive simulation). Compared to the other two conditions, interactive simulation requires action (manipulation of parts of the simulation) in addition to comprehension and as such is more cognitively demanding. We expected that interactive simulation would induce continuous integration of textual and pictorial content which may depend on working memory capacity. We test this hypothesis by tracking eye movements, and use recorded gaze as an indicator of deployment of overt visual attention to the different visualization forms.

Hypotheses

In the present study, we use recorded scanpaths to ascertain how learners make use of graphical and textual information. Relying on the eye-mind hypothesis (Just & Carpenter, 1980), assuming gaze direction is linked with one's overt focus of attention, recorded eye movements offer insights into how and when cognitive load occurs and, in turn, how readers' eyes move in response to interaction.

To evaluate the efficacy of multimedia learning materials, we used an online learning web page with one of three types of visualizations of the Towers Of Hanoi algorithm: static illustration, self-paced animation (video), or interactive simulation. We hypothesized that patterns of attention allocation would differ as a function of type of visualization accompanying text, and of the learner's working memory capacity.

By referring to Schnotz and Bannert's (2003) integrated model of text and picture comprehension, we predict that interactive simulation fosters knowledge integration, inducing more systematic visual examination of the learning material (text and visualization) but at the same time poses more cognitive demands.

Method

The present study is a mixed design eye tracking experiment with the type of visualization (static illustration vs. self-paced animation vs. interactive simulation) as a between-subject manipulated factor and the type of Area Of Interest (lines of text and visualization) as a within-subjects factor with working memory capacity as a main controlled variable.

Participants

Sixty-three senior high school students took part in the study (15 F, 48 M, mean age 19). The school was chosen because it was profiled as one with its main focus on mathematics and computer science, consequently it provided a relatively unified sample due to motivation for learning mathematics and mathematical skills. Participants signed a consent form and were obliged to provide a signed consent from their legal caregiver. Due to technical problems during the experiment (calibration errors and low eye tracking ratio) the final sample consisted of 43 participants. The calibration error was reported separately on the x -axis ($M=0.57$, $SD=0.16$) and on y -axis ($M=0.56$, $SD=0.21$). The average tracking ratio was 88.36% ($SD=9.40$). Participants with poor calibration were invited to complete the experimental procedure for ethical reasons.

Stimuli and Procedure

The experimental procedure consisted of two tasks presented in random order. Participants began either with a working memory task or with a learning task. When learning, participants' eye movements were recorded.

Learning task. For the present study we chose the Towers Of Hanoi (TOH) as the algorithm that participants were asked to learn. The TOH problem is a classical puzzle-solving situation that does not involve domain-specific knowledge and hence is often used to investigate basic cognitive mechanisms such as search and decision-making mechanisms (Richard et al., 1993). The problem is usually presented as a planning task whose difficulties involve planning consecutive and correct moves, not as a problem involving restructuring. According to Richard et al. (1993), solving the TOH problem calls for elimination of misconceptions that are not consistent with the solution process. Although the solution to the problem can lead to modeling of understanding and solving a problem in general, we

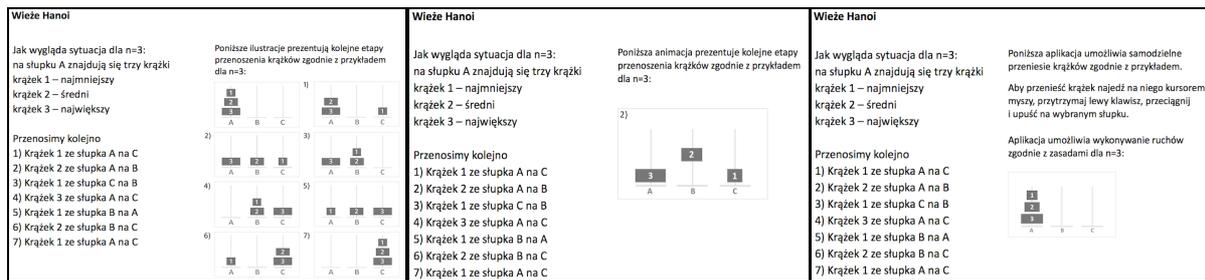
mainly use the problem to investigate gaze switching mechanisms. However, we do not model gaze switching per se, rather we show that it differs when interactive tools are made available to the learner.

Zanga et al. (2004) point out that the TOH is a welldefined problem in that the learner has access to all the information they need, namely a specific goal, described in the form of a state, and the rules for transformation. Often a self-paced animation is shown to participants, and testing situations arise where specific or non-specific goals are also depicted to viewers. We use a self-paced animation of the solution as one of our test conditions.

We use the *classical* TOH problem (Tijus et al., 2006), which consists of a stack of n disks of decreasing diameter stacked on one of three pegs, with two other pegs initially empty. The problem requires relocation of a stack of disks from the first peg to the third observing two rules: only one disk can be moved from one peg to another, and a disk with larger diameter cannot be placed atop one with a smaller diameter. The middle peg can be used in the process. The problem is discussed in many sources, including the text by Graham et al. (1994).

Learners were given the description of the problem in three consecutive web pages presented on separate screens describing the TOH problem. Each web page was presented in MS Internet Explorer 8 in full screen mode. Each participant's task was to learn about the problem. There was no time limit for learning. The first screen included an introduction and presentation of the problem along with plain text information on how to solve it. The next screen included a more specific description of the solution with specific steps that need to be performed in order to arrive at the solution, see Figure 1. On the same screen, together with the textual description of the problem, learners were presented with one of the three variants of visual aid:

1. a static illustrated sequence of 7 consecutive snapshots for each move for $n=3$, see Figure 1(a);
2. a self-paced animation, showing a visualization of continuous movements of 3 disks, which could be repeated on demand, see Figure 1(b); or
3. an interactive simulation (also with 3 disks) allowing the user to manipulate the disks with the use of a mouse, see Figure 1(c).



(a) static illustration

(b) self-paced animation

(c) interactive simulation

Figure 1. Example of Towers of Hanoi visualizations on screen 2.

The analyses presented in this article were performed on data collected during learning from the second page of the learning task, which differentiated the experimental conditions, see Figure 2.

Working memory task. Participants' working memory capacity was measured with a computerized version of the Visual Digit Span Task (backward version) (Conway et al., 2005), wherein a participant is presented with a series of digits, each appearing for one second on the screen. The goal is to remember the digits and then to immediately recall all numbers in reverse order. Successful trials are followed by ones where the number of digits is increased by one. The procedure stops with the second failed trial. The score is calculated as the maximal length of digits correctly recalled during all trials (Woods et al., 2011).

Apparatus

The Visual Digit Span was performed on a standard 15-inch PC screen. During the learning task, eye movements were recorded at 250Hz by an SMI eye tracking system, with spatial resolution of 0.03 degrees of visual angle, and gaze position accuracy of 0.4deg., according to the manufacturer. Participants were seated in front of a computer monitor (1680 × 1050 resolution; 22-inch LCD, 60Hz refresh rate). SMI's *Experiment Center* software was used to present stimuli and to synchronize with recorded eye movements. SMI's *BeGaze* software was used for fixation and saccade detection and raw data cleaning, with default settings used to classify fixations and saccades via *High Speed Event Detection*, a velocity-based algorithm (Salvucci & Goldberg, 2000). The peak velocity threshold was set to 40deg./s, the minimum saccade duration was set to 22ms., and the minimum fixation duration was set to 50ms.

Independent Variables

The experiment followed a factorial design with the type of visualization as a main independent variable at three levels: static illustration vs. self-paced animation vs. interactive simulation accompanying textual description. For the transition matrix analyses a withinsubjects independent variable Areas of Interest was created. AOIs were drawn around each line of the textual algorithm description (eight lines) and one around the visualization. Working memory capacity was a continuous predictor.

Dependent Variables

Results were analyzed in terms of time to learning completion, and eye movement characteristics. Eye movement characteristics were calculated for Areas of Interest around the textual description of the TOH algorithm and its visualization.

Learning completion time. Time to learning completion was calculated as the time from the onset of the second learning page to its offset. The learning time was self-paced. Participants were allowed to spend as much time with the page as they needed. Time to learning completion may be an indicator of effort required to understand the TOH algorithm.

Fixation count. Fixation count is a number of fixations on the second learning page during the whole learning completion period. Fixation count is an indicator of visual processing of selected stimuli during the learning process (e.g., the textual portions of the stimulus). We expected different types of visualizations to have an affect on fixation counts over multimedia learning materials.

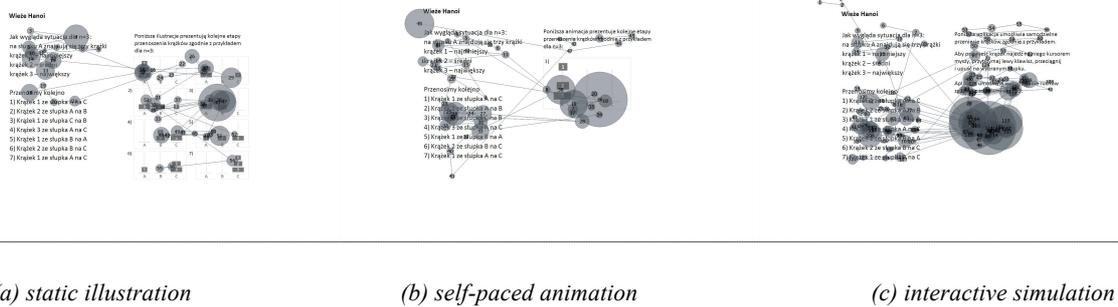


Figure 2. Towers of Hanoi problem with three types of multimedia instructional materials with recorded representative eye gaze scanpaths of relatively high working memory capacity individuals (static text at left is in Polish and explains the recursive algorithm).

Fixation duration. Fixation duration is an average duration of fixation measured in milliseconds. According to the literature, a longer fixation duration is associated with a deeper and more effortful cognitive processing of visual information, see (Just & Carpenter, 1980). For instance, more complicated texts or more complicated grammatical constructs increase fixation durations (Rayner, 1978, 1998; Rayner et al., 2012). Fixation duration may thus be considered as an indicator of effort needed for visual information processing.

Transition matrices. Gaze patterns were summarized for all participants with transition matrices (Ponsoda et al., 1995; Acartürk & Habel, 2012; K. Krejtz et al., 2014, 2015). In the present study, transition matrices were constructed from AOIs drawn around each line of the algorithm description and one AOI around the visualization and white space (rest of the stimuli/second page).

Each transition matrix cell represents the number of transitions from the AOI represented in the row to the corresponding column AOI. The value of each cell was normalized by the marginal sum of each cell's row, resulting in a probability score. Transition matrices were calculated for the second page of the TOH problem (the page with the visual aid for each visualization type).

Empirical entropy. To investigate how careful the reading was, empirical entropy, \hat{H} , was calculated for normalized transition matrices. Empirical entropy \hat{H} was calculated individually for transition matrices for each participant allowing us to treat it as a dependent variable in the statistical analysis. For technical details, see *Gaze Patterns and Transition Matrices* in the *Results*. Low entropy values may be associated with higher

predictability of eye movements (transitions between different AOIs) while high values of entropy indicate more random transition processes.

Results

Statistical analyses were conducted using *R* (R Development Core Team, 2011). To test the hypotheses a series of Multilevel Linear Models (MLMs) was constructed. These models allowed us to verify the influence of different types of visualizations and working memory capacity on dependent variables as well as to control for learning task completion time as a covariate. The MLMs also allow for estimation (and simple comparison) of means across different levels of predictors (Field et al., 2012). Eye tracking data were nested in the Areas of Interest (text vs. visualization). Contrasts for the type of visualization predictor assumed the static illustration as a baseline.

Prior to the main analyses with MLM, a check for outlying data points was performed for each dependent variable separately within each of the experimental conditions. The outlying data points were defined as those which lied beyond the extremes ($upper=Q3+1.5 \times IQR$, where $Q3$ is 3rd quartile and IQR is the inter-quartile range; lower extremes were defined analogously). Three data points for fixation count and four for learning task completion were identified as outliers and changed to mean plus or minus two standard deviations ($M \pm 2SD$). No outliers were found for fixation duration or transition matrix empirical entropy. A normality check for each condition was performed with the use of $K=kurtosis/2SE_{kurtosis}$ and $S=skewness/2SE_{skewness}$ indicators.

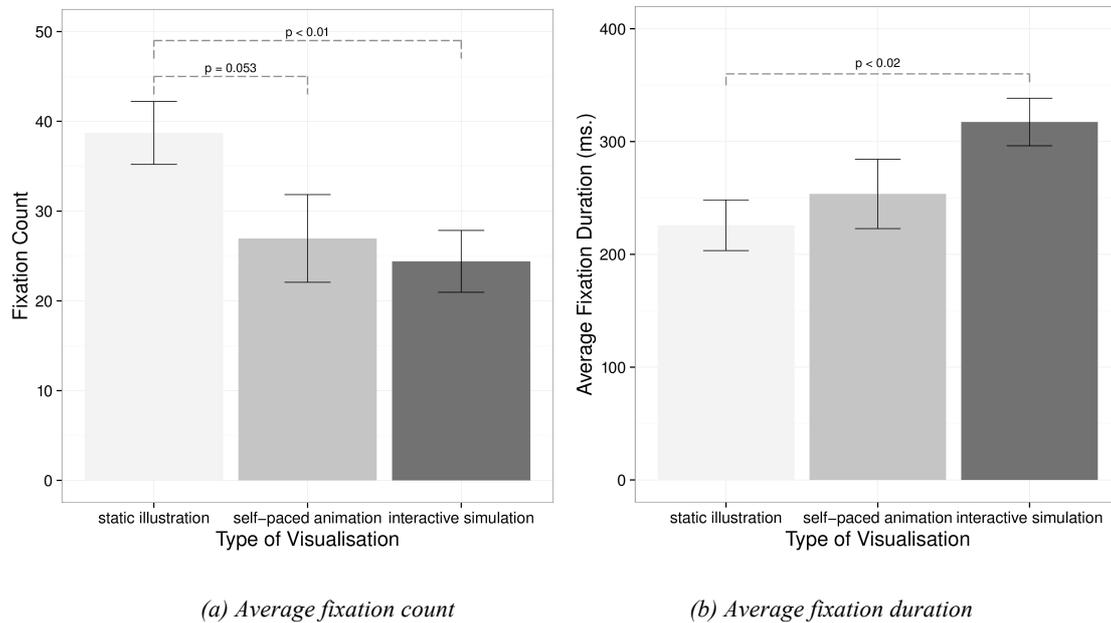


Figure 3. Fixation count and duration depending on visualization type. The whiskers represent $\pm 1SE$.

Time to Complete Learning

Our first hypothesis posited that an interactive simulation would prolong the time spent with the learning materials and that this relation would depend on working memory capacity of the learner. The tested model consisted of two fixed predictors, the type of visualization, and working memory capacity. The analyses revealed that the type of visualization had a statistically significant influence on learning time, $\chi^2(2)=6.49$, $p<0.05$ with $AIC=-1038.75$ and $BIC=-1026.84$. The analysis also revealed that working memory capacity had a marginally significant influence on learning time, $\chi^2(1)=3.61$, $p=0.057$, $AIC=-1040.36$, $BIC=-1026.07$. The contrast coefficients are reported for the latter model.

Time to completion was significantly longer while learning with the interactive simulation ($M=39253.37ms.$, $SE=3376.59$) than with static illustration ($M=28735.28ms.$, $SE=3548.65$), $b=10518.09$, $t(36)=2.21$, $p<0.05$. The difference in learning completion time between self-paced animation ($M=25943.16ms.$, $SE=4842.40$) and static illustration was not significant, $b=-2792.121$, $t(36)=-0.48$, $p>0.1$.

Working memory capacity influenced the learning completion time at a marginally significant level,

$b=22677.86$, $t(36)=1.87$, $p=0.069$. This suggests that the higher the working memory capacity the longer the time spent on learning with multimedia materials.

The interaction term between working memory capacity and the type of illustrations was not significant, $\chi^2(2)=3.73$, $p>0.1$.

Deeper Learning with Interactive Simulation?

We hypothesized that the interactive simulation would elicit more attentive learning and that working memory capacity would moderate this relation. To verify this hypothesis, two MLM analyses were performed on fixation count and fixation duration as dependent variables. In these analyses two predictors, namely visualization type (static vs. self-paced vs. interactive) and working memory capacity, were included. Additionally, for the analysis on fixation count we treated the learning time completion as the covariate.

Fixation count. The MLM analyses on fixation count revealed that learning completion time significantly influenced the overall number of fixations, $\chi^2(1)=30.49$, $p<0.001$, $AIC=711.56$, $BIC=723.47$. In line with the hypothesis, the analyses also showed that the type of visualization influenced fixation count, $\chi^2(2)=9.03$, $p<0.02$, $AIC=706.53$, $BIC=723.21$. The model coefficients are reported for the latter model.

The longer the learning completion time, the more fixations were captured, $b=0.01$, $t(36)=6.37$, $p<0.001$. A significant effect of visualization type revealed that participants exhibited significantly fewer fixations while learning with the interactive simulation ($M=24.40$, $SE=3.45$) than when learning with materials containing the static illustration ($M=38.70$, $SE=3.50$), $b=-14.30$, $t(36)=-2.82$, $p<0.01$, see Figure 3(a). Learning with the self-paced animation differed marginally from learning with the static illustration in terms of fixation count ($M=26.95$, $SE=95$), $b=-11.75$, $t(36)=-2.00$, $p=0.053$.

Finally, neither the working memory capacity predictor, $\chi^2(1)=0.70$, $p>0.1$, nor the interaction term, $\chi^2(2)=1.30$, $p>0.1$, significantly improved the model fit.

Fixation duration. In line with the hypothesis that the interactive simulation condition is cognitively demanding, we predicted longer fixation durations compared to the static illustration. A multilevel linear modeling analysis on fixation duration revealed that the type of visualization significantly predicted the dependent variable, $\chi^2(2)=8.57$, $p<0.02$, $AIC=992.80$, $BIC=1007.09$. Participants in the interactive simulation condition produced significantly longer fixations ($M=317.35ms.$, $SE=21.09$) compared to learning with the static illustration ($M=225.79ms.$, $SE=22.46$), $b=91.56$, $t(37)=2.97$, $p<0.01$, see Figure 3(b). At the same time, the average fixation duration was not significantly different between learning with the self-paced animation ($M=253.57ms.$, $SE=30.75$) or with the static illustration, $b=27.79$, $t(37)=0.73$, $p>0.1$.

The effect of working memory capacity on fixation duration was not significant, $\chi^2(1)=2.46$, $p>0.1$. The addition of the interaction term between visualization type and working memory capacity also did not significantly improve the model fit, $\chi^2(2)=4.81$, $p=0.09$.

Gaze Patterns and Transition Matrices

To discover specific gaze switching patterns during learning with different types of visualizations accompanying the textual description of the algorithm, analysis was carried out with gaze transition matrices. This analysis allowed us to disambiguate whether reading during learning progressed sequentially, similar to when reading regularly (Rayner, 1998), or was more in parallel to include the illustration, i.e., switching between text and visual.

There are potentially three different approaches to learning with the multimedia learning materials:

1. reading the textual description and then focusing on the visualization,
2. viewing the visualization and then reading the algorithm, or
3. systematically switching gaze between the two of them.

Such strategies may be beneficial for knowledge building as they are reflected in more predictable dynamical patterns of visual attention. The least effective strategy for knowledge acquisition would likely be a random (chaotic) pattern of attention switching between different elements of the multimedia learning material. We used fixation transition matrices and their empirical entropy measure to verify whether different types of visualization elicit different gaze transition patterns while learning with multimedia materials, see Figures 4(a)–4(c).

Transition Matrix and Calculation of Entropy

A transition matrix is a tool for sequential gaze pattern analysis (see Ponsoda et al. (1995) for an early example, where they used Z and χ^2 statistics to compare matrices, with matrices limited to cardinal (compass) saccade directions (i.e., N, NE, SE, etc.). In this paper, we present a method of computing transition matrices for any number of AOIs, based on Krejtz et al.'s (2014) Markov model, and provide a statistical method to compare them (Krejtz et al., 2015). Our use of entropy is a simplified form used for data aggregation that is straightforward to implement, resembling Goldberg and Kotval's (1999) suggested use of matrix density.

Formally, given a set of AOIs $S=\{1,\dots,s\}$ (the state space) and denoting a gaze fixation atop the i^{th} AOI, $A_t=i$, at time t , ($t=1,\dots,T$), the process describing a gaze transition from i^{th} source to j^{th} destination AOI is assumed to be modeled by a 1st order Markov process, fully determined by the initial source state. This allows the transition matrix to be defined as $P=(p_{ij})_{s \times s}$, where $p_{ij}=\text{P}(A_{t+1}=j \mid A_t=i)$, $t=0,\dots,T-1$, is the probability of changing gaze position from state i to j over AOI set S .

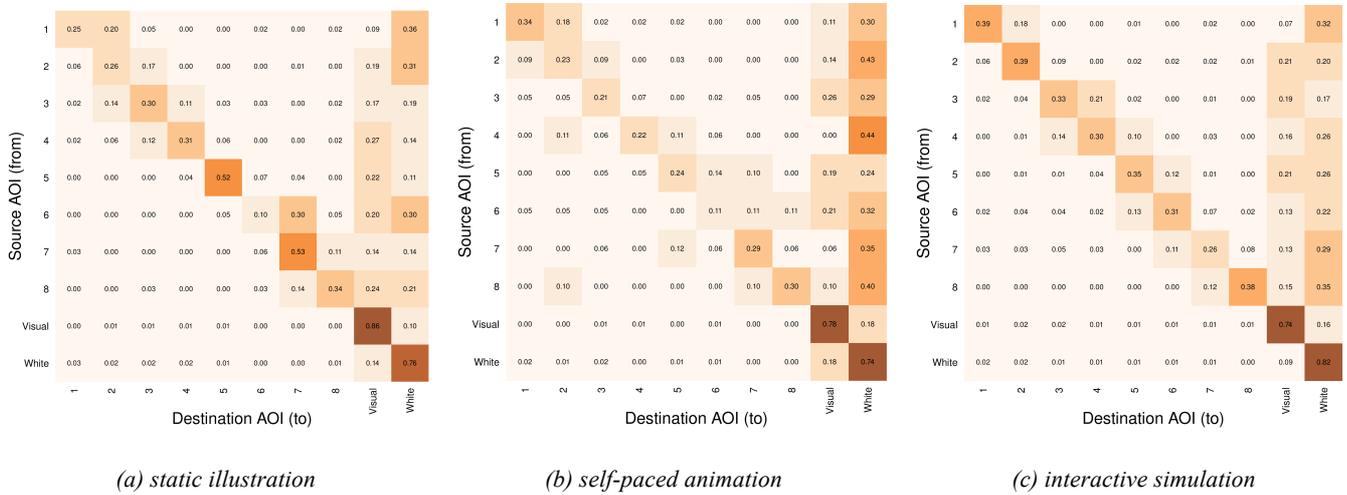


Figure 4. Fixation transitions. Digit labels (1–8) represent AOIs for each consecutive line of algorithm in the textual description. The label Visual refers to the AOI around the different visualization corresponding to the experimental condition, and the label White refers to the white space AOI (regions outside the textual algorithm description or the visualization AOI).

Transition matrix P can be computed in R (R Development Core Team, 2011) for each of the AOIs defined atop the stimulus image. Matrix elements p_{ij} are set to the number of transitions from i^{th} source AOI to j^{th} destination AOI for all participants and then the matrix is normalized relative to each source AOI (i.e., per row), such that p_{ij} represents the estimated probability of transitioning from i^{th} AOI to any j^{th} AOI given the i^{th} AOI as the starting point.

To compare the effect of task on gaze transitions, a statistical comparison of transition matrices is performed. Empirical entropy is computed via maximum likelihood. Empirical entropy \hat{H} , an estimate of Shannon's entropy, is defined as

$$\hat{H} = - \sum_{i,j} p_{ij} \log(p_{ij})$$

To facilitate statistical comparison of mean entropies per condition, \hat{H} is computed per each participant of the condition. Therefore, a transition matrix is computed as above but per individual and per condition. That is, entropy is computed from each individual's transition matrix, resulting in a table of $c \times n$ entropies for each of c experimental conditions and each of n participants.

Statistical procedures can then be employed to test for differences in mean entropy per condition.

Transition matrix entropy is an indicator of the randomness of fixation distributions between AOIs (Acartürk & Habel, 2012; Di Nocera et al., 2006). Entropy can be thought of as the number of differing matrix cells, akin to density dispersion. If every cell in the matrix contained the same probability value, entropy would be maximum, indicating equal likelihood of transitions from a given AOI to any other. Conversely, maximum likelihood of transition to any given AOI would suggest lower entropy. In the present situation, entropy is a convenient metric for numerically characterizing transition matrices such that they can be compared with traditional statistical tools. Note that the present straightforward approach assumes normal distribution of transition matrix cell values and does not consider Markovian stationarity as suggested by Krejtz et al. (2014, 2015).

The diagonal cells of the transition matrix represent subsequent fixations to the same AOI (i.e., line of text or visualization). The numbers above the diagonal represent gaze switching from the current line to the next, whereas values below the diagonal represent gaze switching in the opposite direction. The second-to-last column to the right

shows gaze fixation switches from the text to visualization. The second-to-last row (from the top) shows gaze switches from the visualization to each line of the algorithm description, see Figure 4.

In order to compare transition matrices between experimental conditions, empirical entropy \hat{H} was used. In line with predictions, multilevel linear models revealed that the type of visualization significantly predicted the empirical entropy of gaze transitions, $\chi^2(2)=21.85$, $p<0.001$, $AIC=-13962.38$, $BIC=-13945.10$. Participants exhibited more predictive (organized) patterns of gaze switching while learning with the interactive simulation ($M=0.38$, $SE=0.02$) compared to learning with the static illustration ($M=0.49$, $SE=0.02$), $b=-0.12$, $t(36)=-4.20$, $p<0.001$. Similarly, during learning with the self-paced animation, gaze transition patterns were also less random ($M=0.36$, $SE=0.03$) than in the static illustration condition, $b=-0.14$, $t(36)=-3.77$, $p<0.001$, see Figure 5.

Again, neither working memory capacity, $\chi^2(1)=0.81$, $p>0.1$, nor the interaction term between working memory and the visualization type, $\chi^2(2)=1.72$, $p>0.1$, significantly improved the model fit.

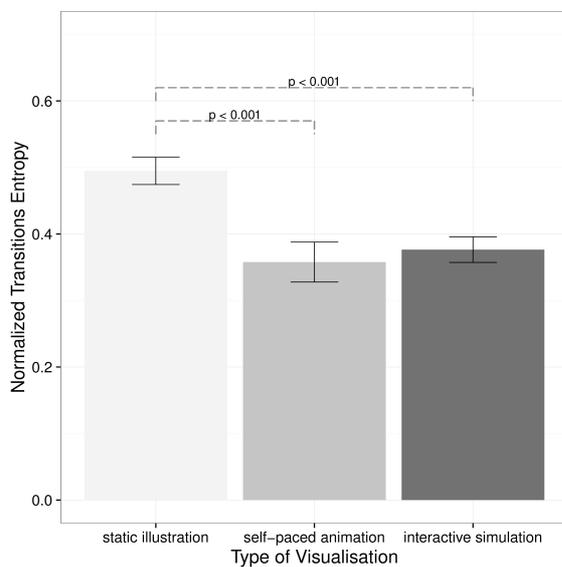


Figure 5. Mean empirical transition matrix entropy \hat{H} dependent on visualization type. The whiskers represents $\pm 1SE$.

Qualitatively, inspection of transition matrices, see Figures 4(a), 4(b), and 4(c), shows differences in attention switching patterns during learning with multimedia materials. The diagonals of the transition matrices tell us about the probabilities of re-fixating the

same AOI (e.g., line of the text or visualization). One may notice that while learning with the interactive simulation participants made consecutive fixations to all lines of the algorithm description with very similar probabilities for fixating each line (ranging between 0.30 and 0.39). At the same time, the distribution of fixation probabilities on each line of textual description accompanied by the static visualization or self-paced animation is more varied (from 0.10 to 0.53 and from 0.11 to 0.34, respectively). We may claim that relatively high probabilities and equally distributed fixations on the textual description lines may suggest a more attentive reading pattern and consequently indicate deeper learning (see Figure 2 for exemplary scanpaths). Further investigation is needed to support this interpretation.

Discussion

Our present study is based to a certain extent on Mayer's *multimedia effect*, which states in essence that it is better to learn from pictures and text than from text alone (Mayer, 2002). We assumed this was so and evaluated different forms of pictorial aids, through analysis of gaze metrics. In particular, we provided a means of analyzing dynamical gaze patterns in the presence of static illustrations as well as self-paced animations and interactive visual aids.

In general, in line with predictions, significant differences between the three types of visual aids in terms of eye movement characteristics and viewing patterns were obtained.

Analysis distinguished impact of the interactive simulation. Compared to the static illustration, the interactive simulation prolonged learning time. Participants also made fewer but longer fixations while learning with the interactive simulation accompanying the textual description. Gaze transition entropy was significantly smaller in the interactive simulation condition, implying less chaotic or more systematic visual inspection of the learning material. Moreover, qualitative investigation of transition matrices showed that participants made consecutive fixations to all lines of the algorithm description while learning with the interactive simulation. They did so consistently (with relatively high similarity and probability).

Impact of the self-paced animation is less obvious. This type of visualization did not influence the learning time compared to the static illustration. Participants in the self-paced animation condition exhibited marginally more fixations but with similar durations, compared to the static illustration. The dynamics of the gaze transitions were similar to the interactive simulation, that is, significantly less chaotic than in the static illustration condition.

Contrary to predictions, working memory capacity did not moderate the effects of the visual aids on eye movement characteristics. Results revealed its marginally significant role in predicting learning completion time, however. It was observed that the higher the working memory, the longer time spent with learning TOH problem. Finally, working memory capacity had no effect on the dynamics of gaze transitions between different Areas of Interest.

Our results suggest that the type of visualization accompanying textual description modifies the characteristics of gaze fixations as well as the pattern of gaze switching. We believe that the interactive simulation induces a more attentive visual investigation of learning material and deeper cognitive processing of the given information.

Results also provide some supporting evidence for the cognitive model of multimedia learning proposed by Schnotz and Bannert (2003). According to this model, the integration and interaction between different modalities (textual and pictorial) starts at the early stages of information processing. Assuming learners have sufficient cognitive resources, continuous and systematic switching between different elements of the learning material fosters integration of information and the creation of a common mental model that is modalityunspecific. The side effect of the early integration of different modalities is a high requirement for working memory resources (longer fixation durations may reflect this). This suggests that when learning with an interactive simulation, learners need to devote more cognitive effort (e.g., in this instance to understand the algorithm).

We also noticed that the interactive aid motivated reading of the problem description through to completion, as indicated by the equally distributed and relatively high probabilities of fixations counted over the consecutive

lines of the textual description. A larger number of fixations is also often associated with higher cognitive load (Henderson & Ferreira, 1990), suggesting that distribution of visual attention provides insight into learners' cognitive processing of the task.

One may speculate that more attentive visual inspection of the textual algorithm description as well as its visualization, when learning with the interactive simulation, is also an indication of the concept of *desirable difficulty* (Bjork, 1994). Bjork showed that learning is more effective when an elementary level of difficulty is provided, e.g., reading very small font. In the interactive condition, the solution to the problem required effort, motivating individuals to engage in more effortful cognitive processing. This claim is supported by gaze metrics, fixation duration, and longer learning time in the interactive simulation condition.

Consequently, one may claim that the interactive simulation elicited activation of learners' visual spatial cognition to test the solution provided by the textual explanation. Visual spatial cognition is an independent component of cognition, distinct from verbal and analytic abilities (Thurstone, 1938). It is the ability to hold the image of an object in mind and to twist, turn, or rotate it to match another object. This involves multiple processes, including perception, selection, organization, and the utilization of location- and object-based information (Possin, 2010). The goal is to structuralize interaction with learning materials (in our case). Thus, we posit that considering learners' working memory capacity yields a richer account of learning behavior than considering gaze dynamics in isolation.

Study Limitations and Future Directions

The present study suggested that interactive simulation leads to more deliberate visual inspection of the learning material (pictorial and textual), affecting the dynamical pattern of gaze switching between different parts of the multimedia learning material. We must note that our analyses focused on *process measures*, e.g., metrics related to eye movements and switching of overt visual attention, and stopped short of evaluation of *performance measures*, i.e., those related to learning outcomes (e.g., comprehension, retention, or learning transfer). However, the focus of our contribution is the provision of tools, namely transition matrix analysis, as a

means for others to use in helping corroborate their interpretations of such metrics.

We offer four directions for future research. First, future studies should verify the short and long term influence of learning with different visual aids on knowledge acquisition by including longitudinal testing (for example). Second, future work should also address the makeup or demographics of the population sample. In the present situation the sample consisted of highly motivated students fond of mathematics. Future analyses should consider different participants, particularly students who might not be as highly motivated. Third, working memory could be analyzed in terms of functional aspects. We relied on a widely used measure of working memory capacity (Conway et al., 2005). Oberauer et al. (2003), however, defines working memory as a set of three cognitive functions: (a) simultaneous storage and processing, (b) supervision (monitoring and control of ongoing cognitive operations), and (c) coordination of information elements into structures. It may be worthwhile to control for different aspects of working memory function in future studies of learning with multimedia. Finally, analysis of transition matrices, complementary to performance measures, may be useful in future studies of instructional material focusing on differences in processing between different types of text.

Conclusions

Can interactive simulations compete with text and *book-style* illustrations accompanying textual description in multimedia learning? We expected that learners would benefit more from interactive visual aids, by being motivated to better strategize their visual inspection of the learning material and to read more attentively the problem description through to completion. Indeed, when learners were allowed to manipulate the visual simulation and take control of the learning process, they would return to the textual information and continue reading the explanations in full. They then kept visually switching between different parts of the material in an organized way. Learners who lacked interaction (static illustrations), or received it in a limited way (self-paced animation), processed the information visually in a significantly different manner: their visual scanning strategy appeared more shallow (not completely reading

the textual elements) and random (as indicated by gaze transition matrix entropy). Our experiment shows that interactivity does not replace reading, however. On the contrary, interaction appears to spur reading, leading to a more complete visual inspection of the material.

References

- Acartürk, C. & Habel, C. (2012). Eye tracking in multimodal comprehension of graphs. In R. Cox (Ed.), *Tedr-2012: Proceedings of the workshop on technology-enhanced diagrams research* (pp. 11-25).
- Andrade, J. (2001). *Working memory in perspective*. Hove: Psychology Press.
- Ashcraft, M. & Kirk, E. (2001). The relationships among working memory, math anxiety and performance. *Journal of Experimental Psychology: General*, 2, 224-237.
- Ayres, P. & Paas, F. (2007). Making Instructional Animations More Effective : A Cognitive Load Approach. *Applied Cognitive Psychology*, 700(21), 695-700.
- Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. (1999). *Essentials of human memory*. Boston, MA, USA: Hove: Psychology Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*(4), 417-423.
- Baddeley, A., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*(105), 158-173.
- Bednarik, R. & Tukiainen, M. (2008). Temporal eye-tracking data: Evolution of debugging strategies with multiple representations. In *ETRA '08: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (pp. 99-102). New York, NY: ACM.
- Bjork, R. (1994). Memory and metamemory considering in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.
- Boucheix, J.-M. & Lowe, R. K. (2010, April). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, 20(2), 123-135.

- Christiansen, S., Loftus, E., Hoffman, H. & Loftus, G. (1991). Eye Fixations and Memory for Emotional Events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 693-701.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O. & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12, 769-786.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P. & Paas, F. (2010, April). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20(2), 111-122.
- Di Nocera, F., Terenzi, M. & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbor as a measure of mental workload. In D. de Waard, K. Brookhuis & A. Toffetti (Eds.), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 1-9). Maastricht, the Netherlands: Shaker Publishing.
- Duchowski, A. T. (2002). A Breadth-First Survey of Eye Tracking Applications. *Behavior Research Methods, Instruments & Computers (BRMIC)*, 34(4), 455-470.
- Dutke, S. & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction*(16), 526-537.
- Eitel, A., Scheiter, K., Schüler, A., Nyström, M. & Holmqvist, K. (2013, December). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48-63.
- Engle, R. W. & Kane, M. J. (2004). Executive attention, working memory capacity and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 145-199). Elsevier.
- Field, A., Jeremy, M. & Z., F. (2012). *Discovering statistics using R*. London, UK: SAGE Publications.
- Goldberg, J. H. & Kotval, X. P. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
- Graham, R. L., Knuth, D. E. & Patashnik, O. (1994). *Concrete mathematics: A foundation for computer science* (2nd ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Gyselinck, V., Ehrlich, M.-F., Cornoldi, C., de Beni, R. & Dubois, V. (2000). Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning*, 16, 166-176.
- Gyselinck, V. & Meneghetti, C. (2011). The role of spatial working memory in understanding verbal descriptions : a window onto the interaction between verbal and spatial processing. In A. Van Dierendonck & A. Szmalec (Eds.), *Spatial Working Memory* (pp. 159-180). Psychology Press.
- Hegarty, M. (2004, June). Dynamic visualizations and learning: getting to the difficult questions. *Learning and Instruction*, 14(3), 343-351.
- Henderson, J. M. & Ferreira, F. (1990, May). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 417-429.
- Holsanova, J., Holmberg, N. & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, 9(29), 1215-1226.
- Holzinger, A., Kickmeier-Rust, M., Wassertheurer, S. & Hessinger, M. (2009). Learning performance with interactive. simulations in medical education: Lessons learned from results of learning complex physiological models with the haemodynamics simulator. *Computers and Education*, 2(52), 292-301.
- Hyönä, J. (2010, April). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20(2), 172-176.
- Jarodzka, H., Scheiter, K., Gerjets, P. & van Gog, T. (2010, April). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146-154.
- Jian, Y.-C., Wu, C.-J. & Su, J.-H. (2014, August). Learners' eye movements during construction of mechanical kinematic representations from static diagrams. *Learning and Instruction*, 32, 51-62.
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.

- Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A. & Duchowski, A. (2012). Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study. In *Proceedings of the symposium on eye tracking research and applications* (pp. 99-106).
- Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A. & Villalobos, N. (2015, December). Gaze transition entropy. *ACM Trans. Appl. Percept.*, 13(1), 4:1-4:20.
- Krejtz, K., Szmidt, T., Duchowski, A. T. & Krejtz, I. (2014). Entropy-based Statistical Analysis of Eye Movement Transitions. In *Proceedings of the 2014 Symposium on Eye-Tracking Research and Applications*. New York, NY: ACM.
- Lowe, R. & Boucheix, J.-M. (2011, October). Cueing complex animations: Does direction of attention foster learning processes? *Learning and Instruction*, 21(5), 650-663.
- Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S. & Doolittle, P. E. (2008). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology*, 40(4), 636-651.
- Marton, F. & Säljö, R. (1976). On qualitative differences in learning: 1. outcome and process. *British Journal of Educational Technology*, 46, 4-11.
- Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, 41, 85-139.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instructions*, 20(2), 167-171.
- Meyer, K., Rasch, T. & Schnotz, W. (2010, April). Effects of animation's speed of presentation on perceptual processing and learning. *Learning and Instruction*, 20(2), 136-145.
- Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91, 358-368.
- Moreno, R. & Mayer, R. E. (2001). Virtual reality and learning: Cognitive and motivational effects of students' sense of presence. In *The Ninth International Conference on Human-Computer Interaction* (pp. 65-67).
- Moreno, R. & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19, 309-326.
- Moreno, R. & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, 53, 35-45.
- Nusir, S., Alsmadi, I., Al-Kabi, M. & Sharadgah, F. (2012). Studying the impact of using multimedia interactive programs at children ability to learn basic math skills. *Acta Didactica Napocensia*, 2(5), 17-31.
- Oberauer, K., Süß, H.-M., Wilhelm, O. & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31(2), 167-193.
- Paas, F., Van Gerven, P. & Wouters, P. (2007). Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology*, 21, 781-793.
- Paik, E. S. & Schraw, G. (2013). Learning with Animation and Illusions of Understanding. *Journal of Educational Psychology*, 105(2), 278-289.
- Ponsoda, V., Scott, D. & Findlay, J. M. (1995). A probability vector and transition matrix analysis of eye movements during visual search. *Acta Psychologica*, 88, 167-185.
- Possin, K. (2010). Visual spatial cognition in neurodegenerative disease. *NEUROCASE*, 16, 466-487.
- R Development Core Team. (2011). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria.
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618-660.
- Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 years of Research. *Psychological Bulletin*, 85, 618-660.

- Rayner, K. & McConkie, G. (1976). What guides a reader's eye movements. *Vision Research*, 16, 829-837.
- Rayner, K., Pollatsek, A., Ashby, J. & Clifton, C. J. (2012). *Psychology of Reading* (2nd ed.). New York, NY: Psychology Press.
- Richard, J.-F., Poitrenaud, S. & Tijus, C. (1993). ProblemSolving Restructurati: Elimination of Implicit Constraints. *Cognitive Science*, 17, 497-529.
- Rosengrant, D. (2010). Gaze scribing in physics problem solving. In *Proceedings of the 2010 symposium on eye tracking research & applications* (pp. 45-48). New York, NY: ACM.
- Salvucci, D. D. & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71-78). New York, NY, USA: ACM.
- Sanchez, C. & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory and Cognition*, 34(2), 344-355.
- Schmidt-Weigand, F., Kohnert, A. & Glowalla, U. (2010, April). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, 20(2), 100-110.
- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*(13), 141-156.
- Schnotz, W. & Horz, H. (2010). Technology and learning— support for skills learning. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (pp. 140-149). Oxford and Elsevier.
- Schuler, A., Scheiter, K. & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review*, 23(3), 389-411.
- Thurstone, L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Tijus, C., Poitrenaud, S. & Richard, J.-F. (2006). Human Problem Solving : Evidence for Contextual Categorization. In *Information processing and management of uncertainty in knowledge-based systems* (pp. 1100-1109). Paris, France: Editions EDK.
- Unsworth, N. & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104-132.
- Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. & van Merriënboer, J. J. (2014, August). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10-21.
- Woods, D., Kishiyama, M. M., Yund, E., Herron, T., Edwards, B., Poliva, O., Hink, R.F. & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*, 33(1), 101-111.
- Yoon, D. & Narayanan, N. H. (2004). Mental imagery in problem solving: an eye tracking study. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (pp. 77-84). New York, NY: ACM.
- Zanga, A., Richard, J.-F. & Tijus, C. (2004, February). Implicit learning in rule induction and problem solving. *Thinking & Reasoning*, 10(1), 55-83.