# Supporting Adequate Processing of Multimedia Instruction: Two Gaze-Based Interventions

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

vorgelegt von
Dipl.-Psych. Carina Schubert
aus Lohr am Main

Tübingen
2016

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

Tag der mündlichen Qualifikation: 12.09.2016
Dekan: Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter: Prof. Dr. Katharina Scheiter
2. Berichterstatter: Prof. Dr. Ulrike Cress

# Acknowledgements

Finishing this dissertation has been an interesting, educational, challenging, and at times exasperating venture. Althoug only my name appears on the cover, there are many people without whom this work would not have materialized.

Of course, I want to thank my supervisors Katharina Scheiter and Anne Schüler. Their expertise, honest feedback, various ideas, and eye for detail contributed greatly to this dissertation.

I would like to express my gratitude to Ulrike Cress for being my second reviewer.

I am indebt to Gottfried Zimmermann, Benjamin Wassermann and Holger Schmidt for developing the computer technology for this work. I would also like to appreciate the collaboration with the members of WissenschaftsCampus Tübingen, particularly cluster *Adaptable and Adaptive Multimedia Systems*.

Furthermore, I want to thank my research assistants, especially Charlotte Prill for being a very patient Eye Movement Model.

My former co-workers at Leibniz Institut für Wissensmedien provided me with helpful comments, interesting ideas, restorative coffee breaks and memorable lunches. Special thanks go out to Kim Stalbovs and Katrin Schleinschok for being my loyal roommates, and for always making me laugh.

None of this would have been possible without my family: Christfried Pumm, and the Schuberts Cornelia, Andrea, Florian, and Anne.

Johannes Blinkle has been my rock during the final stages of this work. Thank you.

Last but not least, I would like to thank my friends for their support and all the good times.

# Table of contents

# 1 Introduction

The use of multimedia instruction in education is ubiquitous. Combinations of text and visualization appear in schoolbooks, learning software, animations and, more recently, in e-books. Teachers and learners commonly suppose that adding static or animated visualizations to text enhances motivation, makes learning more fun and, most importantly, makes the content easier to understand and leads to better learning. In fact, research verifies that multimedia often leads to better learning than text alone (for an overview, see Anglin, Vaez, & Cunningham, 2004). This *multimedia effect* (Mayer, 2005) is explained in theories of multimedia learning like the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009) and the Integrative Model of Text and Picture Comprehension (ITPC; Schnotz, 2002).

However, learners do not always benefit from multimedia instruction (for an overview, see Levie & Lentz, 1982). Learning from multiple sources of information is cognitively demanding; learners need to relate the information from text and pictures (Seufert, Brünken, & Zander, 2005) and translate between representations (Ainsworth, 1999). Consequently, to learn from multimedia instruction successfully, learners have to apply several cognitive strategies. Poor learning outcome might therefore be associated with inadequate knowledge about - or self-regulation of - cognitive strategies.

Beyond identifying individual differences and relevant cognitive processes, research aims at supporting multimedia learning. Thereby, there are basically two different approaches. The first approach provides external support. More specifically, instructional material is designed in a way that prompts adequate processing. For example, color-coding of corresponding elements in text and picture leads to more adequate visual processing and better learning (Ozcelik, Arslan-Ari, & Cagiltay, 2010).

The second approach aims at supporting learners to self-regulate multimedia learning by imparting relevant cognitive strategies. Such interventions encompass strategy worksheets (Kombartzky, Ploetzner, Schlag, & Metz, 2010; Schlag & Ploetzner, 2010) and strategy trainings (Scheiter, Schubert, Gerjets, & Stalbovs, 2014).

In this dissertation, I conducted research on two gaze-based interventions and used eye-tracking methodology to support the learning process.

The first intervention falls into the category of external support: an adaptive system that analyzes learners' gaze behavior online and detects inadequate processing of the material (Experiments 1 and 2). The system then alters the instruction to prompt adequate processing when necessary. For example, assume that a learner does not show enough transitions between text and picture. This might indicate that (s)he does not integrate information from the two representations (Johnson & Mayer, 2012; Mason, Pluchino, & Tornatora, 2015; Mason, Tornatora, & Pluchino, 2013). The system reacts by showing a version of the learning page where corresponding text-picture elements are color-coded to prompt integrative transitions.

The second intervention falls into the category of self-regulation support: an exemplary modeling of adequate visual processing of the instruction, or Eye Movement Modeling Examples (EMME). Learners view a video of a model's eye movements on exemplary material (Experiment 3) or the actual learning material (Experiment 4). They are instructed that the model is a successful learner. The idea is that the learner then applies the adequate processing strategies visualized in the EMME to the learning material.

This dissertation is structured as follows. Chapter 1.1 summarizes theories on multimedia learning. In Chapter 1.2, the eye tracking methodology and empirical findings on processing of multimedia instruction are described. Against the backdrop of these findings, I conducted two lines of research which are described in Chapters 2 and 3. Chapter 2 contains research on external support of multimedia learning by an adaptive system. I first provide a short review on research on external support of multimedia learning (Chapter 2.1) and on adaptive systems in the educational context (Chapter 2.2). I then report Experiment 1, which aimed at developing a framework for the adaptive system, in Chapter 2.3. Based on the findings of this experiment, I derived which gaze parameters and threshold values indicate inadequate processing of the multimedia instruction and the adaptive system was designed accordingly. This system analyzes learners' viewing behavior in a multimedia learning environment and adapts the instruction accordingly. The system is described in detail in Chapter 2.4. To evaluate the adaptive system, it was compared with a static system regarding learning outcome in Experiment 2, which is also reported in Chapter 2.4. In Chapter 3, I report two studies on supporting self-regulated adequate processing by Eye Movement Modeling Examples (EMME). I first provide an overview on research on self-regulation support (Chapter 3.1) and on EMME (Chapter 3.2). Chapters 3.3 and 3.4 report two experiments on the effectiveness of EMME for multimedia learning (Experiments 3 and 4, respectively). In each of the two experiments, a different version of EMME was compared with control groups on learning outcome. In Chapter 4, the findings from the four experiments and their implications are discussed.

## 1.1 Cognitive theories of multimedia learning

The effects of adding pictures to text are explained in theories of multimedia learning like the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2005) and the Integrative Model of Text and Picture Comprehension (ITPC; Schnotz, 2002).

CTML (Mayer, 2009) is the most influential theory in the field of multimedia learning. It conceptualizes multimedia learning as a set of cognitive operations during which both representations are processed individually in different channels - the verbal and the pictorial channel, respectively. This dual-channel assumption is based on Paivio's (1986) dual coding theory and Baddeley's (1986) conceptualization of working memory. Paivio (1986) postulates that pictorial and verbal information are processed in two different channels. According to Baddeley (1986), working memory contains a central executive and two slave systems: the phonological loop and the visuo-spatial sketchpad. In line with these theoretical accounts, Mayer postulates that multimedia instruction is processed in two different channels. CTML (Mayer, 2005) states that pictures and words enter our sensory memory through the eyes or the ears. In sensory memory, exact copies of visual or auditory representations can be held for a very short time. Relevant elements from text and pictures are then selected and enter verbal and visuo-spatial working memory. In CTML, working memory is defined as the memory store where the central processes of multimedia learning take place. It is used to hold and consciously process knowledge for a limited timeframe. In line with the limited-capacity assumption (Mayer, 2005), only a small amount of information can be processed in working memory at a time. This assumption is based on Baddeley's (1986) conceptualization of working memory. The information from both representations is then organized into a verbal and a pictorial mental model of the learning content. Both mental models are integrated with prior knowledge retrieved from long-term memory and result in a coherent representation of the learning content, i.e., a mental model. This mental model is then stored in long-term memory, which is unlimited in capacity. Figure 1 shows a summary of CTML.

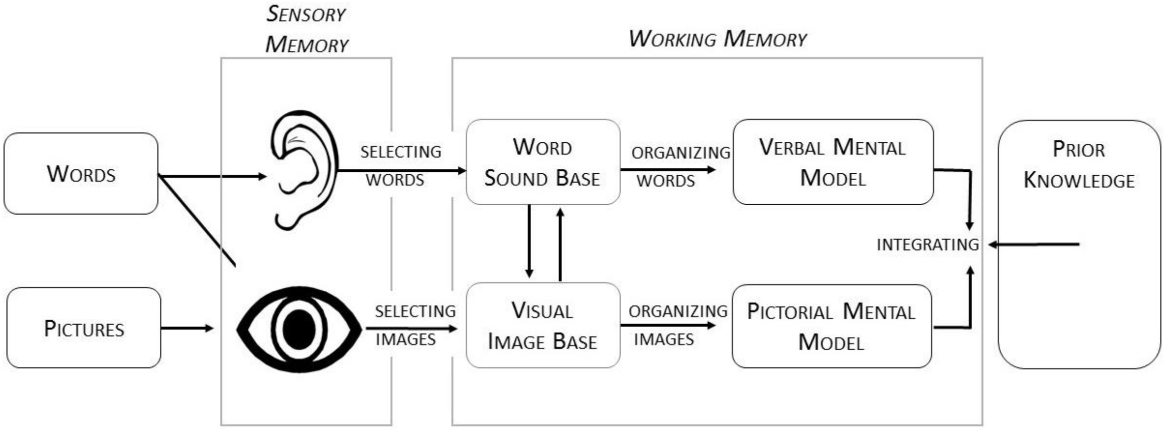


*Figure 1*. Cognitive Theory of Multimedia Learning (adapted from Mayer, 2005).

Similarly, ITPC (Schnotz & Bannert, 2003; Schnotz, 2002) assumes that, in several steps, an integrated mental model is formed based on the text and the picture whose information is connected (*coherence formation*). Like CTML, ITPC refers to Baddeley’s (1986) working memory model and distinguishes a verbal and a visuo-spatial working memory which are limited in capacity (Schnotz, 2002). Other than CTML, however, text and picture comprehension are not conceptualized as parallel processes. Descriptive representations like text or equations and depictive representations like images or other symbols are comprehended in two different branches of cognitive processes (see Figure 2). A learner understands text by first extracting a text surface structure, constructing a text base and finally a mental model of the text content. Picture processing occurs by first creating a visual mental representation of the picture and then constructing a mental model as well as a propositional representation of the content. ITCP postulates a constant interplay between text and picture processing through *coherence formation*. This coherence formation occurs through two processes: model inspection and model construction. In text comprehension, model construction transforms the propositional representation into a mental model. In picture comprehension, model inspection is used to add information from the mental model - which also contains prior knowledge - to the propositional representation.

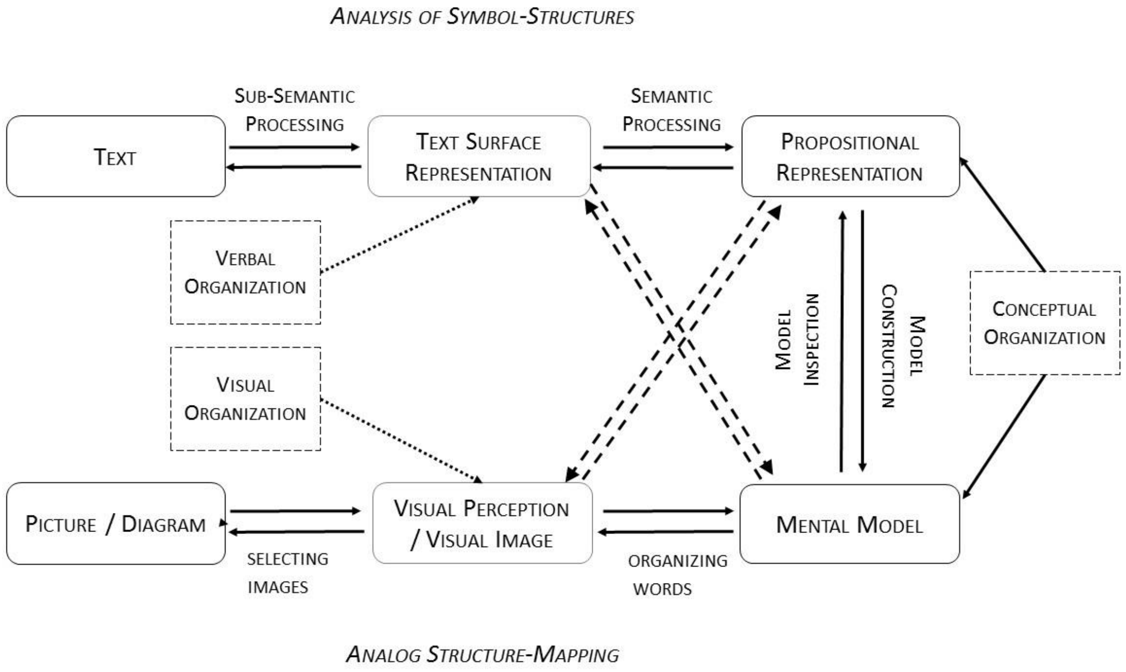


*Figure 2.* Integrative Model of Text and Picture Comprehension (adapted from Schnotz, 2002).

Both models have in common that learning from text and pictures occurs in several steps: the extraction of information from both representations, organization in separate mental representations and some form of interaction between both mental representations (*integration* or *coherence formation*, respectively). The resulting mental representation is richer than one derived from text alone, which explains the benefits of multimedia instruction.

According to CTML, the prerequisite for these steps to be effective is that the learner engages actively in the learning process (active-processing assumption, Mayer, 2005). Active processing, in turn, requires successful self-regulation. Learners may differ regarding how well they self-regulate the relevant cognitive strategies, which could explain why the multimedia effect does not always occur. To understand when and why multimedia learning is effective, it has been investigated which processes occur during successful learning.

The problem is that cognitive processes - like selection, organization and integration of text and picture elements - are covert and cannot be observed directly. Researchers need to find ways to get insight into learners' cognition. In addition, such process assessment should not interfere with the learning process or even alter it.

In the past years, the collection of gaze data, or *eye tracking* has been used in educational research to investigate the processing of materials. In the following section, I describe this methodology (Chapter 1.2.1) and findings of eye tracking research on multimedia learning (Chapter 1.2.2).

## 1.2 Eye tracking in multimedia learning

### 1.2.1 Introduction to eye tracking

Eye tracking can provide insight into the learning process because it shows what the learner attends to when, and in what order (Hyönä, 2010). At the base of interpreting eye data is the *eye-mind assumption*, which states that a learner fixates something as long as he/she is processing it (Just & Carpenter, 1980). Eye tracking has been applied in reading research, scene perception, and in visual search (for overviews, cf. Duchowski, 2007; Rayner, 2009). In the case of multimedia, there is also a growing body of research deploying this methodology (Scheiter & Van Gog, 2009; Van Gog & Scheiter, 2010).

Eye tracking is a method that follows the movements of a person's eyeball(s) in order to find out where the person is looking at a given moment in time. Modern eye tracking devices are relatively easy to use and not invasive, and thanks to advances in both hardware and software development they usually provide good data quality. The device used in this thesis is a remote system; the eye tracker is positioned below the computer screen where the learning material is displayed. The system uses the pupil-and-corneal reflection method (for a detailed description of eye tracking methods see Holmqvist et al., 2011). Thereby, an infrared light is aimed at the eyes, which produces the corneal reflection. The eye tracking camera identifies this reflection as the lightest spot on the eye. The pupil is identified as the darkest spot. When the eyeball moves, the distance between the corneal reflection and the pupil changes, and the software uses this to calculate where the gaze is directed. Other variables like distance of the eyes to the stimulus screen and coordinates of the monitor are used in these calculations. The eye tracking data is then aggregated into two different parameters by the system's software: fixations and saccades (Duchowski, 2007; Holmqvist et al., 2011). Fixations include small eye movements within a very small range; the eye is relatively stable and information intake can occur. In contrast, saccades are fast movements over a larger spatial range that move the eye's focus from one stimulus to another (Duchowski, 2007). Information intake is suppressed during saccades.

To interpret the eye data, it is usually further edited with software. For this purpose, so-called areas of interest (AOI) are defined around relevant aspects of the instruction. In multimedia learning, AOIs can be text segments or (certain parts of) the illustration. The software then computes different parameters that describe a participants' viewing behavior regarding each AOI. For example, the fixation count indicates how often a participant looked at a certain AOI, and the number of transitions indicates how often a participant looked from one AOI to another. There are a variety of gaze parameters, and depending on the research question different parameters are investigated. In the next section, I describe parameters used in multimedia learning research and related empirical findings.

### 1.2.2 Text-picture comprehension: results from eye tracking research

Multimedia research usually employs eye tracking measures that are related to the cognitive processes relevant to text-picture comprehension. As discussed above, these processes encompass selecting, organizing and integrating information from text and picture.

To investigate selection and organization processes, eye tracking is used to determine how much attention a learner pays to text and picture. More specifically, it is analyzed how often and how long a learner looks at text and picture. Therefore, the overall number of fixations and the total fixation times are computed. More frequent and longer fixation of a representation is interpreted as more intense processing.

To investigate integration processes, the number of saccades between the representations, or text-picture transitions, is analyzed. Frequent transitions are interpreted as a learner’s attempt at integrating the information from text and picture.

Table 1 provides an overview of measures used in multimedia research and the associated cognitive processes.

Table 1: *Eye tracking measures and associated cognitive processes (according to Johnson & Mayer, 2012)*

| **Measure** | **Description** | **Cognitive processes** |
|---|---|---|
| Fixation time text | Total duration of all fixations on text segments | Selection/organization: attentional focus on text |
| Fixation time picture | Total duration of all fixations on picture | Selection/organization: attentional focus on picture |
| Fixation count text | Total number of fixations on the text segments | Selection/organization: attentional focus on text |
| Fixation count picture | Total number of fixations on the picture | Selection/organization: attentional focus on picture |
| Transitions text-picture | Total number of looks from text to picture | Integration: attempts at integrating elements from text and picture |
| Transitions picture-text | Total number of looks from picture to text | Integration: attempts at integrating elements from text and picture |

Hegarty and Just (1993) examined learners' viewing behavior when learning from an illustrated text on the functioning of pulley systems and analyzed transitions between the representations. Their results suggest that learners construct mental models of the learning content stepwise by reading several clauses of text and then integrating information from the pictures, as indicated by transitions from text to picture.

Hannus and Hyönä (1999) conducted two experiments to investigate the effects of illustrations on learning and the relation between cognitive abilities and processing of illustrated text. In their second experiment, they had low and high cognitive ability children learn illustrated text passages from a biology textbook. They found that, in general, the viewing behavior was mainly text-driven:

the proportion of fixation time spent on the picture was very small. Furthermore, high-ability children spent more time on relevant text segments and on relevant pictures and made more transitions. This finding could be interpreted as high-ability children engaging in more selection, organization and integration processes.

In a study with fourth graders, Mason et al. (2013) examined viewing behavior in an illustrated science lesson. Using cluster analysis, they identified three different patterns of viewing behavior. Learners in these clusters differed in the amount of transitions between text and picture: one cluster showed very few transitions, one an intermediate amount and one showed many transitions. According to Mason et al., the three clusters can be characterized as low, intermediate and high integrators, respectively. Importantly, learners who fell into different clusters of viewing behavior also achieved different learning outcomes. High learning outcome was associated with a high amount of integrative transitions during learning.

Taken together, the empirical evidence discussed above concurs with the assumptions of theoretical accounts of multimedia learning. More specifically, learners' viewing behavior suggests that processes of selection, organization and integration are relevant to learning. Selection and organization of pictorial information and integration of the information from text and picture seem to be especially important for multimedia learning, as high-ability learners show higher picture fixation times and more transitions. Individual differences in the implementation of these processes, especially integration processes, are reflected in differences in learning success. Poor learners seem to focus on the text and neglect the picture. Therefore, it is possible that they fail to exploit the benefits of multimedia instruction. To enhance multimedia learning, selection, organization and integration of relevant information from text and picture should be supported. As discussed above, this can either be done by external support or by self-regulation support. In the next chapter, I report research I conducted on a gaze based adaptive system that provides external support. In chapter 3, I describe my research on self-regulation support using EMME. .

# 2 External support for multimedia learning: a gaze based adaptive system

In this chapter, I provide an overview on previous research on external support of multimedia learning, or *instructional design*. Evidence suggests that learners benefit from well-designed materials; however, there is also evidence that not all learners need this kind of support (cf. Chapter 2.1). Adaptive systems take individual differences into account and provide tailored support, accordingly. I describe such systems and their use in educational technology in Chapter 2.2. In Chapters 2.3 and 2.4, I report two experiments on a gaze-based adaptive system for multimedia learning.

## 2.1 Research on instructional design

Instructional design aims at presenting multimedia materials in a way that prompts successful processing. More specifically, well-designed materials reduce extraneous processing (needed to cope with a confusing layout); as a consequence, limited working memory is freed, which can then be invested in essential processing (needed to understand the content of the instruction) (Mayer, 2005). For example, according to the spatial contiguity principle, people learn more when corresponding elements from text and pictures are presented close to one another rather than in a split format (Moreno & Mayer, 1999). The finding that a split format inhibits learning is also referred to as the split-attention effect (Ayres & Sweller, 2005; Chandler & Sweller, 1992). This effect is explained within a Cognitive Load Theory (Sweller, 1988) framework by increased extraneous cognitive load which hinders learning (Chandler & Sweller, 1992). A meta-analysis by Ginns (2006) revealed that the detrimental effects of a split presentation of text and picture are a robust empirical finding. Johnson and Mayer (2012) conducted three experiments to investigate the effects of integrated presentation of text and picture on cognitive processes and learning outcome. Learners received a lesson on how car brakes work in either an integrated or a separated format. Learners in the integrated groups achieved better transfer test scores than learners in the separated group. Furthermore, learners made more integrative transitions in the integrated group. This indicates that the integrated presentation of the instruction prompted them to integrate information from both representations.

Another instructional design measure that has proven helpful is *signaling* or *cueing.* A signal can be any kind of typographical or visual aid that is used to make the instructional message clearer. Importantly, a signal does not add any content to the instructional message (Mautone & Mayer, 2001). A common way to signal a multimedia message is to highlight corresponding elements in text and picture, for example by color-coding. Signaling improves learning (Mautone & Mayer, 2001; Ozcelik et al., 2010; Scheiter & Eitel, 2015). Evidence from eye tracking suggests that signals foster text-picture integration processes (Scheiter & Eitel, 2015).

There is a number of other design principles derived from CTML which have been shown to promote multimedia learning (for an overview, see Mayer, 2005). The redundancy principle (Mayer & Johnson, 2008), for example, refers to better learning outcome when printed keywords are added to the picture and the narrated text. The modality principle (Moreno & Mayer, 1999) states that presenting pictures and narrated text yields better learning than pictures and written text.

Although the design principles discussed above have often been found to support multimedia learning, there is also evidence that this is not the case for all learners. The expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) describes the finding that beneficial effects of instructional techniques disappear or even invert for learners with a higher level of expertise. This effect has been reported for the split-attention effect (Kalyuga, Chandler, & Sweller, 1998; Yeung, Jin, & Sweller, 1998) as well as the modality effect (Kalyuga, Chandler, & Sweller, 2000). There is also evidence that signaling only enhances learning outcome for learners with a low level of prior knowledge (Richter, Scheiter, & Eitel, 2016), indicating that there is an expertise reversal effect and that some learners perform well without additional support.

Similar effects are known from reading research: in a study by McNamara and Kintsch (1996), high-coherence text was helpful for low-knowledge students while high-knowledge students learned better from low-coherence text.

Furthermore, eye tracking research suggests that some learners already use pictures extensively and integrate spontaneously, even when the instruction is not presented in an integrated way. For example, in the study by Hegarty and Just (1993), successful learners looked at the related picture element upon finishing a semantically meaningful text unit - even though the materials were presented in a split format. Likewise, Mason, Tornatora, and Pluchino ( 2013) showed that one out of three group of learners spontaneously integrated text and pictures, as indicated by a high number of transitions, whilst the other two groups showed only an intermediate or even a low number of transitions.

In conclusion, these findings suggest that there are individual differences in the processing of multimedia instruction, and that some learners show adequate visual processing without instructional support. Providing those learners with additional support might interfere with their learning process; to enable self-regulated learning, a learner should receive only as much instructional support as strictly necessary (cf. SEASITE principles, Renkl, 2002). This notion is also referred to in literature on computer-supported collaborative learning; too much external support can hinder a learner’s self-regulated application of appropriate scripts in collaborative learning and is referred to as *overscripting* (Fischer, Kollar, Stegman, & Wecker, 2013). Similarly, providing too much external support in multimedia learning may result in some form of overscripting.

Taken together, a one-size-fits-all approach to supporting multimedia learning may not be the best solution. Designing a form of support that takes individual differences into account and provides adaptive support for each learner is a promising research line. The last decades' advances in technology made the development of such adaptive systems possible. Several systems were developed in the educational context; they are addressed in the next section.

## 2.2 Adaptive technologies in educational research

According to a framework by Shute and Zapata-Rivera (2012), adaptive technologies connect the learner to educational resources like learning objects or pedagogical agents. They do so through the use of a learner model in a cycle with four processes: capture, analyze, select, present. *Capture* refers to the collection of information on the learner, like cognitive data or emotional states. *Analyze* means that the learner's state is put into relation to the domain, that is, the computer can infer what the learner knows or can do based on the learner's performance. In the *select* process, content is selected for the learner based on his/her status as represented in the learner model. Finally, in the *present* process, content is presented to the learner based on the results of the select process.

In the context of learning, adaptive technologies are mainly implemented in the design of intelligent tutoring systems. Thereby, the students' learning process is accompanied by the system, which, for example, provides feedback or gives hints on how to proceed.

For instance, Cognitive Tutors (Koedinger & Corbett, 2006) use a cognitive model to monitor students' performance and provide context-specific instruction accordingly; they also monitor students' learning and select problem-solving tasks that are adapted to the individual learner's abilities. Such Cognitive Tutors have been developed for mathematics (Koedinger, 2002), computer programming, and genetics (Corbett, Kauffman, Maclaren, Wagner, & Jones, 2010). In addition, there are some intelligent tutoring systems where metacognitive support was added to the functions of Cognitive Tutor. Help Tutor (Aleven, McLaren, Roll, & Koedinger, 2006; Roll, Aleven, McLaren, & Koedinger, 2011), for example, provides learners with metacognitive feedback on their help-seeking behavior.

Another intelligent tutoring system is AutoTutor (Graesser, Chipman, Haynes, & Olney, 2005). AutoTutor is a conversational dialogue based tutor; the dialogues are centered around so called main questions which the student has to answer via reasoning and explanation. AutoTutor tracks students' knowledge states and adaptively manages the dialogue, for example by providing feedback, prompts, and identifying misconceptions. In addition to monitoring learners' cognitive states, a derivative of AutoTutor called Gaze Tutor (D'Mello, Olney, Williams, & Hays, 2012) also tracks students' emotional states. Thereby, in addition to the functions of AutoTutor, the system analyses learners' eye movements to identify boredom or disengagement and reacts with a dialogue that refocuses the learners' attention towards the pedagogical agent.

Beyond the examples above, there are several adaptive technologies that are applied in educational settings, including but not limited to quantitative modeling (Jameson, 1995), machine learning (Webb, Pazzani, & Billsus, 2001), and Bayesian networks (Conati, Gertner, & VanLehn, 2002). To my knowledge, however, there is no adaptive system supporting cognitive processes relevant to multimedia learning. Furthermore, the systems described above have in common that they diagnose learning progress and/or knowledge gaps as opposed to processing strategies. They use specific tasks like problem-solving tasks or recall questions to model the learner's state of knowledge before and/or after the learning process. In contrast, the adaptive system presented in this thesis aims at diagnosing and supporting adequate processing during the learning process.

In the next two chapters, I describe the design and evaluation of a multimedia learning system that adapts to each learner's viewing behavior and provides individually tailored support.

As eye-tracking research has shown, learners differ in the way they visually process multimedia instruction, and these differences relate to learning outcome (e.g., Mason et al., 2013). Consequently, an online analysis of eye movements can serve as an indicator for learners' success. When a learner shows inadequate visual processing, an adaptive system can alter the instruction in a way that prompts adequate visual processing which, in turn, should support learning.

Based on the four-process adaptive cycle (Shute & Zapata-Rivera, 2012), the design of such an adaptive system encompasses knowing which eye tracking parameters to *capture* and how to *analyze* them. Furthermore, a decision is required on how to *select* the content that is then *presented* to the learner. For this purpose, I conducted Experiment 1 where I investigated learners' viewing behavior with the same learning material that was later used for the adaptive system. I analyzed which visual processes distinguish successful from less successful learners. Based on these findings, a gaze based adaptive system was designed. Experiment 2 investigated whether this system supports learning.

## 2.3 Experiment 1

The aim of this study was to investigate how to design a multimedia learning system that adapts to learners' gaze behavior. More specifically, I wanted to examine which eye tracking parameters are relevant and, as a consequence, how the system should adapt. Therefore, participants completed a learning session on cell division while their eye movements were recorded. To be able to identify successful visual processing in terms of learning, learning outcome was assessed. There is evidence that a learner's level of prior knowledge needs to be taken into consideration when interpreting eye tracking parameters. For example, long fixations may indicate comprehension difficulties in low prior knowledge learners, but elaborated processing in high prior knowledge learners (Schwonke, Berthold, & Renkl, 2009). Therefore, cognitive prerequisites (i.e., pre-existing knowledge) were measured to investigate if a learner's level of prerequisites needs to be taken into consideration for the adaptive system. Cluster analysis was used to identify patterns of viewing behavior and their relation to learning success.

Based on previous research, which indicates that the processes of selecting, organizing and integrating information from text and pictures are relevant to learning outcome, this study employed eye-movement measures that are related to these processes (cf. Table 1). As measures for selection and organization processes, it was analyzed how long learners processed text and picture and how often they did so. To obtain information on integration processes, it was analyzed how often a learner looked from the text to the picture and vice versa.

This experiment addressed the following research questions:

1) Can successful vs. unsuccessful learners in a multimedia learning session be distinguished based on viewing behavior (Research Question 1)?
2) Which gaze parameters are relevant for this distinction and should be implemented into the adaptive system (Research Question 2)?
3) Is the level of cognitive prerequisites a necessary parameter for identifying successful vs. unsuccessful learners and should it therefore be implemented into the adaptive system (Research Question 3)?

### 2.3.1 Method

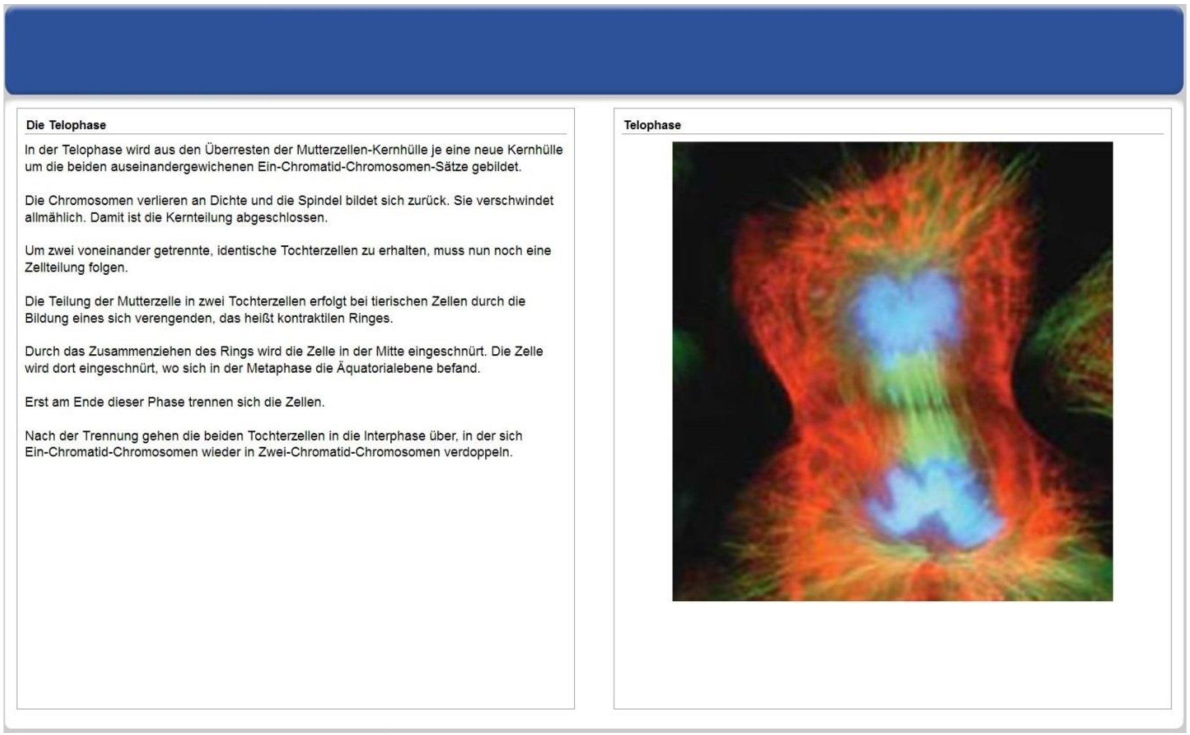
#### *2.3.1.1 Participants and design*

Participants were 32 students of the University of Tuebingen (28 female; *M* = 23.03 years, *SD* = 5.59) enrolled in different courses. All subjects had normal or corrected to normal vision and received either course credit or payment for their participation. Two different measures of learning outcome (recall and transfer) were assessed as dependent variables.

#### *2.3.1.2 Materials and apparatus*

*Learning materials*. Learning materials consisted of a written and illustrated German text on cell division. In part 1, which consisted of 4 pages, important cell structures, the concept of DNA and its storage in chromosomes, and basic concepts relevant to mitosis were explained. In part 2, which consisted of 6 pages, the interphase and the five phases of mitosis were described. This part described (a) how chromatin fibers contained in the nucleus of the cell duplicate into pairs, (b) chromatin fibers condense into chromosomes, the nuclear envelope breaks into fragments, the duplicated centrosomes move away from each other, while microtubules lengthen between them resulting in the mitotic spindle, (c) two chromatids of each chromosome develop a kinetochore, microtubules attach to them and move the chromatids towards the center of the cell, (d) chromosomes become aligned along the equatorial plane, (e) sister chromatids separate from one another and develop into chromosomes which move towards opposite poles in the cell, and (f) segregation of daughter cells with genetically identical material is promoted by a contractile ring.

Each page showed text on the left and one corresponding static picture on the right (cf. Figure 3). The text had an overall length of 1,180 words; text lengths on each slide varied between 51 and 200 words. The text passages were divided into semantically meaningful units and presented as paragraphs. The pictures illustrating the text were schematic pictures of cell structures and processes relevant to cell division. The text segments described the content on a more abstract level, provided the relevant terminology and explained how processes during mitosis are accomplished and why they are important. The pictures illustrated the visuo-spatial aspects of cell structures and of the processes the cell undergoes during mitosis. Thus, text and pictures were complementary and were both necessary to understand the learning content.



**Die Telophase**

In der Telophase wird aus den Überresten der Mutterzellen-Kernhülle je eine neue Kernhülle um die beiden auseinandergewichenen Ein-Chromatid-Chromosomen-Sätze gebildet.

Die Chromosomen verlieren an Dichte und die Spindel bildet sich zurück. Sie verschwindet allmählich. Damit ist die Kernteilung abgeschlossen.

Um zwei voneinander getrennte, identische Tochterzellen zu erhalten, muss nun noch eine Zellteilung folgen.

Die Teilung der Mutterzelle in zwei Tochterzellen erfolgt bei tierischen Zellen durch die Bildung eines sich verengenden, das heißt kontraktilen Ringes.

Durch das Zusammenziehen des Rings wird die Zelle in der Mitte eingeschnürt. Die Zelle wird dort eingeschnürt, wo sich in der Metaphase die Äquatorialebene befand.

Erst am Ende dieser Phase trennen sich die Zellen.

Nach der Trennung gehen die beiden Tochterzellen in die Interphase über, in der sich Ein-Chromatid-Chromosomen wieder in Zwei-Chromatid-Chromosomen verdoppeln.

*Figure 3*. Exemplary page of the learning material in Experiments 1 and 2.

*Apparatus*. Eye movements were collected using a SMI RED 250 eye tracker manufactured by SensoMotoric Instruments. This remote eye tracking system features a 22-inch TFT-Monitor and a sampling rate of 250 Hz. Participants were seated at a 65 cm distance to the screen. Data was recorded using SMI ExperimentCenter$^{TM}$ software and prepared for statistical analysis using SMI BeGaze$^{TM}$ software.

#### *2.3.1.3 Measures and scoring*

Measures encompassed cognitive prerequisites, posttest performance, and gaze data. The measures will be described in the following.

*Cognitive prerequisites.* The questionnaire was paper-pencil based and without time limit. To achieve a broad assessment of cognitive prerequisites, I used two different measures: a test of general scientific literacy and a test of prior knowledge on cell division.

The Test of Basic Scientific Literacy (Laugksch & Spargo, 1996) taps general knowledge about scientific concepts. I used 24 items from the Life Sciences scale. Participants read statements and have to judge if they are correct or incorrect, or unknown (e.g., “Life on earth has only been existent since a few thousand years ago”). Each correctly answered item was scored one point, resulting in a maximum total score of 24.

The test of prior knowledge on cell division consisted of 15 multiple choice items. Each item had four alternatives and one correct answer. The items asked about cell elements, genetics, and mitosis (e.g., “What are Mitochondria? (a) structures which are responsible for energy generation, (b) enzymes, (c) proteins, (d) structures which transport nutrients). Correct answers were scored one point, incorrect answers were scored zero; maximum total score was 15.

Both scales were z-standardized and summed to achieve one comprehensive measure of cognitive prerequisites. Internal consistency of the collapsed measure was Cronbach’s α = .74.

*Posttest.* The paper-pencil based posttest consisted of three subtests: a free recall question, a fill-in-the-gaps task and 16 multiple-choice-items.

The free recall question asked participants to write down everything they know about mitosis. Time limit for this open question was six minutes. Answers were scored using a coding scheme that awarded one point for each correctly mentioned concept (max. score 38). Two independent raters scored all participants and discussed their ratings until total agreement was achieved.

The fill-in-the-gaps task had no time limit and assessed recall; it consisted of segments from the learning text where key words were left out and had to be completed (e.g., “The period between two cell divisions, i.e., two mitoses, is called _________”). Each correctly filled gap was scored one point.

Multiple-choice items had four alternatives. Seven items assessed recall and nine assessed transfer. The items were either text- or picture-based (e.g., text-transfer: “Colchicin is a poison that inhibits the formation of microtubules. Which process would be impaired as a result of colchicine-poisoning? (a) duplication of the chromosomes, (b) condensation of the chromosomes, (c) dissolving of the nuclear envelope, (d) separation of the sister chromatids). Again, there was no time limit. Each correctly answered item was scored one, incorrect answers were scored zero points.

Mean scores from the free recall task, the fill-in-the-gaps task and from the seven multiple choice recall items were averaged and transformed into percentages to obtain a measure of recall test performance. The mean score of the nine multiple-choice transfer items was computed and transformed into percentages to obtain a measure of transfer test performance.

#### *2.3.1.4 Eye-movement measures*

For the analysis of learners’ eye movement data, areas of interest (AOIs) were defined for each page. To determine the number and duration of fixations on both representations as well as the number of transitions between them, one AOI was created around the text and one around the picture. Eye movement data was averaged across all pages of the learning material. For each participant, the mean time per page spent on the text (fixation time text) and on the picture (fixation time picture), the mean number of times the text/picture was looked at (fixation count text and fixation count picture) as well as the mean number of transitions between text and pictures and vice versa were computed.

##### *2.3.1.5 Procedure*

Data collection took place in individual sessions. Participants first were handed out written information on the proceedings of the experiment, signed a consent form, and completed the cognitive prerequisites test. They were then seated in front of the eye tracker, which was calibrated for each participant using a nine-point calibration. The written, onscreen instruction informed participants that they could learn at their own pace and proceed to the next page by pressing the space key but not go back in the material. Participants were also informed that they would be tested on the content. Afterwards, participants completed the learning session on mitosis while their eye movements were recorded. After the learning phase, they filled out the paper-pencil posttest, were debriefed and received compensation.

### 2.3.2 Data analysis

With the aim of designing an adaptive multimedia system, this experiment addressed individual differences in the processing of multimedia material and their relation to learning outcome. More specifically, I wanted to investigate if there are patterns of viewing behavior that characterize successful and less successful learners. For this purpose, I used cluster analysis, a method that can identify the underlying structure of data. Since this procedure is not very common in educational research, I describe cluster analysis and subsequent analyses in this section.

Cluster analysis is a procedure that groups sets of objects based on their proximity in multivariate space. This proximity reflects the similarity among objects along multiple dimensions (i.e., eye tracking parameters). Thus, clusters are formed in a way that objects in the same cluster are more similar to each other than to objects in different clusters. I employed Ward's method, a commonly used hierarchical model based on the squared Euclidian distance as proximity measure. Each case is first regarded as a group and groups are then successively combined in a way that keeps within-cluster variance as low as possible. That means that with each step, the most similar clusters are combined to a new cluster, until only one cluster remains. The optimal number of clusters can be determined by identifying the step in the cluster analysis where a further merging of clusters would result in an unacceptable large increase in within-cluster variance (cf. stepsize criterion, Johnson, 1967).

To assess the quality of the cluster solution, homogeneity of the clusters can be investigated by computing an *F*-score for each variable where the within-cluster variance is divided by the overall variance. A ratio < 1 indicates that the variance between clusters is larger than within clusters, thus indicating good homogeneity. Furthermore, as a measure of effect size, the difference between cluster mean and overall mean is calculated for each variable and divided by the overall standard deviation. This measure indicates if the cluster profile is distinct from the overall sample.

To investigate which parameters contribute most to the distinction of the clusters, Backhaus, Erichson, Plinke, and Weiber (1996) suggest submitting the data to a discriminant analysis. Thereby, participants are assigned to the previously defined clusters based on the parameters used for the clustering. The standardized discriminant coefficients are a measure for the relative contribution of each variable to the distinction of the clusters with high absolute values indicating a high contribution.

To investigate how the cluster profiles differ from each other, and how they can be interpreted, an ANOVA with cluster as factor is conducted for each of the clustering variables.

### 2.3.3 Results

For all statistical analyses reported here, the level of significance was set at $\alpha = .05$. Statistical analyses were conducted with IBM SPSS Statistics 20. Due to poor gaze data quality, three subjects were excluded from analyses.

#### *2.3.3.1 Clusters of gaze behavior*

To identify patterns of viewing behavior, the eye tracking measures were submitted to a cluster analysis. Since I was interested in the role of cognitive prerequisites in the processing of multimedia materials, the same cluster analysis was run twice, once with and once without submitting cognitive prerequisites as an additional variable.

For both cluster analyses, based on the stepsize criterion it was decided to terminate the clustering after three clusters had been formed. In both cases, cluster 1 consisted of 11 participants, cluster 2 of 13 participants, and cluster 3 of 5 participants. Comparison of the two cluster solutions with and without cognitive prerequisites showed that the very same participants were allocated to the same clusters in both cases (100% agreement). Thus, additionally considering cognitive prerequisites did not yield a different cluster definition than defining clusters based on eye movement parameters alone. In the following, I will nevertheless report results from the cluster analysis containing cognitive prerequisites to further investigate its relative contribution for arriving at meaningful clusters.

To evaluate the quality of this three cluster solution, each cluster's homogeneity for all parameters was investigated. The resulting $F$-scores are shown in Table 2 and indicate that cluster 1 and cluster 2 were very homogenous, since within-cluster variance is smaller than overall variance for all variables except cognitive prerequisites in cluster 2. Cluster 3 was not as homogenous which was likely due to the small cluster size.

Table 2: *Homogeneity indices and standardized discriminant coefficients for the eye tracking variables as a function of cluster in Experiment 1*

| **Variables** | **Cluster 1 (*n* =11)** | **Cluster 2 (*n* =13)** | **Cluster 3 (*n* =5)** | **Discriminant coefficient** |
|---|---|---|---|---|
| Cognitive prerequisites | .65 | 1.02 | 1.80 | -.20 |
| Fixation time on text (ms) | .48 | .83 | .83 | .43 |
| Fixation time on picture (ms) | .54 | .15 | .83 | .98 |
| Fixation count text | .18 | .15 | .74 | 1.30 |
| Fixation count picture | .60 | .18 | 1.11 | -.01 |
| Transitions text_picture | .51 | .25 | 1.62 | -.61 |
| Transitions picture_text | .74 | .59 | 1.47 | -.77 |

*Note.* Homogeneity is the ratio of within-cluster variance relative to the overall variance.

Furthermore, as suggested by Backhaus, Erichson, Plinke, and Weiber (1996), the eye tracking data and cognitive prerequisites were submitted to a discriminant analysis. The seven parameters were used to allocate each subject to one of the previously defined three clusters. In all 29 cases, the discriminant analysis assigned students to the cluster that had been determined in the cluster analysis, indicating that the three groups identified in the cluster analysis discriminate the underlying patterns in the data very well. Furthermore, the discriminant coefficients (cf. Table 2) suggest that all variables except for cognitive prerequisites and fixation count on the picture, which had negative discriminant coefficients, contributed to the distinction of the three clusters.

As a measure of effect size, the difference between cluster mean and overall mean was calculated for each gaze variable and divided by the overall standard deviation. This measure indicates if the cluster profile is distinct from the overall sample. Negative values indicate that a variable is underrepresented in the cluster compared to the overall sample (i.e., lower cluster mean than grand mean) whilst positive values mean that a variable is overrepresented (i.e., higher cluster mean than grand mean). As can be seen in Table 3, all gaze parameters were overrepresented in clusters 1 and 3 and underrepresented in cluster 2. This pattern is reversed for cognitive prerequisites.

Table 3: *Standardized differences between cluster means and overall mean for cognitive prerequisites (z-standardized) and eye tracking variables as a function of cluster in Experiment 1*

| **Variables** | **Cluster 1 (*n* =11)** | **Cluster 2 (*n* =13)** | **Cluster 3 (*n* =5)** |
|---|---|---|---|
| Cognitive prerequisites | -.10 | .35 | -.53 |
| Fixation time on text (ms) | .22 | -.83 | 1.66 |
| Fixation time on picture (ms) | .60 | -.85 | .89 |
| Fixation count text | .20 | -.79 | 1.61 |
| Fixation count picture | .36 | -.75 | 1.15 |
| Transitions text_picture | .44 | -.73 | .93 |
| Transitions picture_text | .42 | -.57 | .55 |

To compare the three clusters on cognitive prerequisites and the eye tracking measures, multiple one-factor ANOVAs and Bonferroni-adjusted post-hoc tests were performed.

Table 4 shows means and standard deviations for the six eye-movement parameters and cognitive prerequisites as a function of cluster, as well as the results from ANOVAs and post-hoc comparisons. All overall comparisons on the gaze variables were highly significant, indicating that learners in the different clusters showed distinct patterns of viewing behavior. Cognitive prerequisites did not differ significantly between clusters.

For fixation times and fixation count on text, all post-hoc comparisons were significant. Fixation times and fixation count were highest in cluster 3, followed by cluster 1, and lowest in cluster 2. Fixation times on pictures were significantly higher in clusters 1 and 3 than in cluster 2; clusters 1 and 3 did not differ significantly. The same pattern emerged for fixation count on pictures. Learners in cluster 1 and 3 looked more often from text to picture than learners in cluster 2. Cluster 1 and 2 also differed significantly on transitions from picture to text, with more transitions in cluster 1.

Table 4: *Means and standard deviations of cognitive prerequisites (z-standardized) and gaze behavior as a function of cluster in Experiment 1*

| **Variables and ANOVA results** | **Cluster 1 (n =11)** | | **Cluster 2 (n =13)** | | **Cluster 3 (n =5)** | | **Post-Hoc (Bonferroni adjusted)** |
|---|---|---|---|---|---|---|---|
| | **M** | **SD** | **M** | **SD** | **M** | **SD** | |
| Cognitive prerequisites F(2, 26) = 1.49, p =.25, n.s. | -0.10 | 0.81 | 0.35 | 1.01 | -0.53 | 1.34 | — |
| Fixation time on text (ms) F(2, 26) = 63.20, p <01, $\eta_p^2$ = .83 | 50217.78 | 542.47 | 24338.41 | 7092.00 | 85682.24 | 22401.80 | 1 vs. 2 **<br>1 vs. 3**<br>2 vs. 3** |
| Fixation time on picture (ms) F(2, 26) = 21.72, p < .01, $\eta_p^2$ = .63 | 4551.39 | 1405.89 | 1751.81 | 744.98 | 5110.32 | 1746.34 | 1 vs. 2 **<br>2 vs. 3 ** |
| Fixation count text F(2, 26) = 42.20, p < .01,$\eta_p^2$ = .76 | 199.65 | 34.41 | 119.35 | 32.04 | 315.02 | 7.42 | 1 vs. 2**<br>1 vs. 3**<br>2 vs. 3** |
| Fixation count picture *F*(2, 26) = 15.87, *p* < .01, $\eta_p^2$ = .55 | 16.32 | 5.53 | 8.38 | 3.02 | 21.92 | 7.51 | 1 vs. 2**<br>2 vs. 3** |
| Transitions text_picture *F*(2, 26) = 12.01, *p* < .01, $\eta_p^2$ = .48 | 6.23 | 1.95 | 3.02 | 1.37 | 7.56 | 3.48 | 1 vs. 2**<br>2 vs. 3** |
| Transitions picture_text *F*(2, 26) = 4.87, *p* < .05; $\eta_p^2$ = .27 | 3.85 | 1.12 | 2.56 | 1.01 | 4.02 | 1.59 | 1 vs. 2* |

**p<.01; *p<.05; 1, 2 ,3 = Cluster 1, 2, 3, respectively.

In conclusion, the three patterns of viewing behavior as found with the clustering algorithm can be described as follows (cf. Figure 4).

Learners in *cluster 1* exhibit a viewing behavior that is at an intermediate level between the other clusters. This behavior is characterized by long fixation times and high fixation counts on text and pictures as well as many transitions between text and pictures. In contrast, *cluster 2* comprises learners who fixate text and pictures shorter and less often and also do not switch as often between representations. Learners in *cluster 3* show the viewing behavior with the most and longest fixations both on text and pictures as well as the most transitions.

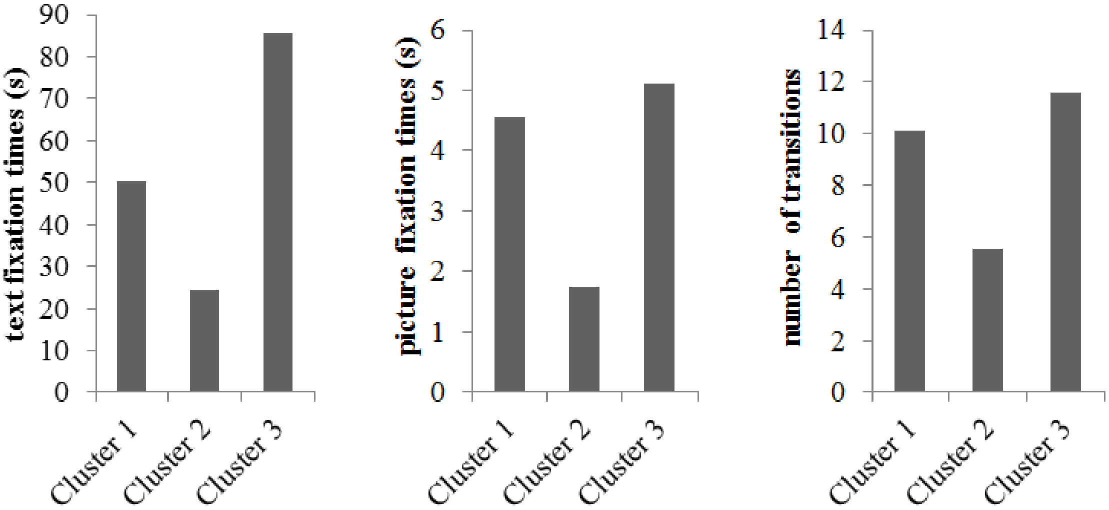


*Figure 4*. Eye tracking parameters as a function of cluster in Experiment 1.

### *2.3.3.2 Eye movements and learning outcome*

To investigate the relation of viewing behavior and learning success, cluster allocation was used as factor in a MANOVA with recall and transfer performance as dependent variables. For transfer, one participant was excluded from analysis due to missing data. Learners in the three clusters did not differ in recall performance, $F(2, 26) = 2.48$, $p = .10$, $\eta_p^2 = .01$. They differed significantly in transfer performance $F(2, 25) = 3.40$, $p = .049$, $\eta_p^2 = .15$. Bonferroni-adjusted posttests revealed that students in cluster 1 had significantly better learning outcomes than students in cluster 2 (*p* =.046). Although clusters 1 and 3 differ greatly on a descriptive level, this difference is not statistically significant, probably due to the small size of cluster 3. Table 5 reports means and standard deviations of learning outcome in the three clusters.

Table 5: *Means and standard deviations of learning outcome measures as a function of cluster in Experiment 1*

| | **Cluster 1 (*n* =11)** | | **Cluster 2 (*n* =13)** | | **Cluster 3 (*n* =5)** | |
|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Recall % | 59.89 | 13.26 | 47.14 | 17.89 | 43.80 | 18.10 |
| Transfer % | 68.69 | 16.34 | 50.43 | 21.09 | 48.89 | 16.85 |

### 2.3.4 Discussion of Experiment 1

Against the backdrop of designing an adaptive multimedia system based on learners' gaze behavior, the purpose of Experiment 1 was to investigate patterns of eye movements that can identify successful vs. unsuccessful learners in a multimedia environment (Research Question 1). Furthermore, I wanted to examine which eye tracking parameters need to be considered for the adaptive system (Research Question 2). Another research objective was the relative contribution of cognitive prerequisites to the distinction of learners (Research Question 3). Therefore, learners' cognitive prerequisites and their eye movements in a multimedia learning session on cell division were submitted to a cluster analysis, resulting in three groups of learners with different patterns of viewing behavior. However, the third cluster was too small to be interpreted in a meaningful way. Therefore, it was decided to refer to the research questions based on the first two clusters, which differed in their patterns of viewing behavior as well as in their learning outcome. More specifically, learners in cluster 1 fixated both text and pictures more often and longer than learners in cluster 2, and also showed more transitions between text and picture, and this viewing behavior resulted in better learning outcome in the transfer test. Therefore, with regard to the first research question, namely how to identify successful vs. unsuccessful learners, the findings show that successful learners show more and longer fixations on both text and picture, and they also exhibit more transitions between representations. This is in line with findings from previous eye tracking studies (Hegarty & Just, 1993; Johnson & Mayer, 2012; Mason et al., 2013; Scheiter & Eitel, 2015). Regarding the second research question, namely which eye tracking parameters to consider for the adaptive system, discriminant analysis showed that all gaze parameters that were analyzed contributed to the distinction of the clusters. However, fixation count and fixation time are interrelated as a learner who looks at text or picture more often will also have a longer total fixation time. With the purpose of designing an adaptive system that works with as few variables as possible, it was decided to implement transitions and fixation times but not fixation count.

According to the third research question, I wanted to investigate if cognitive prerequisites are a necessary parameter for the adaptive system. Interestingly, cognitive prerequisites made no significant contribution to the distinction of the clusters: cluster analysis yielded the same solution with gaze parameters alone as with cognitive prerequisites as additional variable. Furthermore, the clusters did not differ regarding their cognitive prerequisites, and discriminant analysis showed that cognitive prerequisites made no significant contribution to the distinction of the clusters. Consequently, learners' level of cognitive prerequisites was excluded from the definition of the adaptation algorithm.

In conclusion, this study showed that long fixations on text and pictures as well as a high number of transitions are associated with good learning outcome. Fixation time on text and picture might be an indicator for selection and organization processes, whilst transitions are associated with integration of the representations. Therefore, this study indicates that in line with CTML (Mayer, 2005), learners who actively process the materials benefit from multimedia instruction.

These findings were used to develop a gaze-based adaptive system. More specifically, based on the four-process adaptive cycle (Shute & Zapata-Rivera, 2012), the adaptive system *captures* text/picture fixation times as well as the number of transitions between both representations. It then *analyzes* if these parameters are above or below the threshold values and *selects* and *presents* content accordingly. The design and evaluation of the adaptive system are described in the following.

## 2.4 Experiment 2

Based on the findings from Experiment 1, an algorithm for an adaptive system was developed. The system analyses learners' eye movements online and alters the presentation of the multimedia instruction accordingly. The basic system itself was developed by computer scientists and can be adapted to different purposes (Schmidt, Wassermann, & Zimmermann, 2014; Wassermann, Hardt, & Zimmermann, 2012).

Since it was found that poor learners processed both text and pictures shorter and had fewer transitions between both representations than good learners, the system adapts to short fixation times and few transitions. When a learner exhibits inadequate visual processing on a page of the material, the system selects and presents the same content, but in an instructional design that, based on empirical evidence, should prompt adequate processing. In the case of short text/picture fixation times, this design is a zoom-out of the respective representation. Zoom-outs of the text should prompt the rereading of text segments (regressions), which are positively correlated to learning outcome (Rayner, 1998). Zoom-outs of the picture should lead to longer picture fixation times, which are positively related to learning outcome (Hegarty & Just, 1993; Scheiter & Eitel, 2015). In the case of a lack of transitions, this content is a version of the learning material that includes color coding of corresponding elements of text and picture since this form of cueing has been shown to enhance integration processes (Ozcelik et al., 2010).

This experiment investigated if the adaptive multimedia learning system has beneficial effects on learning outcome. For this purpose, a no intervention control group learned with the same, static material as in Experiment 1. The experimental group learned with the same material, but received adaptations of the instruction when necessary.

Receiving adaptations of the learning material based on one’s viewing behavior should compensate for poor visual processing and thus improve learning. Accordingly, I expected learners with inadequate visual processing in the experimental group to learn better than learners with inadequate visual processing in the control group, as indicated by higher posttest scores (Hypothesis 1).

Additionally, I was interested in the influence of cognitive prerequisites on the effect of the adaptive system. Although in Experiment 1 cognitive prerequisites did not differentiate between successful and less successful learners, they can influence the effect an intervention has on multimedia learning (Mason, Pluchino, & Tornatora, 2015b). Possibly, the effect of an adaptive learning environment interacts with a learners’ level of cognitive prerequisites. This interaction could work in two ways. On the one hand, especially learners with weak cognitive prerequisites might benefit from the adaptive system because it compensates for their lack of preexisting knowledge. On the other hand, the support offered by the adaptive system might need a certain level of cognitive prerequisites to build on. Thus, the influence of cognitive prerequisites on the effects of an adaptive system could be twofold; therefore, I addressed the potential moderating role of cognitive prerequisites as an exploratory research question (Research Question 4). Since the adaptive system was a novel learning environment, I also assessed ease of use to assess if it has good usability (Research Question 5).

### 2.4.1 Method

#### *2.4.1.1 Participants and design*

Participants were 79 students of the University of Tuebingen (54 female; *M* = 25.19 years, *SD* = 6.45) enrolled in different courses. All subjects had normal or corrected to normal vision and received a compensation of ten euros for their participation. Experimental condition was manipulated as a between-subjects factor: participants were randomly assigned to either the control (*n* = 40) or the experimental group (*n* = 39). Two different measures of learning outcome (recall and transfer) were assessed as dependent variables.

#### *2.4.1.2 Materials and apparatus*

*Materials.* The same learning material on cell division as in Experiment 1 was used. It was presented in two different versions: a static version identical to that in Experiment 1 (see Figure 3), and an adaptive version, which is described in Chapter 2.4.1.3.

*Apparatus.* Eye movements were collected using a SMI RED 250 eye tracker manufactured by SensoMotoric Instruments. This remote eye tracking system features a 22-inch TFT-Monitor and a sampling rate of 250 Hz. Participants were seated at a 65 cm distance to the screen.

#### *2.4.1.3 Description of the adaptive system*

The *Adaptive Learning Module* (ALM) used in this study was developed by computer scientists within an interdisciplinary project (Schmidt et al., 2014; Wassermann et al., 2012). The eye tracking application ALM connects the eye tracking hardware and the web based e-learning environment. It consists of several components. The driver of the SMI RED 250 uses a TCP-IP-interface to send raw data to downstream applications. A newly developed server application enters this interface and converts the data stream into a web socket connection. For details regarding the server application's architecture, please see Wassermann, Hardt and Zimmermann (2012). Based on *ILIAS 4 Open Source Framework*[1] a PlugIn was developed; it uses the infrastructure of an established ILIAS installation to present learning content. The ALM PlugIn builds on the course structure of ILIAS and extends them with additional components like the ALM learning unit. Beyond the ability to display learning content as interactive presentation within ILIAS, such a learning unit can receive, analyze and react to eye tracking data via the web socket interface.

The application's adaptivity is based on the capture of gaze fixations on Areas of Interest (AOI). When a user fixates an AOI for a predefined time, a fixation is recognized and logged. Each learning unit can react to individual situations by dynamically changing the learning content or the sequence of the presentation. For example, it is possible to count the fixation on an AOI and to trigger a modification of the learning content when a threshold value is reached. For detailed information on the adaptivity of ALM, see Schmidt, Wassermann, and Zimmermann, 2014.

Based on Experiment 1, threshold values for each page of the learning material were derived as follows. To identify inadequate processing of the instruction as indicated by short total fixation times and few transitions, the mean fixation time on text and picture and the mean number of transitions were computed for cluster 2, that is, for the non-successful learners from Experiment 1. One standard deviation was added and the resulting value was used as threshold. To obtain the most conservative threshold values, the numbers were rounded downwards. In those cases where

---

[1] http://www.ilias.de/docu/ilias.php?baseClass=ilrepositorygui&reloadpublic=1&cmd=frameset&ref_id=1

the resulting value was higher than the mean in cluster 1, that is, the successful learners from Experiment 1, the latter value was used. For example, on the first page of the material explaining the structure of an animal cell, the non-successful learners fixated the text for a mean duration of 6789.82 ms (*SD* = 3135.36), the picture for a mean duration of 2170.42 ms (*SD* =1205.74), and exhibited a mean of 5.92 (*SD* = 4.46) transitions. Adding one standard deviation resulted in the following threshold values: 9925 ms for text fixation time, 3376 ms for picture fixation time, and 10 transitions. These calculations were repeated for all pages of the learning material.

The resulting adaptive system worked as follows. When learners fixated both text and picture long enough, they could proceed to the next page by clicking on the forward button. When they did not reach the threshold value for either the text or picture fixation time and clicked the forward button, the text or picture zoomed out to prompt text or picture processing, respectively (see Figure 5, left panel). After ten seconds, a closing button appeared and the learners could close the adaptation. When learners' number of transitions between text and picture was too low, they were presented with a color-coded version of the learning content where corresponding text-picture elements were highlighted using the same color (see Figure 5, right panel). The system then analyzed the number of transitions on the color-coded material, and the adaptation could be closed after three transitions had been made.

When a learner was below the threshold value for two or all three parameters, s(he) received all corresponding adaptations in the following order: zoom-out of the text, zoom-out of the picture, and finally the color-coded version of the instruction.

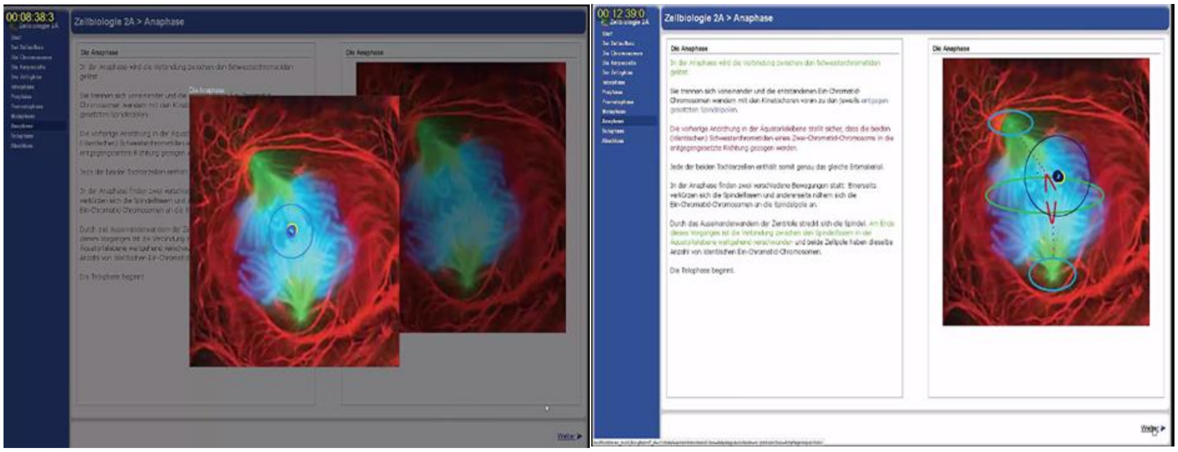


*Figure 5*. Exemplary pages of the adapted learning material in Experiment 2: Zooming-out of the picture (left) and presentation of the color-coded version (right).

### *2.4.1.4 Measures and scoring*

Measures encompassed cognitive prerequisites, posttest performance, and user satisfaction, which will be described in the following section.

*Cognitive prerequisites test.* The cognitive prerequisites questionnaire consisted of a measure for scientific literacy and a measure for domain-specific prior knowledge; it was identical to the one used in Experiment 1 (cf. Chapter 3.1.3) and was also scored accordingly. As in Experiment 1, one comprehensive measure of cognitive prerequisites was computed by adding up the z-standardized total scores of scientific literacy and domain-specific prior knowledge (Cronbach’s $\alpha = .68$).

*Posttest*. The paper-pencil based posttest assessed recall and transfer. It consisted of three subtests: a free recall question, a forced-choice verification task and 16 multiple-choice-items.

As in Experiment 1, the free recall question asked participants to write down everything they know about mitosis. Time limit for this open question was six minutes. Answers were scored using a coding scheme that awarded one point for each correctly mentioned concept (max. score 38).

Two independent raters scored all participants and discussed their ratings until total agreement was achieved.

The multiple-choice items were identical to Experiment 1. Again, seven items assessed recall and nine assessed transfer. The items were either text- or picture-based. There was no time limit. Each correctly answered item was scored one point, incorrect answers were scored zero.

In the forced-choice verification task, participants had to state if sentences or pictures were either true or false (e.g., ‘During mitosis, each daughter cell gets 46 one-chromatid chromosomes’). Two items assessed transfer, 20 items assessed recall. Each correct answer was awarded one point, incorrect answers were scored zero.

Mean scores from the free recall task, the 20 forced-choice verification recall items and from the seven multiple choice recall items were averaged and transformed into percentages to obtain a measure of recall test performance. The mean scores of the nine multiple-choice transfer items and of the two forced-choice verification transfer items were averaged and transformed into percentages to obtain a measure of transfer test performance.

*User satisfaction*. To assess the ease of use of both the static and the adaptive system, I administered the System Usability Scale (SUS) (Brooke, 1996) in German translation. This questionnaire consists of eleven items addressing different usability aspects (e.g., *“I thought the system was easy to use”*). Participants were asked to express their level of agreement for each statement on a scale of 0 (*“I do not agree at all*”) to 4 (*“I totally agree”*). Since five items were reversely coded (e.g., *“I found the system unnecessarily complex”*), they were reversed and, to obtain a general score of user satisfaction, the ratings on all items were summed up. The overall score had a range of zero to 44.

#### *2.4.1.5 Procedure*

Data collection took place in individual sessions. Participants were seated in front of the eye tracker and received written information on the proceedings of the experiment, signed a consent form, and completed the cognitive prerequisites test. Afterwards, the eye tracker was calibrated for each participant with a 9-point calibration. In both groups, participants were informed that they could learn each page at their own pace and proceed to the next page by pressing the space key, but not go back in the material. Participants were also informed that they would be tested on the content. Additionally, participants in the experimental group were informed that upon clicking the forward button, elements from the actual learning page would be re-presented if they had not been sufficiently processed. After the learning phase, participants completed the paper-pencil posttest, were debriefed and received compensation.

#### *2.4.1.6 Data analysis*

To investigate the effects of the adaptive system and the influence of cognitive prerequisites on learning outcomes, I used regression analyses with effect coding to analyze the data. Thereby, the experimental condition is coded in a way that reflects the hypothesis regarding its effect. In this case, I expected the experimental group to outperform the control group. Consequently, the experimental group was coded +1, and the control group was coded -1. Experimental condition, the z-standardized cognitive prerequisites score, and the interaction term between cognitive prerequisites and condition were entered simultaneously as predictors. User satisfaction in the two versions of the system was compared using one-way ANOVA.

### 2.4.2 Results

For all statistical analyses reported here, the level of significance was set at $\alpha = .05$. Statistical analyses were conducted with IBM SPSS Statistics 20.

Due to large amounts of missing data, five participants were excluded from statistical analysis. One participant was excluded due to a total score of zero in the cognitive prerequisites test. Another three participants were excluded due to extreme values (i.e., 2 standard deviations above or below the mean) in text fixation times, picture fixation times or the number of transitions.

To obtain a fair comparison, those participants whose visual processing was inadequate in both groups were selected. That means, those participants in the experimental and the control group whose number of transitions was below the threshold value, which indicates inadequate visual processing, were compared. For this purpose, areas of interest around text and pictures were defined on each page of the learning material and computed the number of transitions. I chose to use this measure rather than text or picture fixation times because eye tracking research has shown that integration processes are very important for multimedia learning (e.g., Hegarty & Just, 1993; Mason et al., 2013). For each participant, the number of transitions was compared to the threshold value implemented into the adaptive system. Only participants whose number of transitions was below the threshold value on at least 2 pages were included. This resulted in 64 participants (45 female; *M* = 23.64 years, *SD* = 2.98), that is, six participants were excluded from analyses.

Table 6 shows means and standard deviations for the cognitive prerequisites measures, the learning outcome measures and user satisfaction.

Table 6: *Means and standard deviations for cognitive prerequisites, learning outcome measures and user satisfaction as a function of experimental condition in Experiment 2*

| | **Experimental ($n$ = 33)** | | **Control ($n$ = 31)** | |
|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Cognitive prerequisites (z-value) | -.14 | .96 | .15 | 1.04 |
| Recall % | 44.28 | 9.41 | 45.80 | 8.69 |
| Transfer % | 48.57 | 13.20 | 49.37 | 13.56 |
| User satisfaction | 33.59 | 5.74 | 35.39 | 5.67 |

#### *2.4.2.1 Learning outcomes*

For recall, the overall regression model was marginally significant, $F(3,60) = 2.27$, $MSE = 8.77$, $p = .09$, *adjusted* $R^2 = .06$. Cognitive prerequisites significantly predicted learning outcome in that learners with stronger cognitive prerequisites achieved higher recall scores than learners with weaker prerequisites.

For transfer, the overall regression model was not significant, $F < 1$. Table 7 shows results of the regression analyses for the learning outcome measures.

Table 7: *Results of the regression analyses for the learning outcome measures in Experiment 2*

| | **Recall** | | | |
|---|---|---|---|---|
| | **Beta** | ***SE*** | ***β*** | ***p*** |
| Constant | 44.99 | 1.11 | | <.001 |
| Condition | -.36 | 1.11 | -.04 | .75 |
| Cognitive prerequisites | 2.78 | 1.12 | .31 | .02 |
| Condition x cognitive prerequisites | -.27 | 1.12 | -.03 | .81 |
| | **Transfer** | | | |
| | **Beta** | ***SE*** | ***β*** | ***p*** |
| Constant | 49.10 | 1.70 | | <.001 |
| Condition | -.12 | 1.70 | -.01 | .95 |
| Cognitive prerequisites | 2.02 | 1.71 | .15 | .24 |
| Condition x cognitive prerequisites | 0.94 | 1.71 | .07 | .59 |

#### *2.4.2.2 User satisfaction*

Means and standard deviation of user satisfaction are displayed in Table 6. Mean user satisfaction was high in both systems. There was no significant difference on user satisfaction between the static and the adaptive version of the system, $F(1,61) = 1.56$, $MSE = 32.54$, $p = .22$, $\eta_p^2 = .03$.

### 2.4.3 Discussion of Experiment 2

In Experiment 2, a control group learning with a static multimedia learning environment was compared to a group learning with a gaze-based adaptive learning environment. I expected the experimental group to learn better than the control group, as indicated by higher posttest scores (Hypothesis 1). Furthermore, I was interested in the influence of cognitive prerequisites on the effect of the adaptive system (Research Question 4). Research Question 5 addressed the usability of the adaptive system.

Contrary to the hypothesis, there were no group differences on learning outcome, indicating that in this experiment, the gaze-based adaptive system did not support multimedia learning. Possible reasons for this finding will be discussed in the next section.

For recall, cognitive prerequisites predicted learning outcome in that stronger prerequisites led to higher learning outcome. Regarding Research Question 4, there were no interaction effects between experimental group and cognitive prerequisites, indicating that preexisting knowledge did not influence the effect of the adaptive system. The usability of the adaptive system was comparable to that of the static system; participants had no difficulties regarding the ease of the adaptive system's use.

## 2.5 Discussion: external support for multimedia learning

With the aim of developing and evaluating a gaze-based adaptive system, two experiments were conducted. In the first experiment, gaze behavior and learning outcome in a multimedia learning environment were assessed. Furthermore, the role of cognitive prerequisites, i.e., domain-specific prior knowledge and scientific literacy for the distinction of good and poor learners was investigated. Cluster analysis revealed two meaningful patterns of viewing behavior, while a third cluster could not be interpreted due to a small amount of cases. The two main patterns of viewing behavior also differed in learning success. Compared to the successful learners in cluster 1, the less successful learners in cluster 2 showed shorter text and picture fixation times as well as fewer transitions. Interestingly, cognitive prerequisites did not differ between the clusters. Based on these results, an adaptive system was designed; it analyzes a learner's text/picture fixation times and number of transitions online and alters the instruction accordingly. In the second experiment, this system was compared to a static, non-adaptive version of the same learning environment regarding learning outcomes. Furthermore, the role of cognitive prerequisites on the effects of the adaptive system was investigated. There was no main effect of the adaptive system and no interaction with cognitive prerequisites, indicating that learners did not benefit from the adaptive system, irrespective of their pre-existing knowledge on the domain.

The system was based on the results of Experiment 1, where successful learners were characterized by longer text/picture fixation times and more transitions than less successful learners. Consequently, the system reacted when a learner showed values below certain thresholds. These thresholds were derived by adding one standard deviation to the mean fixation times and number of transitions of the cluster with inadequate visual processing. This decision was not based on empiric evidence because to my knowledge, this is the first adaptive multimedia learning system based solely on gaze behavior. Using the cluster mean would result in fewer learners receiving adaptations. It is possible that the current algorithm included learners whose visual processing would have resulted in good learning outcome. For these learners, the adaptations may have resulted in poorer learning outcome because they were disturbed in their (adequate) learning process. As research on the expertise reversal effect (Kalyuga et al., 2003) and overscripting (Fischer et al., 2013) shows, successful learners do not necessarily benefit from additional support, or might even be impaired by it.

Possibly, only some of the learners benefitted from the adaptations while others did not or were even inhibited, which would explain why there were no significant group differences.

Furthermore, it is possible that the algorithm does not identify all learners who need support as not only short, but also very long fixation times and many transitions might indicate inadequate visual processing (Rayner, 2009; Schwonke et al., 2009). In Experiment 1, there was a third cluster with even longer fixations and more transitions that was not interpreted because it contained only five cases. Learners in this cluster had poorer learning outcome than learners in the other two clusters. However, this difference was found only on the descriptive level, probably due to the small cluster size. Possibly, the algorithm of the adaptive system should react to values below and above certain threshold values.

It is also possible that learners in the experimental group did not learn better because they were confused by the adaptation. The instruction for the experimental group provided very little information on how the adaptive system works in order to investigate the system`s effect under realistic circumstances and to avoid reactive behavior. A more detailed instruction or even a short training period with the adaptive system may lead to different results and is therefore an interesting direction for future studies. However, the high scores on user satisfaction indicate that participants received the adaptive system overall as easy to use.

In addition, the algorithm is based on the assumption that learners in Experiment 2 are comparable to those in Experiment 1 regarding their viewing behavior. There may be more than 2 patterns of viewing behavior that did not come up in Experiment 1 due to the small sample size. A replication of Experiment 1 with more cases might provide a more fine-grained algorithm.

In the reported experiment, the influence of cognitive prerequisites on both viewing behavior in a multimedia learning session (Experiment 1) and the effect of the adaptive system (Experiment 2) were investigated. More specifically, I examined the role of domain-specific prior knowledge and scientific literacy. However, there are additional learner characteristics that are related to multimedia learning. For example, working memory is important for multimedia learning (Schüler, Scheiter, & van Genuchten, 2011), as is reading comprehension (Mason et al., 2013; Scheiter, Schubert, et al., 2014) and spatial ability (Hegarty & Kriz, 2008). The aim of this dissertation was to develop an adaptive system that collects as little data on the learner as possible in non-intrusive ways while still providing tailored support. Furthermore, the system adapts to online learning behavior rather than previously assessed learner characteristics which are then assumed to remain constant over time. However, the results presented here suggest that it may be necessary to add at least some additional variables to the algorithm.

# 3 Self-regulation support for multimedia learning: Eye Movement Modeling Examples

In the previous chapter, I presented an intervention that provided external support for multimedia learning. As discussed in the introduction, a different approach is to support self-regulation of adequate processing. The general idea is to provide learners with strategies, which they can then apply to the instruction in a self-regulated way. Other than with external support, learners should benefit from self-regulation support beyond the learning context where the support is provided. A learner who is supported in the self-regulated processing of a specific instruction might apply these skills to other instructions as well. In this dissertation, I used Eye Movement Modeling Examples to illustrate adequate visual processing of multimedia instruction.

This chapter provides an overview on research on supporting self-regulated adequate processing (Chapter 3.1) and an introduction to Eye Movement Modeling Examples (Chapter 3.2).

## 3.1 Research on supporting self-regulated adequate processing

Self-regulated learning can be defined as systematically orienting ones thoughts, feelings and actions towards the learning goal (Zimmerman & Schunk, 1989). Boekaerts (1999) describes self-regulated learning as a three-layered concept (cf. Figure 6). In her model, self-regulated learning encompasses three types of regulation: regulation of processing modes, of the learning process, and of the self. Regulation of processing modes is represented in the model’s inner layer and encompasses the selection of cognitive strategies. Regulation of the learning process is represented in the middle layer; it refers to the use of metacognitive knowledge and skill to direct learning. Regulation of the self, that is, motivational control, comprises the outer layer of Boekaerts’ model.

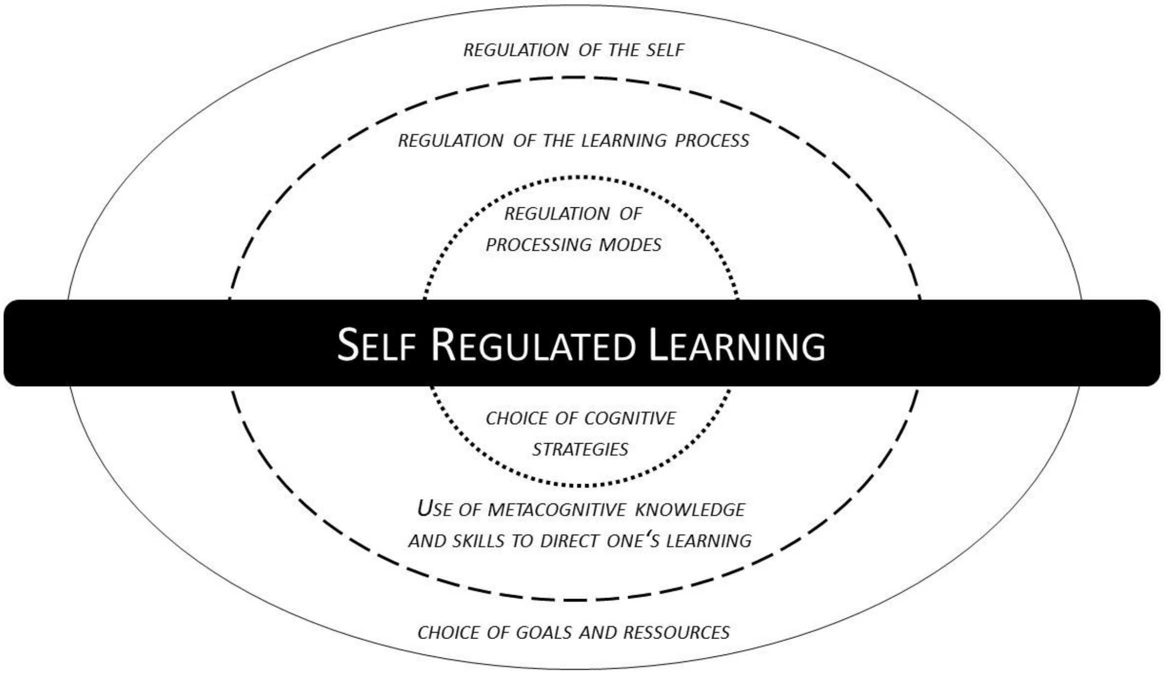


*Figure 6.* Boerkaerts' three-layered model of self-regulated learning (adapted from Boekaerts, 1999).

The inner layer, or the student's ability to select and apply cognitive strategies, is essential for self-regulated learning (Boekaerts, 1999). Consequently, helping students to choose and use adequate processing strategies should benefit learning.

There is some research on such strategy interventions in multimedia learning. Kombartzky et al. (2010) provided school students in the 6th grade with a worksheet explaining several effective cognitive processes for learning from an animation (selection, organization, and integration/transformation processes). Learners received either the strategy worksheet or instructions for the preparation of an essay about the animation's content. Children in the strategy worksheet group achieved better learning than children in the essay group. Similar results were found by Schlag and Ploetzner (2010) who implemented an analogous strategy support for text with static visualizations. Again, their participants were 6th graders and the learning topic was the dance of the honeybee. The strategy worksheet improved learning as compared to an essay control group.

While there is evidence for beneficial effects of strategy interventions for children, research with teenagers and adults paints a different picture. Schlag, Florax, and Ploetzner (2007) administered a strategy intervention similar to the one used by Schlag and Plötzner (2010) to students in their mid-twenties. While the intervention was effective for 6th graders, think aloud protocols showed that the older learners did not pick up the new strategy but retained their usual learning habits.

Scheiter, Schubert, Gerjets, and Stalbovs (2014) developed a strategy training for 10th graders where several processing strategies were conveyed. Strategies included selection, organization and integration processes. A model explained and applied the processes using exemplary material; learners then applied them to a new set of materials while they were supported by the model when necessary. Finally, learners received a third set of materials and were instructed to learn from them. Results indicate that the learners had better knowledge of adequate processing, which was assessed by a questionnaire. However, they did not benefit in terms of learning outcome: a control group who received a training on general learning techniques achieved the same learning outcome as the experimental group.

The strategy interventions discussed above presuppose that learners will apply the conveyed strategies. Most learners, however, have their own (maladaptive) strategies, which they apply in an automated way. Strategy interventions like the one used by Scheiter et al. (2014) are rather abstract. Mapping the strategies to processing behavior might be too challenging, and learners might therefore rely on existing strategies. This may especially be the case with adults, while children - who do not yet have preexisting strategies – use the ones provided by the intervention.

The dominance of preexisting, possibly inadequate strategies might be overcome by presenting the new strategies in a way that is closer to the actual learning behavior. By doing so, the strategies may be easier to apply, making it more likely that learners give up on their habitual strategies and try the new ones. Eye Movement Modeling, which is described in the next section, conveys the strategies on the perceptual level and therefore as close to the learning process as possible.

## 3.2 Eye Movement Modeling Examples

EMME consist of a video of a skilled person's eye movements which are recorded while (s)he performs a task and which are superimposed onto the material (Jarodzka et al., 2012; Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; Mason et al., 2015, in press; van Gog et al., 2009). The basic idea is that displaying an expert's eye movements offers perceptual guidance on how to effectively process the stimuli. It is assumed that learners internalize how to process the information, thereby acquiring skills that can later be applied to new stimuli. A benefit of EMME compared with verbal strategy instruction is that learners can look at the strategies that they are supposed to acquire. Although these observations are related only to the cognitive strategies' manifestation at the visual processing level, it is assumed that the perceptual input will also trigger the cognitive processes underlying these manifestations.

EMME have been successfully applied in several domains. Jarodzka, van Gog, Dorr, Scheiter, and Gerjets (2013) used an expert's eye movements to model the classification of fish locomotion patterns. The modeling examples were shown during a training phase, whereas test performance was assessed with novel stimuli. The control group received the same videos with verbal explanations and without eye movements superimposed onto the stimulus. Results show that EMME during training increased performance. In a study by Jarodzka et al. (2012), EMME were used to train medical diagnosis making, more specifically, the occurrence of seizures in infants. Like in the study by Jarodzka et al. (2013), the actual task was performed on novel stimuli, and the control group received only the video without eye movements but with verbal explanations. The study revealed that EMME supported performance on the diagnostic task.

First evidence that EMME can also support multimedia learning came from a study by Mason, Pluchino and Tornatora (2015) who compared an EMME group with a control group regarding visual processing and learning outcomes. The EMME showed a skilled learner's gaze behavior while studying a single page with illustrated text; the model first gained an overview of the whole text by scanning it and then put verbal and pictorial information in relation to each other by shifting the attention from one representation to the other. No verbal explanations of the strategies were provided. After watching the EMME, the learners were given a second, one-page illustrated text unrelated to the one used for generating the EMME, and their eye movements were recorded while they studied it. Then they were tested regarding their recall and comprehension of this material. The control group received only the second part of the material and was tested afterwards. Results revealed that learners in the EMME group showed more integrative rereading of the material and also had better learning outcomes, thus yielding evidence for the effectiveness of EMME as strategy support for multimedia learning. These findings were replicated in a second study; moreover, this study revealed that especially students with poorer reading comprehension skills benefitted from EMME, whereas there was no EMME effect for students with stronger comprehension skills (Mason et al., 2015b). These results indicate that EMME are beneficial to multimedia learning. However, both studies were conducted with children, and the modeling examples aimed only at integration processes. It is unclear if EMME are of assistance to adult learners, who may already possess, albeit maladaptive, strategies of processing text and picture.

Moreover, it is still an open question if promoting more than one cognitive process is helpful. The latter requires more complex EMME illustrating multiple processes. Learners may find it difficult to extract the information that is relevant to skill development from such complex EMME. However, evidence that providing learners with several processes is beneficial comes from a study by Stalbovs et al. (2015). In a multimedia learning session, instructional support featuring a combination of selection, organization and integration processes was superior to an intervention with only one type of process. In the two experiments presented in the following, I investigated the effects of EMME that aimed at promoting several cognitive processes that are related to successful multimedia learning. Based on cognitive theories of multimedia learning, the processes of selection, organization and integration were conveyed. In addition, the EMME illustrated several processes successful learners were shown to execute in previous research. For example, there is evidence that an initial viewing of the picture prior to reading text is helpful (*pictorial scaffold*). It provides a coarse representation of the picture which can subsequently serve as a scaffold for text and picture processing, thus supporting mental model construction (Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; Eitel, Scheiter, & Schüler, 2013). Likewise, Hegarty and Just (1993) found that successful learners took a global last look at the picture *(final picture inspection)*, presumably to check their mental representation for misconceptions. When realizing that there are misconceptions, it is a useful strategy to revisit the problematic concept by studying the respective text and picture segments (*reaction to comprehension problems*).

Beyond deciding which processes the EMME illustrate, there are several design options for the EMME.

First, the eye movements can be visualized by adding visual information, for example by a circle display that moves across the page. They can also be visualized by reducing visual information by blurring or greying out the areas where the model's gaze is not directed (spotlight display). On the one hand, there is evidence that for perceptual tasks, a spotlight display was better suited to improve performance than a circle display (Jarodzka et al., 2012). On the other hand, spotlight displays were shown to enhance selection of information while circle displays support organization and integration processes (Jarodzka et al., 2013).

Second, the EMME can contain a didactic audio commentary (Jarodzka et al., 2012; Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; van Gog et al., 2009) or provide just the visual information (Mason et al., 2015a, 2015b).

Third, the EMME can be presented on exemplary material, or on the actual learning material. While presenting EMME on exemplary material allows for assessing the transfer of adequate processing to novel learning contexts, displaying EMME on the learning material might make it easier to apply the processing strategies.

In this dissertation, two different EMME interventions were tested for their effectiveness. The first intervention is reported in Chapter 3.3; it featured a circle display and an audio commentary and was presented on exemplary material. The second intervention is reported in Chapter 3.4; it featured a spotlight display on the first four pages of the actual learning material and no audio commentary.

## 3.3 Experiment 3

In this experiment, I investigated the effects of EMME on students' multimedia learning. For this purpose, in the experimental group, learners received a video-based pre-training containing a model's eye movements and verbal explanations (*EMME group*). In this experiment, the EMME were designed similarly to previous studies on EMME in multimedia learning (Mason et al., 2015a, 2015b). That means, a circle display was used to illustrate adequate visual processing, and the

EMME were displayed on exemplary material. However, contrary to Mason et al., whose EMME illustrated only integration processes, several cognitive processes were conveyed. To make it easier for learners to distinguish the processes, an explanatory audio commentary was added to the EMME.

Although there is first evidence that EMME can support multimedia learning, an open question is if EMME are more beneficial than an intervention that conveys the same processes but illustrates them differently, for example by highlighting or color-coding. EMME directly reflect a successful learner's visual processing of the instruction and are very close to a learner's behavior. They might have specific beneficial effects because learners can easily transfer the adequate processing strategies to their own learning. Therefore, I developed a second, analogous video-based intervention displaying signals instead of eye movements (*Cueing group*).

In addition, a third group received no intervention (*Control group*). Subsequently, all students learned with text and static visualizations. Importantly, the materials for interventions were different from the actual learning materials for which subjects' learning outcomes were assessed.

I expected the EMME and the Cueing group to be more effective in processing the multimedia instruction, resulting in higher learning outcomes in the intervention groups (Hypothesis 2). Research Question 6 addressed differential effects of the two interventions.

### 3.3.1 Method

#### *3.3.1.1 Participants and design*

Participants were 80 students of a German university (53 female; M = 22.98 years, SD = 3.73) enrolled in different courses. Students of physics and related fields were excluded; participants received 8 euros for their participation. Experimental condition was manipulated as a between-subjects factor: participants were randomly assigned to either the EMME group (*n* = 26), the

Cueing group ($n$ = 28), or the Control group ($n$ = 26). Three different measures of learning outcome (recall, transfer, and drawing) were assessed as dependent variables. Reading speed, reading comprehension, prior domain knowledge and learning time were measured as control variables.

#### *3.3.1.2 Materials and apparatus*

*Video based interventions.* The Eye Movement Modeling and the Cueing interventions were digital videos (.mp4 format) with a size of 720 x 576 pixels.

*Eye Movement Modeling Examples.* This video showed the eye movements of a successful learner moving along as he processed an illustrated text on the circulatory system. The model was a post-doc from our lab familiar with effective cognitive processes in multimedia learning. The model was instructed to process the materials in line with selection, organization and integration processes. Fixations were visualized as red circles with larger circles indicating longer fixations (scanpath). I added an explanatory audio commentary recorded by myself. In this commentary, the processing strategies and their relevance were explained as the model performed them.

To illustrate the EMME intervention, the models' eye movements on the material and the corresponding processing strategies are described in the following section.

Upon entering the page, the model globally inspected the picture (*pictorial scaffold*) and then read the title. Afterwards, he read the whole text, fixating relevant words and looked at the picture thoroughly to get an overview of the representations' contents (*selection of relevant words and picture elements*). He then read the text again section by section (*text organization*), and looked at the corresponding picture elements (*picture organization*), switching between the representations (*integration*). In the end, he thoroughly looked at the picture again (*final picture inspection*) while using it as a scaffold for checking his mental representation of the pages learning content and reread the corresponding sections in the text where necessary (*reaction to comprehension problems*). Total duration of the video was 12:48 min.

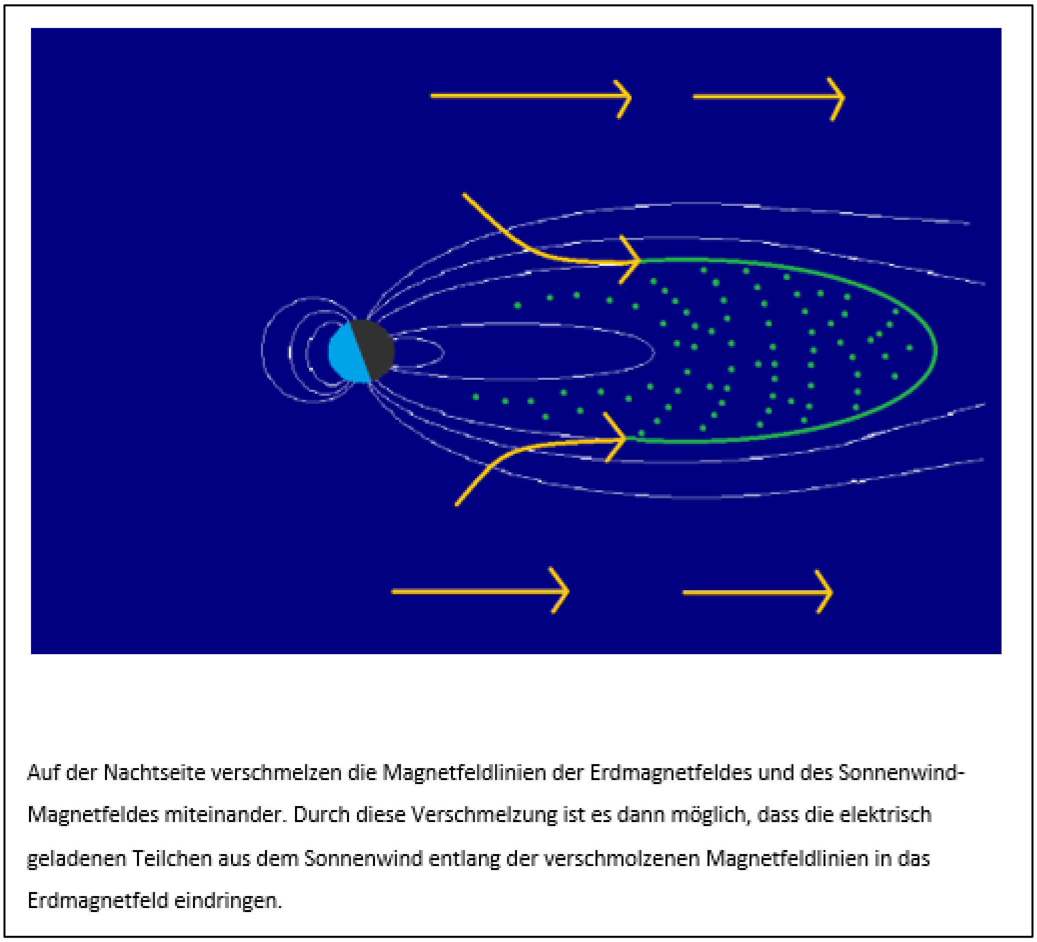
Table 8 provides an overview on the cognitive processes and the corresponding eye movements employed in the Eye Movement Modeling.

*Cueing.* In this video, the eye movements were exchanged for signals such as highlights and zoom-outs (cf. Table 8). Apart from that, the video was identical to the EMME video, with the same audio commentary and the same duration.

Table 8: *Description of the Cognitive Strategies and the corresponding eye movements / signals in Experiment 3*

| **Cognitive strategy** | **Eye movements / Cueing** |
|---|---|
| Pictorial scaffold | Scanning the picture upon entering a page / Zoom-out of picture |
| Selection of relevant words | Reading the text sentence by sentence with longer fixations on relevant words / relevant words in red font |
| Selection of relevant pictures | Taking a close look at the picture with long fixations on relevant components / red circles around relevant components |
| Organization of relevant words | Fixations on each semantic unit (reading) / zoom-out of semantic units and summarizing in own words |
| Organization of relevant pictures | Long fixations on / red circles around relevant picture elements and naming them |
| Integration of verbal and pictorial information | Transitions between / color coding of corresponding text and picture elements |
| Final inspection of the picture | Scanning / Zoom-out of the picture |
| Reaction to comprehension problems | Fixations / highlights on the relevant picture element if a text passage is not understood and vice versa |

*Learning materials*. Learning materials consisted of a written and illustrated German text on the subject of polar lights. It was printed on DIN-A4 paper. The learning content was divided into topics and distributed across ten pages. Each page showed text on the bottom half and one corresponding static picture on the top half. The text had an overall length of 1,047 words; the text passages were divided in semantically meaningful units and presented in paragraphs. First, the text explained what polar lights are and where they occur. Afterwards, it described prerequisites for the formation of polar lights, namely the existence of solar winds and the terrestrial magnetic field as well as the magnetic field's specific characteristics. It then explained the processes that lead to the formation of polar lights, where polar lights can be observed and why they can have a variety of colors. The illustrations were colored, schematic pictures that depicted the phenomena and processes described in the text and illustrated the respective visuo-spatial aspects. Thus, text and pictures were complementary and were both necessary to understand the learning content. Figure 7 shows an exemplary page of the learning material.



Auf der Nachtseite verschmelzen die Magnetfeldlinien der Erdmagnetfeldes und des Sonnenwind-Magnetfeldes miteinander. Durch diese Verschmelzung ist es dann möglich, dass die elektrisch geladenen Teilchen aus dem Sonnenwind entlang der verschmolzenen Magnetfeldlinien in das Erdmagnetfeld eindringen.

*Figure 7*. Exemplary page of the learning material in Experiment 3.

*Apparatus.* The model's eye movements were recorded using an SMI remote eye tracker with a sampling rate of 250 Hz and iView X$^{TM}$ 2.2 and Experiment Center$^{TM}$ 3.1 software. The eye tracking data were edited with BeGaze$^{TM}$ 3.1 (www.smivision.com). The video interventions were presented on 19" notebooks using Windows Media Player. Sound was played via headphones.

#### *3.3.1.3 Measures and scoring*

Since reading comprehension can influence learning success in a multimedia learning session (Scheiter et al., 2014), and to control for learning time and pre-existing knowledge on polar lights, these three variables were assessed as control variables.

Learning outcome was measured with four subtests. All tests were paper-pencil questionnaires. Open questions were scored by an experienced rater who had coded the same questions before in a different study, using the same coding scheme. All inter-rater reliabilities reported in the following were obtained with said rater in the previous study.

*Prior knowledge*. Domain specific prior knowledge was measured with three multiple-choice questions with four alternatives each (e.g., "What is solar wind?" (a) a stream of electrically charged particles which the sun attracts from space, (b) a stream of electrically charged particles that flows around the sun, (c) a stream of electrically charged particles which gets flowing by being warmed by the sun, (d) a stream of electrically charged particles which flows from the sun towards space). Each correct answer was scored one point, omissions and incorrect responses yielded zero points, resulting in a max. total sum of 3 points.

*Reading comprehension*. Reading comprehension was assessed by the LGVT 6-12 (Schneider, Schlagmüller & Ennemoser, 2007). In this test, students are asked to read a text with 1,727 German words for four minutes and to select (i.e., underline) the word that fits best into the text context amongst three alternatives in 23 sentences. Correct responses are scored two points, omissions zero points, and incorrect responses minus one point. Reading comprehension is defined by the number of correctly selected words (max. 46). The retest reliability (after six weeks) for reading comprehension is $r = .87$.

*Learning outcome*. The posttest consisted of a free recall task, 12 multiple-choice questions, four open transfer questions, and four drawing tasks.

The free recall task asked participants to write down everything they knew about polar lights. Each correctly mentioned concept was assigned one point, resulting in a maximum total score of 124. Inter-rater reliability was Cohen's $\kappa = .80$.

The 12 multiple choice items assessed knowledge on the concepts and processes relevant to the formation of polar lights (e.g., "What speed does the slow solar wind reach?" (a) 200 kilometers per second, (b) 300 kilometers per second, (c) 400 kilometers per second, (d) 500 kilometers per second). Eleven items assessed recall, one item assessed transfer. Correctly selected options were awarded one point, omissions or wrong selections zero points. Thus, the maximum total score was 12. Internal consistency of the scale was Cronbach's $\alpha = .79$.

The four transfer questions asked the participants to explain different phenomena related to polar lights (e.g., "What would be the consequence of solar winds occurred sporadically instead of regularly?"). Correctly mentioned concepts were awarded one point. The resulting maximum total score was 13. Inter-rater reliability was Cohen's $\kappa = .81$.

The four drawing items assessed recall of the learning content. Participants had to sketch different aspects of the learning content (e.g., "Please sketch how electronic particles move along magnetic field lines"). The content, but not the quality of the sketches was evaluated: correctly sketched parts were awarded one point. Participants could achieve a maximum total score of 35. Inter-rater reliability was Cohen's $\kappa = .72$.

Mean scores from the free recall task and from the eleven multiple choice recall items were averaged and transformed into percentages to obtain a measure of recall test performance. The mean scores of the multiple-choice transfer item and of the four open transfer items were averaged and transformed into percentages to obtain a measure of transfer test performance. Likewise, mean performance on the four drawing items was computed and transformed into percentages.

*Learning time.* The examiner wrote down the start and end times of the learning phase for each participant and learning time in minutes was computed.

#### *3.3.1.4 Procedure*

The experiment took place in groups of five to ten participants. Participants were randomly assigned to the three conditions. In one third of the group sessions, all participants were allocated to the Control group, whereas in two third of the sessions, participants received either the EMME or the Cueing intervention.

Participants were seated at individual desks separated by portable walls. They received written information about the experiment and signed a consent form. Afterwards, all participants completed the reading comprehension test, followed by the prior knowledge questionnaire; both tests were paper-based. In the experimental groups, each participant received the respective written instruction followed by the instructional video, which was started by the experimenter. The instruction in the EMME group included an introduction to EMME, including a screenshot of the material with the models' scanpath to familiarize them with the format. Participants in the Cueing group were instructed that they would see a video explaining several useful strategies for learning from text and pictures.

The videos were presented on individual 19" notebooks with headphones. Participants could adjust the volume, but could not stop or rewind the videos.

All participants were then handed out the paper-based learning material and instructed to learn for as long as they want but to not go back in the material. Upon completing the learning phase, participants were handed out the paper-based posttest, which they completed at their own pace. Afterwards, they received payment, were debriefed and dismissed. Each session lasted between 50 and 60 minutes.

### 3.3.2 Results

For all statistical analyses reported here, the level of significance was set at $\alpha = .05$. Statistical analyses were conducted with IBM SPSS Statistics 20.

#### *3.3.2.1 Control variables*

To investigate if the groups differed with respect to the control variables, I conducted a MANOVA with experimental condition as factor and reading speed, reading comprehension, prior knowledge and learning time as dependent variables. Table 9 shows means and standard deviations of the control variables.

There were no group differences in reading speed ($F < 1$), reading comprehension ($F < 1$) and prior knowledge ($F < 1$). The groups differed significantly in learning time, $F(2,77) = 8.95$, $p < .001$. Bonferroni-adjusted post-hoc analysis revealed that the Cueing group learned significantly longer than the Control group, $p < .001$. The difference between the Cueing and the EMME group was only marginally significant, $p = .06$. The EMME and the Control group did not differ significantly ($p = .21$).

I then investigated the correlation between learning time and our dependent variables. For recall, there was no correlation ($r = .09$, p = .44). For the transfer task, correlation was also not significant with $r = .01$, p = .90. Likewise, learning time and the drawing task did not correlate significantly ($r = .07$, p = .53). As a consequence, I did not include learning time as covariate for further analyses.

#### *3.3.2.2 Learning outcome*

I conducted a MANOVA with experimental condition as factor and the three learning outcome variables as dependent measures. Table 9 shows means and standard deviations for the dependent variables.

Table 9*: Means and standard deviations of the control variables and the three scales of the posttest as a function of experimental condition in Experiment 3*

| | **EMME (*n* = 26)** | | **Cueing (*n* = 28)** | | **Control (*n* = 26)** | |
|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Reading speed | 849.81 | 210.99 | 890.68 | 249.92 | 858.81 | 197.42 |
| Reading comprehension | 18.65 | 6.09 | 19.54 | 7.45 | 19.35 | 4.80 |
| Prior knowledge | 1.54 | .86 | 1.57 | .79 | 1.31 | .97 |
| Learning time (min) | 20.42 | 7.87 | 26.29 | 12.74 | 15.77 | 4.54 |
| Recall (%) | 51.23 | 12.89 | 50.78 | 11.54 | 52.52 | 10.61 |
| Transfer (%) | 37.09 | 20.41 | 31.63 | 15.24 | 42.58 | 18.24 |
| Drawing (%) | 61.87 | 25.83 | 64.39 | 17.79 | 61.97 | 21.28 |

There was a statistically significant difference in learning outcome based on experimental condition, $F(6,150) = 2.17$; $p = .049$; Wilk's $\Lambda = .85$, partial $\eta^2 = .08$.

However, there were no significant group differences for recall ($F < 1$), transfer ($F(2,77) = 2.49$; $p = .09$), and the drawing task ($F < 1$).

### 3.3.3 Discussion of Experiment 3

In this experiment, I investigated the effect of two interventions that aimed at supporting effective processing during multimedia learning. Both the Cueing and the EMME group watched a short video that showed nine processes relevant to effective learning from text and visualizations. The two intervention groups were compared with a control group regarding learning outcome in a learning session on the formation of polar lights. Prior knowledge, reading comprehension and learning time were assessed as control variables. Hypothesis 2 predicted that the intervention groups achieve higher posttest scores than the control group. In contrast to this hypothesis, results showed that there were no significant differences between the three groups on learning outcome. Research Question 6 addressed differential effects of the two interventions; in this experiment, both interventions had no influence on learning outcome.

There is first evidence that EMME have beneficial effects on multimedia learning (Mason et al., 2015a, 2015b), so why did the strategic support in this experiment not work?

Since learning outcome on the recall and transfer task was poor in all groups (with overall means of 51.5 % for recall and 52.21 % for transfer), it is possible that there was a floor effect for these tasks. Consequently, I used different learning materials in Experiment 4.

Another explanation is that learners may have failed at transferring the processing strategies to the new set of materials. In the areas where EMME have been shown to work, for example, medical diagnosis, the task with the modeling examples was very similar to the actual task, making it easy to transfer the relevant aspects of viewing behavior onto the new task. For example, in the study by Jarodzka et al. (2012), the participants’ task was to diagnose the occurrence of seizures in infants, and the crucial aspect was which body part to look at. This can easily be applied to a new patient. Likewise, the crucial aspect in the classification of fish locomotion patterns (Jarodzka et al., 2013) is which part of the fish to look at. In the case of multimedia learning, the relevant processing strategies are more abstract, and learners first have to identify the strategies and then apply them. In the studies by Mason et al., only integration processes were promoted, which might make the support easier to transfer to the new material. In this study, applying the more extensive strategic support to the novel learning material may have resulted in increased cognitive strain, which in turn prevented learning benefits.

This effect may even have been enlarged by an interference with pre-existing, automated strategies which learners actively tried to replace with the new ones. According to Hasselhorn and Körkel (1986), pre-existing strategies influence how well a new strategy is picked up; since multimedia learning is a task which students face on an everyday basis, they may well have a set of (inadequate) processing strategies that interfered with our interventions. This might also explain why EMME worked in the studies by Mason et al. (2015a, 2015b) whose participants were children who may not yet have pre-existing strategies.

The design choices made for this experiment may also be a reason why the EMME had no effect.

First, displaying the EMME on parts of the actual learning material might make it easier for learners to apply the processing strategies, thus freeing cognitive capacity for the learning task itself. Importantly, if the modeling is faded out throughout the learning phase, processing strategies will still have to be transferred to novel stimuli. Allowing for transfer to novel material is one the main intentions for investigating EMME as an alternative to previous approaches in supporting multimedia learning.

Second, the model’s eye movements were visualized with a scanpath, whereby circles are moving across the material. There is, however, evidence that a spotlight display is the better design choice (Jarodzka et al., 2012, 2013). Consequently, the EMME used in this study might not have worked because they featured a circle display. In the case of the Cueing condition, the same explanation may apply since the highlights and zoom-outs added rather than reduced visual complexity.

Third, adding audio commentary might have hindered the effect of the EMME. According to van Gog, Jarodzka, Scheiter, Gerjets, and Paas (2009), learners could experience increased working memory load when attending to both the visualization and the audio commentary simultaneously. Furthermore, in the studies by Mason et al. (2015a, 2015b) where EMME helped multimedia learning, no audio commentary was used.

Consequently, in Experiment 4, the EMME were presented on the learning material, featured a spotlight display and had no audio commentary.

In Experiment 3, domain-specific prior knowledge was only measured with three multiple-choice items, which might not have assessed the full extent of individual differences. Prior knowledge is an important factor when investigating the effectiveness of instructional support. Previous research has shown that the effects of instructional support are influenced by learners' prerequisite knowledge (Scheiter, Schüler, Gerjets, Huk, & Hesse, 2014; Seufert, 2003).

Mason et al. (2015b) found that EMME were effective only for students with poorer comprehension skills, thereby compensating for a lack of processing skills. In Experiment 4, I investigated whether domain-specific prior knowledge rather than comprehension skills would moderate the effects of complex EMME. Given that both the EMME and the materials were more complex than those used by Mason et al. and that a different cognitive construct was addressed, I expected a different type of moderating effect. Particularly, I assumed that visual guidance on how to effectively process multimedia materials would be helpful only for students who already possess relevant background knowledge that helps them interpret the information for which they receive processing support.

I did not collect process data in the present experiment, so it is unclear if the interventions influenced the way learners process the materials. If EMME and Cueing did not result in different visual processing, the finding that they did not support learning is not surprising. Analyzing learners' eye movements provides information on visual processing of the materials and could therefore help to understand if and how EMME have beneficial effects on multimedia learning.

The issues discussed above are addressed in Experiment 4, which I describe in the following.

## 3.4 Experiment 4

I conducted Experiment 4 to investigate the effects of a modified version of the EMME. For this purpose, I compared learning outcome and gaze behavior in an experimental group, which received modeling examples, with a no intervention Control group. The instructional support was provided on the learning material on cell division, a domain where I expected to find considerable individual differences in prerequisite knowledge.

I collected eye movement data to investigate the effect of EMME on learners' visual processing of the multimedia lesson. I assumed that seeing an expert's eye movements on the learning material leads to more effective processing of the material, as indicated by longer text and picture fixation times (*selection and organization processes*) and more transitions (*integration processes*).

I did not include a Cueing group like in Experiment 3 because the collection and analysis of gaze data is an extensive task and I was primarily interested in the effect of EMME.

To sum up, in the present experiment I investigated the effect of EMME on visual processing and learning outcome in a multimedia learning session. In the experimental group, learners saw an expert's eye movements on the first pages of the material. A control group was presented with the same material minus the modeling examples. Eye tracking data was recorded during the learning phase and cognitive prerequisites were assessed as moderators.

Watching EMME should improve learners' visual processing of the material, as demonstrated by more effective viewing behavior. More specifically, I expected longer overall text and picture fixation times and more transitions in the experimental than in the control group (Hypothesis 3). According to Hypothesis 4, I expected EMME to have beneficial effects on learning outcomes, such as higher posttest scores in the experimental condition compared with the control condition. Furthermore, I was interested in the potential moderating influence of cognitive prerequisites, namely students' domain knowledge, on the effect by EMME. In particular, I expected that improvements in learning outcomes due to EMME would be more pronounced in students with stronger rather than weaker cognitive prerequisites (Hypothesis 5). Finally, I assumed that changes in students' visual processing due to EMME would be suited to explain differences in learning outcomes (mediation, Hypothesis 6).

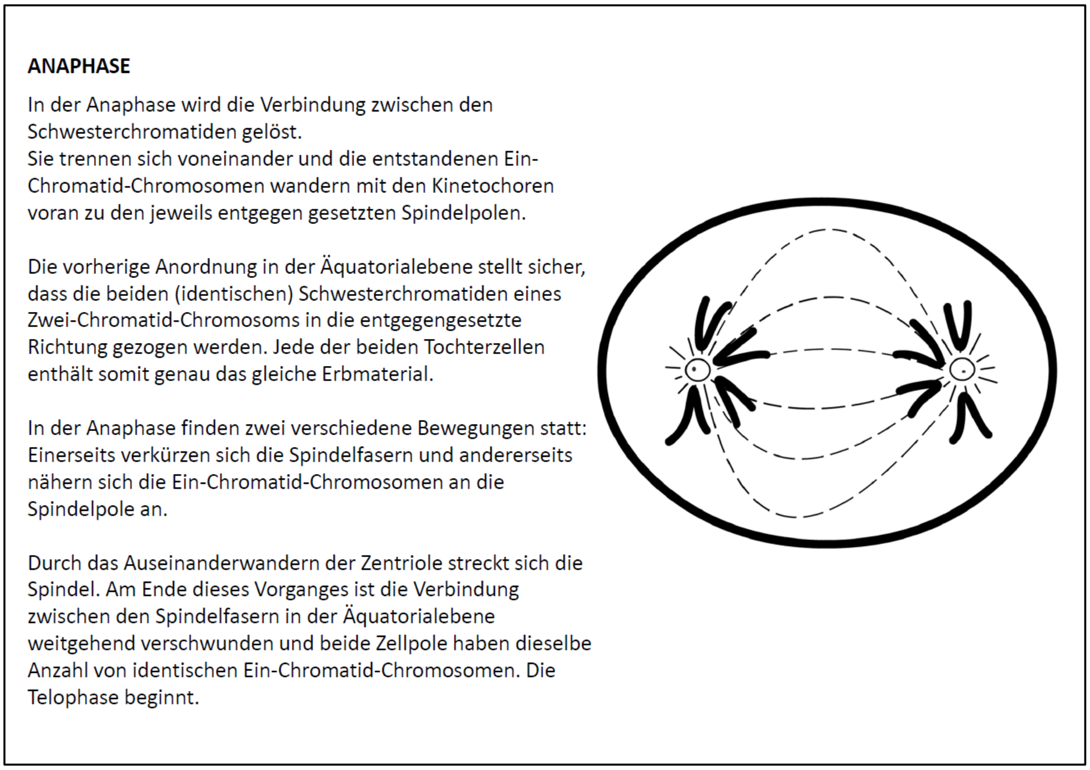
### 3.4.1 Method

#### *3.4.1.1 Participants and design*

Participants were 53 students of the University of Tuebingen (41 female; *M* = 26.89 years, *SD* = 6.81) enrolled in different courses. Students of biology and related fields were excluded from participation. Participation was voluntary and reimbursed with ten euros. Experimental condition was manipulated as a between-subjects factor: participants were randomly assigned to either the EMME (*n* = 27) or the control (*n* = 26) condition. Two different measures of learning outcome (recall and transfer) were assessed as dependent variables.

#### *3.4.1.2 Materials and apparatus*

The learning material was the one used in Experiments 1 and 2 (cf. section *2.3.1.2*). However, the realistic pictures were exchanged for schematic illustrations to reduce visual complexity and make it easier to interpret the model’s gaze behavior (cf. Figure 8).



**ANAPHASE**

In der Anaphase wird die Verbindung zwischen den Schwesterchromatiden gelöst.
Sie trennen sich voneinander und die entstandenen Ein-Chromatid-Chromosomen wandern mit den Kinetochoren voran zu den jeweils entgegen gesetzten Spindelpolen.

Die vorherige Anordnung in der Äquatorialebene stellt sicher, dass die beiden (identischen) Schwesterchromatiden eines Zwei-Chromatid-Chromosoms in die entgegengesetzte Richtung gezogen werden. Jede der beiden Tochterzellen enthält somit genau das gleiche Erbmaterial.

In der Anaphase finden zwei verschiedene Bewegungen statt: Einerseits verkürzen sich die Spindelfasern und andererseits nähern sich die Ein-Chromatid-Chromosomen an die Spindelpole an.

Durch das Auseinanderwandern der Zentriole streckt sich die Spindel. Am Ende dieses Vorganges ist die Verbindung zwischen den Spindelfasern in der Äquatorialebene weitgehend verschwunden und beide Zellpole haben dieselbe Anzahl von identischen Ein-Chromatid-Chromosomen. Die Telophase beginnt.

*Figure 8.* Exemplary page of the learning material in Experiment 4.

In the experimental group, EMME were displayed on the first four pages of the learning materials which served only to introduce definitions of terms relevant to cell division like the structure of a cell and chromosomes. The EMME demonstrated various multimedia processing strategies that had been derived from the literature. Upon presentation of each of the four pages, the model comprehensively inspected the picture by scanning it and then read the title (*construction of a pictorial scaffold*). Afterwards, she read the whole text while fixating relevant words and looked carefully at the picture to get an idea of the representation contents (*selection of* *relevant words and picture elements*). She then read the text again section by section (*text organization*) and looked at the corresponding picture elements (*picture organization*), switching between the representations (*integration*). In the end, she looked again carefully at the picture (*final picture inspection*) while using it as a scaffold for checking her mental representation of the learning content and reread in some instances the corresponding sections of the text (*reaction to comprehension problems*). Focus maps were used to visualize the skilled learner's eye movements, while a white light spot representing a gaze fixation moved across the otherwise shaded page. The skilled learner was a student assistant who was instructed how to process the materials in line with the processing strategies. Duration of the presentation of the modeling examples on the first four pages of the learning material was 39 to 87 seconds (M = 71.75 s).

After the EMME had been displayed on the first page, the word "next" appeared in the right bottom corner of the page. Upon clicking the space bar, participants were shown the first page again, but this time without EMME. This was done to allow them to study the page as long as they wanted before proceeding to the second page. This procedure was repeated for each of the first four pages of the learning material.

In the control group, no EMME were displayed on the first four pages of the learning material. To keep learning times across conditions comparable and to have similar presentation procedures across both conditions, each of the four pages was displayed for the same length of time as in the EMME group before the word "next" appeared and learners could proceed to the subsequent page.

In both conditions, each of the six remaining pages was displayed for 50 seconds before learners could continue. This ensured a minimum learning time for each page in both conditions, but it was up to the learner to spend more time on each page if wanted.

*Apparatus.* The model's and the participants' eye movements were recorded using an SMI remote eye tracker (RED 250) with a sampling rate of 250 Hz and the iView X 2.2 and Experiment Center 3.3 software. The learning material was presented on a 22" widescreen monitor. Eye tracking data were edited and prepared for statistical analysis with BeGaze 3.3 software (www.smivision.com).

### *3.4.1.3 Measures and scoring*

Measures encompassed cognitive prerequisites (i.e., domain-specific prior knowledge and scientific literacy), posttest performance and gaze data.

*Cognitive prerequisites test.* The cognitive prerequisites questionnaire consisted of a measure for scientific literacy and a measure for domain-specific prior knowledge; it was identical to the one used in Experiments 1 and 2 (cf. section *2.3.1.3*) and was also scored accordingly. As in Experiment 1, one comprehensive measure of cognitive prerequisites was computed by adding up the z-standardized total scores of scientific literacy and domain-specific prior knowledge (Cronbach's $\alpha = .71$).

*Posttest*. The paper-pencil based posttest assessed recall and transfer. It consisted of three subtests, which were identical to Experiment 2 (cf. section *2.4.1.4*): a free recall question, a forced-choice verification task and 16 multiple-choice-items. Four items from the multiple-choice test and one item from the forced-choice verification task were removed from the analysis to ensure that the posttest would refer only to contents which had been explained on pages with no modeling in the experimental conditions, thereby assessing students' ability to transfer their newly acquired strategies to novel content. For the free recall question only those aspects were scored that addressed non-modeled contents.

Mean scores from the free recall task, the 19 forced-choice verification recall items and from the four multiple choice recall items were averaged and transformed into percentages to obtain a measure of recall test performance. The mean scores of the eight multiple-choice transfer items and of the two forced-choice verification transfer items were averaged and transformed into percentages to obtain a measure of transfer test performance.

*Eye-movement measures***.** Eye movement data from the four pages on which the EMME were displayed in the experimental group were excluded since viewing behavior was externally guided on those pages. Therefore, I analyzed the eye movement data from six pages.

Areas of interest (AOIs) were defined for each page. To determine the number and duration of fixations on both representations as well as the number of transitions between them, one AOI was created around the text and one around the picture. Eye movement data was averaged across the six pages of the learning material. For each participant, the mean time per page spent on the text (fixation time text) and on the picture (fixation time picture), as well as the number of transitions between text and pictures were computed.

#### *3.4.1.4 Procedure*

Data collection took place in individual sessions. Participants were randomly assigned to the experimental or control condition, received written information on the experiment procedures, signed a consent form, and completed a paper-pencil test to assess their cognitive prerequisites. They were then seated in front of the eye tracker at a distance of approximately 60 cm. The eye tracker was calibrated for each participant using a nine-point calibration. The experimenter then started the learning environment either with or without EMME. The onscreen instruction informed participants that they could learn at their own pace and proceed to the next page by pressing the space bar as soon as the word “next” appeared on the screen, but that they could not go back to the (learning) material. Participants were also informed that they would be tested regarding the content. In the experimental group, learners were instructed that they would see a successful learner's eye movements on the first pages of the material. They were further told that these eye movements would be illustrated by light spots moving on the grey background of the page. They were also informed that the size of the spot corresponded to the models' fixation times, with larger spots illustrating longer fixations. Eye movements were recorded during learning. After the learning phase, they filled out the posttest using paper and pencil, were paid and debriefed. In total, each session lasted about 75 min.

### 3.4.2 Results

Due to poor gaze data quality, three subjects were excluded from analysis. For all statistical analyses reported here, the level of significance was set at $\alpha$ = .05. Statistical analysis was conducted with IBM SPSS 20.0 software.

#### *3.4.2.1 Learning outcome*

To investigate the effects of EMME and the influence of cognitive prerequisites, I used regression analyses with effect coding to analyze the data. Thereby, the experimental condition is coded in a way that reflects the hypothesis regarding its effect. In this case, I expected the EMME group to outperform the control group. Consequently, the experimental group was coded +1, and the control group was coded -1. Experimental condition, cognitive prerequisites (z-standardized), and the interaction terms between cognitive prerequisites and condition were entered simultaneously as predictors. Table 10 shows means and standard deviations of test performance and prerequisite knowledge.

Table 10: *Means and standard deviations for the cognitive prerequisite measures and learning outcome measures as a function of experimental condition in Experiment 3*

| | **EMME (*n* = 26)** | | **Control (*n* = 24)** | |
|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Prior knowledge % | 44.62 | 13.03 | 50.56 | 17.57 |
| Scientific literacy % | 72.28 | 12.19 | 74.65 | 18.55 |
| Recall % | 43.62 | 11.83 | 41.08 | 9.76 |
| Transfer % | 55.68 | 12.06 | 48.96 | 17.35 |

For recall, the regression model was significant, adj. $R^2 = .18$, $MSE = 9.81$, $F(3,45) =4.53$, $p = .01$. There was no main effect of experimental condition but a significant effect of students' cognitive prerequisites: students with stronger cognitive prerequisites achieved higher learning outcomes. There was also a marginally significant interaction between condition and students' cognitive prerequisites (cf. Table 11 for the results of the regression analysis). To interpret this interaction, a simple slopes analysis was conducted (at -1 and +1 standard deviation of the continuous variable cognitive prerequisites). It revealed that for learners with weaker cognitive prerequisites, the EMME had no effect on learning outcome ($B = -.99$, $SE = 2.06$, $\beta = -.09$, $p = .63$), whereas they improved learning outcome for learners with stronger cognitive prerequisites ($B = 4.96$, $SE = 2.12$, $\beta = .46$, $p = .02$; cf. Figure 9).

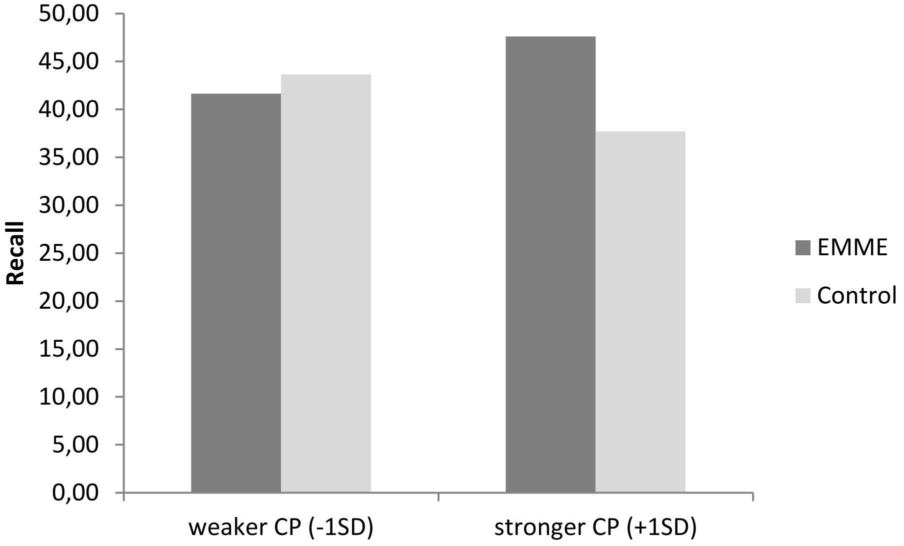


*Figure 9*. Recall performance as a function of experimental condition for learners with weaker and stronger cognitive prerequisites (CP) in Experiment 4.

Table 11: *Results of the regression analyses for the learning outcome measures in Experiment 4*

| | **Recall** | | | |
|---|---|---|---|---|
| | **Beta** | ***SE*** | ***B*** | ***p*** |
| Constant | 42.65 | 1.42 | | <.001 |
| Condition | 1.98 | 1.42 | .19 | .17 |
| Cognitive prerequisites | 5.35 | 1.53 | .49 | .001 |
| Condition x cognitive prerequisites | 2.97 | 1.53 | .27 | .06 |
| | **Transfer** | | | |
| | **Beta** | ***SE*** | ***B*** | ***p*** |
| Constant | 52.55 | 2.15 | | <.001 |
| Condition | 3.38 | 2.15 | .23 | .12 |
| Cognitive prerequisites | .87 | 2.29 | .06 | .71 |
| Condition x cognitive prerequisites | 2.06 | 2.29 | .14 | .37 |

For transfer, the regression model was not significant, adj. $R^2$ = .00, *MSE* = 14.94, *F*(3,46) = 1.06, *p* = .38. None of the predictors had a significant influence on learning outcome.

### 3.4.2.2 Eye movements

Since based on the Kolmogorov-Smirnov Test, picture fixation times (*p* < .001) and the number of transitions (*p* < .01) were not normally distributed, the eye tracking data was submitted to a log-transformation. Multiple regression analyses were conducted for text fixation times, picture fixation times, and the number of transitions, respectively. Again, experimental condition (control group coded -1; experimental group coded +1), cognitive prerequisites (z-standardized) and the interaction term of cognitive prerequisites and experimental condition were entered simultaneously as predictors. For means and standard deviations of the eye tracking measures, see Table 12. The results of the regression analyses are displayed in Table 13.

Table 12: *Means and standard deviations for the eye tracking measures as a function of experimental condition in Experiment 4*

| | **EMME (n = 26)** | | **Control (n = 24)** | |
|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Text fixation time (log)* | 12.59 | .30 | 12.49 | .32 |
| | (305942.03) | (84045.32) | (279561.93) | (92576.15) |
| Picture fixation time (log)* | 11.00 | .55 | 10.58 | .96 |
| | (69548.28) | (38511.01) | (49823.96) | (26854.99) |
| Number of transitions (log)* | 3.29 | .27 | 2.11 | .73 |
| | (27.83) | (6.70) | (9.67) | (4.40) |

*Note: To allow for easier interpretation of the log-transformed eye tracking data, means and standard deviations without log-transformation are additionally provided in brackets (in milliseconds).

The overall regression models for text fixation times and picture fixation times were not significant (text fixation times: F < 1; picture fixation times: *F*(3,49) = 1.68, *MSE* = 0.61, *p* = .19, *adjusted* $R^2$ = .04). However, EMME increased the time students spent studying the pictures.

For the number of transitions, there was a significant overall model, *F*(3,49) = 19.36, *MSE* = 0.30, *p* < .001, *adjusted* $R^2$ = .53. The number of transitions was predicted by experimental condition: Learners in the EMME group made more transitions between text and pictures than learners in the control group. Neither cognitive prerequisites nor their interaction with experimental condition had an effect on students’ attempts to integrate text and pictures.

Table 13: *Results of the regression analyses for the eye tracking measures in Experiment 4*

| | **Text fixation times** | | | |
|---|---|---|---|---|
| | **Beta** | ***SE*** | ***B*** | ***p*** |
| Constant | 12.55 | .05 | | <.001 |
| Condition | .05 | .05 | .16 | .27 |
| Cognitive prerequisites | <.01 | .05 | .02 | .89 |
| Condition x cognitive prerequisites | .04 | .05 | .13 | .39 |
| | **Picture fixation times** | | | |
| | **Beta** | ***SE*** | ***B*** | ***p*** |
| Constant | 10.81 | .11 | | <.001 |
| Condition | .23 | .11 | .29 | .04 |
| Cognitive prerequisites | .11 | .12 | .14 | .35 |
| Condition x cognitive prerequisites | .11 | .12 | .13 | .38 |
| | **Number of transitions** | | | |
| | **Beta** | ***SE*** | ***B*** | ***p*** |
| Constant | 2.71 | .08 | | <.001 |
| Condition | .60 | .08 | .76 | <.001 |
| Cognitive prerequisites | .05 | .08 | .06 | .55 |
| Condition x cognitive prerequisites | .02 | .08 | .03 | .78 |

#### *3.4.2.3 Linking learning outcomes and eye movements.*

In a final step I analyzed whether the changes in the way students processed the materials as a consequence of having viewed EMME prior to learning were suited to explain effects on learning outcomes by means of several mediation analyses (Preacher & Hayes, 2008). As mediators I chose the fixation time for pictures and the number of transitions, for which there had been effects by EMME. These were used to explain the effect of EMME on recall performance which was moderated by students' cognitive prerequisites (cf. conditional process model, Hayes, 2013). The mediation analysis generated 95% bias-corrected and accelerated bootstrap confidence intervals for the indirect effects using 5000 bootstrap samples. There were no significant indirect effects of EMME on recall performance using either picture fixation time or transitions between text and pictures as mediators (with picture fixation time: *coefficient* = 0.36, SE = .54, 95% CI [-.39, 1.88]; with transitions: *coefficient* = 1.47, SE = 2.40, 95% CI [-2.69, 76.38]).

### 3.4.3 Discussion of Experiment 4

This study aimed at investigating the effects of an EMME intervention on gaze behavior and learning outcome. According to Hypothesis 3, I expected learners in the EMME group to show a more effective viewing behavior, that is, longer text and picture fixation times and more transitions. This hypothesis is partially supported by the data: learners in the experimental group had longer picture fixation times and more transitions. There were no group differences on text fixation times. In light of evidence indicating that most learners process multimedia material in a text-driven manner (Hannus & Hyönä, 1999; Scheiter & Eitel, 2010; Schmidt-Weigand, Kohnert, & Glowalla, 2010a, 2010b), however, it is not surprising that also learners in the control group showed long text fixation times.

Contrary to Hypothesis 4, there was no main effect of EMME on learning outcome.

Regarding the role of cognitive prerequisites, and in line with Hypothesis 5, the analysis yielded an interaction between cognitive prerequisites and learning outcome on recall. Only learners with stronger cognitive prerequisites benefited from the modeling examples on learning outcome. This interaction was not found for transfer. Cognitive prerequisites also predicted learning outcome for recall. These results strengthen the notion that cognitive prerequisites are an important variable in the investigation of multimedia learning. Hypothesis 6 predicted that learners' viewing behavior mediates learning outcome, which was not confirmed in my analysis. Although EMME led to more transitions and longer picture fixation times and to higher recall performance for learners with stronger cognitive prerequisites, the improved viewing behavior did not mediate the effect of the EMME on learning.

## 3.5 Discussion: self-regulation support for multimedia learning

In this chapter, I reported two experiments that examined the effects of EMME on learning from multimedia instruction.

In Experiment 1, learners in the EMME group were presented a video with the eye movements of a successful learner on exemplary learning material. This video contained an explanatory audio commentary. The model's eye movements were visualized with a circle display. In the Visual Cues group, the same processes as in the EMME intervention were conveyed. However, they were visualized with signals (e.g., highlights) instead of eye movements. A third group received no intervention. All three groups subsequently learned with new learning material and their learning outcome was compared. Contrary Hypothesis 2, there were no group differences: Neither the EMME nor the Visual Cues intervention had an effect on learning. Consequently, regarding Research Question 6, there were also no differential effects of the two interventions.

Some design choices for the EMME in Experiment 3 may explain the lack of an effect, so I conducted another experiment where the EMME were designed differently. More specifically, the EMME support was given on the first pages of the actual learning material on cell division instead of exemplary material; furthermore, the eye movements were visualized with a spotlight display instead of a scanpath. There was no explanatory audio commentary. To investigate the role of cognitive prerequisites in the effectiveness of EMME, I included them as moderator variable. Gaze data and learning outcome were assessed as dependent measures.

In line with Hypothesis 3, the EMME group had longer picture fixation times and more transitions than the control group, indicating that seeing an expert's eye movements improved visual processing of the instruction. This effect was especially large for integration processes. While learners in the control group showed a mean of ten transitions between the representations, this number went up to 28 in the EMME group. These results support the assumption that modeling eye movements can support effective cognitive processing in multimedia learning.

However, this advanced processing was only partially reflected in learning success, as there was no main effect of EMME (contrary to Hypothesis 4). Yet, in line with Hypothesis 5, there was a significant interaction between cognitive prerequisites and experimental condition for recall. While all learners in the EMME condition applied the new processing strategies to the instruction (as indicated by the improved viewing behavior in the experimental condition), only learners with stronger cognitive prerequisites were better able to memorize the learning content, as indicated by higher recall scores. In this study, there seems to be only a weak linkage between eye movements and learning. This is evidenced by the fact that picture fixation times and the number of transitions did not mediate the effect of the EMME on recall, even when taking the moderating effect of cognitive prerequisites into account. It seems that the increased picture fixation times and number of transitions did not result in improvements in learning for learners with weak cognitive prerequisites. For learners with stronger prerequisites, the EMME improved viewing behavior and memorization, but not deeper understanding. One might argue that changes in visual attention are only a by-product of the intervention. However, that is unlikely as other studies found a connection between visual behavior and learning outcome (Johnson & Mayer, 2012; Mason et al., 2013; Scheiter & Eitel, 2015) and the EMME intervention was tailored to support these very processes.

A possible explanation for why transitions were not predictive for learning in the present task may be that the effect of EMME on recall was quite small in the first place. A mediating effect of visual behavior may therefore have failed to reach significance.

The results may also be attributed to the nature of the materials. The texts per page were relatively short and the pictures were highly schematized. Thus, integration of text and pictures might be achieved by holding active in memory information from one representation while processing the other representation, thus requiring fewer switches between the two representations (Bauhoff, Huff, & Schwan, 2012). Therefore, even though students did switch between text and pictures very frequently after having viewed EMME, this may in the present study not have been required for achieving better performance. On the other hand, children who served as participants in the studies by Mason et al. (2015a, 2015b) in which transitions proved predictive for learning, may not yet possess the working-memory capacity to resort to memory-based integration strategies.

It is also possible that the relation between number of transitions and learning outcome is not linear. An increased number of transition may support learning until a cutoff point, after which additional transitions do not help with learning. Learners in the EMME condition may have made more transitions than necessary to integrate text and picture.

An alternative explanation is that learners managed to “copy” the model’s viewing behavior, but the cognitive strain resulting from applying the processing strategies interfered with learning. This argument is strengthened by the finding that only learners with stronger cognitive prerequisites benefitted from the EMME on retention. One might reason that for learners with weak prerequisites, cognitive strain induced by applying the new strategies was too high to learn well regarding both recall and transfer. Learners with stronger prerequisites had sufficient cognitive capacity left to perform well on recall, but not enough to perform well on transfer of the learning content.

Another explanation why the EMME supported recall in some learners, but not transfer can be derived from CTML. According to Mayer and Johnson (2008), successful multimedia learning occurs when extraneous processing (i.e., processing that does not contribute to learning, like visual search) is reduced, and essential processing (needed to select the relevant information) as well as generative processing (deeper processing related to organization and selection processes) are fostered. It is possible that the EMME were able to guide students' attention to relevant aspects of the learning content, thereby reducing extraneous and fostering essential processing and resulting in effective selection of the relevant information. As selection processes are mainly relevant for retention, recall performance was supported by the EMME, at least for learners with stronger cognitive prerequisites. The EMME may not have fostered generative processing, explaining the lack of an effect on transfer test performance.

Another interesting result of Experiment 4 is the main effect of cognitive prerequisites for recall performance. This finding, and the interaction between cognitive prerequisites and experimental condition on recall performance, strengthen the notion that prerequisite knowledge is an important variable in multimedia learning.

Taken together, the results of this experiment indicate that, for learners with a certain level of prerequisite knowledge, EMME can support effective processing of multimedia instruction. Most importantly, this support worked beyond the pages of the material where the modeling examples were displayed. Therefore, learners were able to transfer the processing strategies to new stimuli. However. contrary to Hypothesis 6, eye movements did not mediate learning success.

Other than Mason et al. (2015a, 2015b), I used EMME to convey not one, but several effective processes for multimedia learning. Although the EMME were complex, learners were able to benefit from them, at least those with stronger cognitive prerequisites.

In addition, the findings suggest several guidelines for the design of EMME interventions in multimedia learning. First, providing the EMME on the actual learning material instead of exemplary material seems to facilitate implementing the new processing strategies. Second, in line with previous findings (Jarodzka et al., 2012, 2013), displaying the EMME in a way that reduces rather than adds visual complexity seems to be important. Third, in the experiments reported here, EMME without audio commentary (Experiment 4) improved performance, whilst EMME with audio commentary (Experiment 3) did not. These findings may be attributed to increased working memory load due to the simultaneous attendance to visual and auditory instruction (cf. van Gog et al., 2009).

In total, the results of Experiment 4 show that EMME are a promising instructional tool to support multimedia learning.

# 4 General Discussion

The multimedia effect (Mayer, 2005) states that adding pictures to text improves learning. The most prominent theoretical account of multimedia learning, CTML (Mayer, 2009), postulates that to learn effectively, learners have to engage actively in selection, organization and integration processes. Although there is manifold evidence that learners benefit from multimedia instruction (Anglin et al., 2004), some learners do not. Eye tracking research suggests that, as stated by CTML, successful multimedia learning is associated with several cognitive processes, and that some learners do not process the materials adequately.

In this dissertation, I investigated two different gaze based approaches to support multimedia learning. The first approach provided external support that aimed at prompting adequate processing of the instruction. More specifically, I developed and evaluated an adaptive system that diagnoses inadequate visual processing of a multimedia session on mitosis and alters the instruction accordingly (Experiments 1 and 2).

The second approach provided self-regulation support by showing learners Eye Movement Modeling Examples (EMME). The EMME used a replay of a successful learner's eye movements to illustrate adequate processing of the instruction (Experiments 3 and 4).

## 4.1 Summary of results

In Experiment 1, learners' viewing behavior and its relation to learning success was investigated. The objective of this experiment was to examine which parameters to use for the gaze-based adaptive system, and to derive threshold values. The experiment was guided by three research questions. Research question 1 addressed whether viewing behavior can distinguish successful from less successful learners. The results indicate that learners differ in the way they visually process the instruction, and that these differences are reflected in learning outcome. Two main clusters of viewing behavior were identified. Successful learners were characterized by more and longer text and picture fixation times and more transitions between the representations. These findings are in line with CTML (Mayer, 2005) which postulates that successful multimedia learning is associated with selection, organization and integration processes. Research Question 2 was which parameters to implement into the adaptive system; based on the results, I decided to use text and picture fixation times as well as the number of transitions. Research Question 2 addressed the role of cognitive prerequisites for the distinction of successful from unsuccessful learners. Interestingly, the two clusters of viewing behavior did not differ in cognitive prerequisites; as a consequence, cognitive prerequisites were not implemented into the algorithm of the adaptive system.

Based on the findings of Experiment 1, the adaptive system was programmed to detect inadequate visual processing, that is, too short text or picture fixation times or too few transitions. The system then altered the instruction to prompt adequate processing when necessary. For example, short picture fixation times triggered a zoom-out of the picture to promote picture selection and organization processes. In Experiment 2, the adaptive system was compared to a control (static) version of the learning environment on learning outcome. Hypothesis 1 predicted better learning outcome in the experimental group. Research Question 4 addressed the role of cognitive prerequisites on the effects of the adaptive system. However, there was no main effect of experimental condition and no interaction of experimental condition with cognitive prerequisites. Research Question 5 addressed the usability of the adaptive system; results indicate that ease of use is perceived as in the traditional, static version of the learning environment.

In Experiment 3, EMME were used to convey nine adequate processes for multimedia learning, including selection, organization and integration. The EMME were presented on exemplary material using a circle display with an explanatory audio commentary. A second experimental group received a Visual Cues intervention, which was identical to the EMME, but the processes were visualized with cues instead of Eye Movements. A third group received no intervention. All three groups then learned with novel material and were tested for learning outcome. I expected both interventions to lead to better learning outcome (Hypothesis 2); differential effects of the EMME compared to the Visual Cues were addressed in Research Question 6. However, there were no group differences on learning outcome. Based on these results, I conducted Experiment 4, where I used EMME that differed from the ones in Experiment 3 in several aspects. First, they were presented on the first pages of the learning material instead of exemplary material. Second, there was no audio commentary. Third, the eye movements were visualized with a spotlight display instead of a circle display. Learning outcome and eye movement were assessed as dependent variables. I expected the EMME to improve viewing behavior, as indicated by longer fixation times and more transitions (Hypothesis 3). I also expected learners in the EMME group to achieve higher learning outcome (Hypothesis 4). Furthermore, I measured cognitive prerequisites as a possible moderator variable; Hypothesis 5 predicted that especially learners with stronger cognitive prerequisites would benefit from the EMME. The hypotheses were partially supported by the data. The EMME group showed longer picture fixation times and more transitions than the control group. Furthermore, there was an interaction between experimental condition and cognitive prerequisites in the predicted direction: the positive effect of EMME on recall performance occurred only for learners with stronger cognitive prerequisites.

## 4.2 External vs. self-regulation support in multimedia learning: comparing the two gaze-based interventions

In this dissertation, I investigated two gaze-based interventions to support multimedia learning. Both interventions aimed at promoting processes that had been found to be important for effective learning from illustrated texts. The adaptive system diagnosed inadequate visual processing as indicated by short text and picture fixation times and insufficient transitions. It then altered the instruction to prompt selection and organization processes by presenting zoom-outs of text and/or picture, and integration processes by presenting a color-coded version of the learning page. The adaptive system had no beneficial effect on learning outcome, which may be explained by several aspects as discussed in section *2.4.3*. For example, it is possible that the algorithm included learners whose processing would have resulted in successful learning, and whose learning was hindered by the adaptations.

The EMME aimed at supporting self-regulated learning. More specifically, it provided support for the inner layer of Boekaerts' model of self-regulated learning, the choice and application of cognitive strategies, which is essential for self-regulated learning (Boekaerts, 1999). Therefore, nine cognitive processes were conveyed by the eye movements of a successful learner and could then be applied to the learning material in a self-regulated way. The results indicate that learners picked up the processes and applied them to the materials, as learners in the EMME group had longer picture fixation times and more transitions. For learners with stronger cognitive prerequisites, this viewing behavior also resulted in better learning outcome.

Although the EMME were effective, they had a less clear-cut and pronounced effect than similar interventions in studies with children (Mason et al., 2015a, 2015b). The information conveyed by EMME is very implicit; maybe adult learners need more explicit instructions.

Furthermore, it is possible that regulation of the learning process is more problematic for adult learners due to interference with pre-existing strategies. According to the WWW&H (Veenman, Van Hout-Wolters, & Afflerbach, 2006) rule, instructional support should tell a learner *what* to, *when*, *why*, and *how*. The *what* aspect is the most explicit in EMME; the other aspects are approached very implicitly, if at all. The *why* component is not part of EMME. Possibly, extending EMME in a way that includes the other components may be necessary, for example by adding explanatory audio commentary. In Experiment 3, the EMME with commentary had no effect; however, many aspects of the intervention were changed between Experiments 3 and 4, and the audio commentary may have been beneficial. Furthermore, the EMME may be combined with other methods such as implementation intentions (Gollwitzer & Brandstatter, 1997). Implementation intentions are “if-then” plans which connect an opportunity to act with an action; in other words, they explicitly support the learners in *what* to do *when*. There is first evidence that implementation intentions (e.g., “IF I have finished reading a paragraph, THEN I will search for corresponding information in the picture”) support multimedia learning (Stalbovs et al., 2015).

Regarding the WWW&H rule (Veenman et al., 2006), the adaptive system is even more implicit than EMME, as it prompts effective processing indirectly by altering the instruction. Therefore, the learner receives implicit information on *what* to do, and no information on *when*, *why* and *how* to do it. Such information might be included into the adaptive system, for example by providing an explanatory instruction (possibly even combined with EMME) before the learning session; the adaptations would then provide feedback on the learning process (e.g., when the highlighted version is presented, the learner knows that (s)he made too few transitions). Furthermore, the adaptation could be accompanied by explanatory prompts which explain why the adaptation occurs and which process the learners should monitor more closely, and why (s)he should do so.

Taken together, my results indicate that supporting self-regulation might be the more successful approach since the adaptive system had no effect, while the EMME did. However, the adaptive system is not fully evolved yet and it is too early to make definitive statements on its effectiveness.

The main advantage of supporting self-regulation is that once learners have internalized the processes, they can apply them to new materials. This means that they are able to adequately process even materials which are not well-designed. In contrast, providing external support always relies on materials which are specifically designed to prompt effective processing. However, an improved version of the adaptive system might be helpful to learners who struggle with self-regulating their learning process and who therefore would not benefit from EMME. Furthermore, it could support especially those learners who have weaker cognitive prerequisites, which may be a result of poor self-regulated learning in the first place.

## 4.3 The role of eye tracking in multimedia learning

In the experiments presented here, eye tracking was used as an instructional tool, which is an innovative approach to the use of eye tracking. Usually, eye tracking tools are used to investigate the learning process. Using them beyond collecting process data in order to support the learning process is a new research angle, although there is some previous work on EMME as an instructional method (Jarodzka et al., 2012; Mason et al., 2015b; Pluchino et al., 2013; van Gog et al., 2009). Evidence from the present experiments suggests that supporting learners' viewing behavior does not necessarily result in improved learning. Most importantly, Experiment 4 shows that EMME improved viewing behavior, but not learning outcome for all learners. The interaction of experimental condition with cognitive prerequisites on recall suggests that the relation between viewing behavior and learning success differs individually.

Transitions are discussed to be an indicator of integration processes, which in turn are a key process to successful multimedia learning. Therefore, transitions are of much interest in a multimedia learning context. However, previous research on the relation between transitions and learning outcome has produced ambiguous results. On the one hand, there is evidence that transitions between text and picture indicate successful learning (Hegarty & Just, 1993; Mason et al., 2013). Furthermore, promoting transitions by instructional design (Johnson & Mayer, 2012) or EMME (Mason et al., 2015a, 2015b) was shown to be associated with improved learning. On the other hand, Scheiter and Eitel (2015) found evidence that signals foster learning by promoting text-picture integration, but no effect on the number of transitions in two experiments. In the research presented here, transitions were not statistically linked to learning outcome (as discussed in section *3.4.3*). Taken together, empirical evidence on the role of transitions is not conclusive and more research is needed.

A general issue with eye tracking as a research tool is that multimedia learning is a very complex process – which we try to understand using very global measures like total fixation duration or number of transitions. It is not sufficient for successful learning to allocate enough visual attention to each representation or to switch between them often enough. Rather, it is crucial that learners process the instruction in a certain temporal order, thereby using the text to understand certain picture elements and vice versa. It is possible that two learners with very similar total fixation times or number of transitions have a quite different learning behavior and very different learning outcomes. Successful self-regulation of the learning process is a very complex matter which we, however, try to understand and promote using aggregated eye tracking measures. To really understand and successfully support multimedia learning, more fine-grained measures may be necessary. For example, in reading research, first- pass and second-pass indicators are distinguished (Hyönä & Nurminen, 2006). First-pass indicators are related to basic attentional processes, for example differences in visual salience of the stimuli, while second-pass indicators describe revisits to AOIs and are related to more intentional processes of comprehension (Scheiter & Eitel, in press). Very few studies in multimedia research have employed second-pass indicators (for an exception, see Mason et al., 2015a). Likewise, measures reflecting the temporal specifics of learners' gaze behavior, like scanpath analysis, have rarely been used (for exceptions, see Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Skuballa, Fortunski, & Renk, 2015).

Using global parameters is, thanks to advances in eye tracking software, relatively easy. More elaborate analyses are time-consuming and cost-intensive, which is probably the main reason for the lack of such research. However, studies using such measures would greatly benefit our understanding of multimedia learning.

In conclusion, while eye tracking is a valuable research instrument and a promising instructional tool for multimedia learning, more research using fine-grained analyses is called for.

## 4.4 Strengths

Experiment 1 provides a systematic examination of adults' natural viewing behavior in a multimedia lesson. It extends on findings from previous research that showed that individual differences in children's viewing behavior are related to learning success (Mason et al., 2013) and shows that the same is true for adults. Furthermore, it strengthens the assumptions of theoretical accounts of multimedia learning by indicating that selection, organization and integration processes are crucial for successful learning from multimedia instruction.

Within this dissertation, the first adaptive system for multimedia learning based solely on learners' gaze behavior was developed. This system can be adapted to new learning content, and the adaptation algorithm can easily be modified. The work presented here provides a great starting point for extensive research on the adaptive system with adjustments to the adaptation algorithm as discussed in section *2.4.3*. The research presented here was a first attempt at a gaze-based adaptive system. Although in this study, the system did not result in higher learning outcome and the algorithm will need to be revised, the tailored support provided by the adaptive system is a promising approach to supporting multimedia learning. The adaptive system is running and the algorithm can be altered easily, which is a great starting point for future research.

Furthermore, my research on EMME (Experiments 3 and 4) indicates that using eye tracking as an instructional tool is an auspicious way to promote multimedia learning, It extends on previous research (Mason et al., 2015a, 2015b) in showing that EMME also work for adults learners, at least to some extent.

Taken together, the work presented here employs an innovative and indicatory use of eye tracking, even when the research is still in its infancy.

## 4.5 Limitations and directions for future research

The main limitation of the research on the adaptive system is that it had no effect on learning performance, so the adaptation algorithm is obviously not sound yet. Based on this finding, there are several implications for future research. For example, one could collect additional data and rerun the cluster analysis to obtain more information on inadequate visual processing, especially on long fixations/many transitions as shown by learners of cluster 3 in Experiment 1. These learners had even longer fixation times and more transitions than the successful learners in cluster 1, but poor learning outcome. The cluster size was too small to be interpreted, but possibly there is a third group of learners who also need support from the adaptive system. For example, the algorithm could be programmed to detect not only too short, but also too long fixations and adapt accordingly. Also, adding more variables on learners' prerequisites like working memory capacity, reading comprehension or spatial ability may result in a better algorithm. The aim of this dissertation was to develop an adaptive system that uses as little variables as possible. However, my results indicate that it might be necessary to add some variables beyond viewing behavior to provide the best support for each individual learner.

Alternatively, the derivation of the threshold values based on Experiment 1 could be revised, for example by using the mean values of cluster 2 without adding a standard deviation, thereby lowering the threshold values. In this case, fewer learners would receive an adaptation, which might prevent successful learners from getting confused by unnecessary support. On the other hand, it is possible that more learners would benefit from an adaptation, which would be achieved by increasing the threshold values. In the system developed here, the algorithm can easily be altered, so that a variety of versions of the algorithm can be programmed and evaluated in a short period of time.

Furthermore, the effects of a more detailed instruction of the experimental group could be investigated, maybe even including a short training period to get used to the adaptive system. Verbal data could be included in future studies to gain more information on the system's effect on the learning process.

In the case of the EMME, three design choices were modified between Experiments 3 and 4: in the latter, the EMME were presented on the actual learning material, without audio commentary, and employed a spotlight display. Only the EMME with these design features supported learning. Since I changed all three aspects between the two experiments, it is still unclear which of them is most important to the effectiveness to EMME instruction. In future studies, it would be interesting to systematically manipulate them to identify the key aspects of successful modeling examples.

Likewise, the EMME conveyed a combination of nine cognitive strategies, and it is unclear if all of them are equally helpful for learners. Future studies could manipulate which processes the EMME convey and detect the best combination of strategies. Evidence that the combination of cognitive processes which are promoted by an intervention matters comes from a study by Stalbovs et al. (2015). They found that prompting a combination of widespread cognitive processes was better suited to support multimedia learning than text or picture processes only. At the same time, EMME should contain only those processes that really help learners. Including redundant processes puts an unnecessary strain on learners' working memory.

In the research presented here, EMME improved recall, but not transfer performance. Future research could address if EMME have to include different processes to foster transfer of the learning content. Alternatively, it is possible that the material used here and the posttest failed to show an effect that EMME do have on understanding, so replications with different learning domains are called for.

Although the findings from Experiment 4 suggest that there was short-term transfer of the processes the intervention promoted, it is yet to be examined if there will be long-term transfer. In light of the results from Experiment 3, it is unlikely that one single EMME intervention will foster learning in later multimedia learning sessions. An interesting research question would be the effects of several consecutive EMME interventions like the one I employed in Experiment 2. For example, learners could receive several multimedia instructions, each with modeling examples on the first pages, over an extended period of time. This approach would enable learners to internalize the processing strategies. Consequently, learners could eventually apply them in a more automated way, thus freeing cognitive capacity for mental model construction. As a result, also learners with low prerequisite knowledge could benefit from the modeling examples.

Based on the findings presented here, EMME seem to be best suited for learners with stronger cognitive prerequisites; as I discussed above, an improved adaptive system might be helpful especially for learners with weaker prerequisites. It would be interesting to conduct research that directly compares EMME with the adaptive system and also takes learner prerequisites into account. Possibly, the two approaches are helpful for different groups of learners – the EMME for learners with stronger, and the adaptive system for learners with weaker cognitive prerequisites.

In addition, the findings from Mason et al. (2015b) suggest that reading comprehension moderates the effects of EMME differently than pre-existing knowledge. As discussed in section 2.5, other cognitive variables influence multimedia learning, such as working memory (Schüler et al., 2011) or spatial ability (Hegarty & Kriz, 2008). Future research on both EMME and the adaptive system should include these variables to investigate for which learners external vs. self-regulation support is helpful.

## 4.6 Theoretical and practical implications

The findings presented here are in line with previous findings on adequate processing of multimedia instruction. In Experiment 1, I could show that selection, organization and integration processes are associated with good learning outcome. Furthermore, supporting said processes with EMME was helpful to certain learners, which in turn stresses the importance of the processes. In line with CTML, especially integration processes seem to be essential for multimedia learning, since the helpful EMME had a great impact on transitions between text and picture. Therefore, from a theoretical perspective, my findings support the account of multimedia learning as provided by CTML. They are also in line with ITPC (Schnotz & Bannert, 2003) which assumes that information is extracted from text and picture, which on the perceptual level would be reflected on text and picture fixations. ITPC also postulates processes of coherence formation whereby information from text and pictures is mapped and integrated with prior knowledge; on the perceptual level, these processes relate to transitions between the representations.

From a theoretical perspective the findings presented here strengthen the notion that a learner has to actively engage in the learning process, as stated by the active-processing assumption (Mayer, 2009). However, the present findings raise the question what active processing entails. It is not sufficient that learners know about and apply the adequate processes on a perceptual level. Even when a learner shows a viewing behavior consistent with effective processing of multimedia instruction, this does not necessarily result in successful learning (cf. Experiment 4). Likewise, the adaptive system in Experiment 2 was designed to identify insufficient visual processing and to prompt adequate visual processing. However, this intervention failed to result in learning benefits. It is possible that in both cases, learners performed the visual equivalent of the adequate cognitive processes without actively engaging in the processes on a cognitive level. In Experiment 4, copying visual processes was not sufficient to improve learning outcome, but learners needed prerequisite knowledge to conduct meaningful actions. Active processing on a behavioral level, as indicated by adequate visual processing, is not enough – there has to be a cognitive correlate. Further evidence that behavioral change is not sufficient for an intervention to be effective comes from a study by Kriz and Hegarty (2007). They found that signaling in an animation changed visual processing: participants in the signaled condition spent significantly more time on the signaled (relevant) content than learners in the control group. However, this improved visual processing was not mirrored in enhanced learning.

From a practical perspective, these findings raise the question to what extent eye tracking is suitable to infer cognitive processes. In other words, the right pattern of gaze behavior may be necessary, but not sufficient evidence for adequate processing of multimedia instruction. To extend our understanding on cognitive processes during multimedia learning, and on how to support them, additional measures may be necessary. For example, pupil dilation and EEG patterns may provide information on cognitive load during reading (Scharinger, Kammerer, & Gerjets, 2015). These findings might extend to the understanding on multimedia instruction. If that is the case, these variables might be added to the adaptation algorithm to provide a more specific adaptive support. Furthermore, they could be assessed as dependent measures when investigating the effects of EMME and contribute to our understanding on how to best support adequate processing.

Nevertheless, my findings provide evidence that EMME can be successfully applied in multimedia learning, at least for learners with a certain basic understanding of the learning domain. In Experiment 4, the total duration of the intervention was less than five minutes, but had considerable impact, especially on integration processes. This is highly promising from a practical perspective, since EMME can be generated rather easily, and they are short interventions that can be applied in a variety of learning contexts. For example, they could easily be implemented into learning software or e-books, and thus provide a simple approach to supporting multimedia learning.

## 4.7 Concluding remarks

In the work presented here, self-regulation support with EMME was more effective than external support with a gaze based adaptive system. EMME are a very promising instructional tool and future research should definitely follow up on the findings presented here. Regarding the adaptive system, although it was not helpful in its first version, I still think that further research should be conducted to improve the system and explore its possible benefits for multimedia learning.

As my research shows, eye tracking methodology can not only be used to investigate learning processes, but also to support them.

# 5 Summary

Although multimedia is often beneficial to learning, there is evidence that not all learners profit from this kind of instruction. Theoretical accounts of multimedia learning and evidence from eye tracking research suggest that successful multimedia learning is associated with processes of selection, organization and transformation/integration of information from both representations. Individual differences in learning success might be associated with differences in the processing of the instruction.

In this dissertation, I conducted research on two gaze-based interventions that aimed at supporting adequate processing of multimedia instruction. The first intervention, an adaptive system, provides external support: it detects inadequate processing and alters the instruction to prompt adequate processing. The second intervention provides self-regulation support: learners are presented with EMME of adequate processing, which they can then apply in a self-regulated way. I conducted four experiments to investigate if these interventions improve multimedia learning.

My results show that the adaptive system with its current adaptation algorithm does not support multimedia learning. EMME led to more adequate processing for all learners and better learning for learners with strong cognitive prerequisites. These findings suggest that self-regulation support may be a better approach than external support. However, further research is needed to examine if an improved version of the adaptive system can support multimedia learning.

# 6 Zusammenfassung

Illustrierte Texte sind häufig lernförderlich; es gibt allerdings Hinweise darauf, dass Lernende nicht immer davon profitieren. Kognitive Theorien zum multimedialen Lernen und Ergebnisse aus der Blickbewegungsforschung implizieren, dass erfolgreiches Lernen mit Multimedia mit der Selektion, Organisation und Integration von Information aus Text und Bild zusammenhängt. Individuelle Unterschiede im Lernerfolg könnten daher damit zusammenhängen, wie Lernende das multimediale Material verarbeiten.

In der vorliegenden Dissertation untersuchte ich die Wirksamkeit von zwei blickbasierten Interventionen, welche die adäquate Verarbeitung multimedialen Materials fördern sollten. Die erste Intervention, ein blickbasiertes adaptives System, liefert externale Unterstützung. Das System diagnostiziert unzureichende Verarbeitung und passt die Instruktion so an, dass eine bessere Verarbeitung angeregt wird. Die zweite Intervention unterstützt die Selbstregulation. Lernende erhalten eine Blickbewegungsmodellierung (BBM), welche die erfolgreiche Verabeitung des Materials illustriert. Diese erfolgreiche Verabeitung kann dann selbstreguliert angewandt werden. Es wurden vier Studien durchgeführt, um die Wirksamkeit der beiden Interventionen zu überprüfen.

Meine Ergebnisse zeigen, dass das adaptive System mit seinem derzeitigen Algorithmus nicht lernförderlich ist. Die Blickbewegungsmodellierung verbesserte für alle Lerner die Verarbeitung des Materials. Nur für Lernende, die über höheres Vorwissen verfügten, war sie auch lernförderlich. Diese Ergebnisse legen nahe, dass es besser ist, die Selbstregulation zu unterstützen, als externale Unterstützung zu liefern. Es ist jedoch weitere Forschung nötig um zu überprüfen, ob eine verbesserte Version des adaptiven Systems das Lernen mit Multimedia unterstützt.

# 7 References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*, 131–152.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101–130.

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

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia earning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge: Cambridge University Press.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (1996). Multivariate Analysemethoden. Eine anwendungsorientierte Einführung [Multivariate methods of analysis]. Berlin: Springer.

Baddeley, A. (1986). *Working Memory*. Oxford, England: Oxford University Press.

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology, 26*, 863–871.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*, 445–457.

Brooke, J. (1996). SUS - a quick and dirty usability scale. *Usability Evaluation in Industry*, *189*, 4–7.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*, 233–246.

Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, *12*, 371–417.

Corbett, A., Kauffman, L., Maclaren, B., Wagner, A., & Jones, E. (2010). A Cognitive Tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research*, *42*, 219–239.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 377–398.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer Science & Business Media.

Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology*, *27*, 451–461.

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

Fischer, F., Kollar, I., Stegman, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48,* 56–66.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*, 511–525.

Gollwitzer, P. M., & Brandstatter, V. (1997). Implementation intentions and effective goal pursuit. *Joural of Personality and Social Psychology*, *73*, 186–199.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*, 612–618.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, *24*, 95–123.

Hasselhorn, M., & Körkel, J. (1986). Metacognitive versus traditional reading instructions: The mediating role of domain-specific knowledge on children's text-processing. *Human Learning: Journal of Practical Research & Applications*, *5*, 75–90.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis. New York, NY: Guilford*. New York: The Guilford Press.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717–742.

Hegarty, M., & Kriz, S. (2008). Effects of knowledge and spatial ability on learning from animation. In W. Schnotz & R. K. Lowe (Eds.), *Learning with animation: Research implications for design* (pp. 3–29). Cambridge: Cambridge University Press.

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

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176.

Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, *97*, 31–50.

Jameson, A. (1995). Numerical uncertainty management in user and student modeling: An overview of systems and issues. *User Modeling and User-Adapted Interaction*, *5*, 193–251.

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40*, 813–827.

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.

Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, *25*, 62–70.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology. Applied*, *18*, 178–91.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*, 1–17.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*, 126–136.

Koedinger, K. R. (2002). Toward evidence for instructional design principles: Examples from Cognitive Tutor Math 6. In *Proceedings of PME-NA XXXIII (The North American Chapter of the International Group for the Psychology of Mathematics Education)*.

Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge Handbook of The Learning Sciences.* (pp. 61–77). New York: Cambridge University Press.

Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction*, *20*, 424–433.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies*, *65*, 911–930.

Laugksch, R. C., & Spargo, P. E. (1996). Construction of a paper-and-pencil Test of Basic Scientific Literacy based on selected literacy goals recommended by the American Association for the Advancement of Science. *Public Understanding of Science*, *5*, 331–359.

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

Mason, L., Pluchino, P., & Tornatora, M. C. (2015a). Eye movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*, 172–187.

Mason, L., Pluchino, P., & Tornatora, M. C. (2015b). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*. doi:10.1111/bjet.12271

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

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*, 377.

Mayer, R. (2005). *Cognitive theory of multimedia learning*. *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning :

Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 183–200). Cambridge: Cambridge University Press.

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

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, *100*, 380–386.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, *26*, 110–117.

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

Pluchino, P., Tornatora, M. C., & Mason, L. (2013). Improving text and picture integration during reading through eye-movement modeling. Talk presented at the15th Biennal EARLI Conference. Munich, Germany.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search.

*The Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*, 529–556.

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*, 267–280.

Scharinger, C., Kammerer, Y., & Gerjets, P. (2015). Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PloS One*, *10*, e0130608.

Scheiter, K., & Eitel, A. (n.d.). The use of eye tracking as a research and instructional tool in multimedia learning. In C. A. Was, F. J. Sansoti, & B. J. Morris (Eds.), *Eye tracking technology applications in educational research*. Hershey, PA: IGI Global.

Scheiter, K., & Eitel, A. (2010). The effects of signals on learning from text and diagrams: How looking at diagrams earlier and more frequently improves understanding. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (Vol. 6170, pp. 264–270). Heidelberg: Springer.

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, *36*, 11–26.

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

Scheiter, K., Schüler, A., Gerjets, P., Huk, T., & Hesse, F. W. (2014). Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior*, *31*, 73–84.

Scheiter, K., & Van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, *1214*, 1209–1214.

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

Schlag, S., & Ploetzner, R. (2010). Supporting learning from illustrated texts: conceptualizing and evaluating a learning strategy. *Instructional Science*, *39*, 921–937.

Schmidt, H., Wassermann, B., & Zimmermann, G. (2014). An adaptive and adaptable learning platform with real- time eye-tracking support : Lessons learned. In S. Trahash, R. Ploetzner, G. Schneider, C. Gayer, D. Sassiat, & N. Wöhrle (Eds.), *Tagungsband DeLFI 2014* (pp. 241–252). Bonn, Germany: Köölen Druck & Verlag GmbH.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*, 100–110.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology*, *24*, 226–237.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*, 101–120.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156.

Schüler, 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*, 389–411.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, *23*, 1227–1243.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*, 227–237.

Seufert, T., Brünken, R., & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen. *Zeitschrift Für Pädagogische Psychologie*, *19*, 61–75.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In *Adaptive technologies for training and education* (pp. 7–27). New York: Cambridge University Press.

Skuballa, I. T., Fortunski, C., & Renk, A. (2015). An eye movement pre-training fosters the comprehension of processes and functions in technical systems. *Frontiers in Psychology, 6*, 1–13.

Stalbovs, K., Scheiter, K., & Gerjets, P. (2015). Implementation intentions during multimedia learning: Using if-then plans to facilitate cognitive processing. *Learning and Instruction*, *35*, 1–15.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*, 257–285.

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.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, *20*, 95–99.

Veenman, M. V. J., Van Hout-Wolters, B., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3–14.

Wassermann, B., Hardt, A., & Zimmermann, G. (2012). Generic gaze interaction events for web browsers: Using the eye tracker as input device. In *WWW2012 Workshop: Emerging Web Technologies, Facing the Future of Education*.

Webb, G. I., Pazzani, M. J., & Billsus, D. (2001). Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, *11*, 19–29.

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, *23*, 1–21.

Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement: Theory, research, and practice.* New York: Springer.

# 8 List of Figures

# 9 List of Tables