Signaling Text-Picture Relations in Multimedia Learning: The Influence of Prior Knowledge

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

vorgelegt von
Juliane Richter, M.Sc.
aus Hohenmölsen

Tübingen
2016

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

Tag der mündlichen Qualifikation:	26.01.2017
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Katharina Scheiter
2. Berichterstatter:	Prof. Dr. Stephan Schwan

Acknowledgements

I want to thank everyone who has supported me during my PhD project by discussing study designs, analyses and results and supporting me with the present project. Moreover, I want to thank the IWM and in particular the Multiple Representations Lab as well as the LEAD Graduate School and Research Network for offering me an outstanding research environment with great opportunities.

Special thanks goes to

- my supervisors
  - Prof. Dr. Katharina Scheiter who did a great job supervising me
  - Dr. Alexander Eitel
- my colleagues and all student assistants at the Multiple Representations Lab who supported me with my studies whereby I would like to particularly thank
  - Emely Hoch who was always there for in-depth discussions and encouraged and motivated me a lot
  - Horst Mesch who offered great support with data collection in schools
  - Dr. Johann Jacoby for his methodological support and helpful discussions
- Dr. Martin Merkt and Prof. Dr. Stephan Schwan for helpful feedback
- my family and friends.

I also want to thank the awards committee of the Society for Text and Discourse and the EARLI Special Interest Group 2 for awarding me prizes for my studies. I am very grateful for these recognitions of my work, which motivated me a lot to continue my research.

Table of Contents

1. Introduction

Students increasingly learn from digital materials like digital textbooks, open educational resources, e-learning environments or educational apps. This trend of using information and communication technology (ICT) such as tablets in classrooms continues all around the world. The latest IEA International Computer and Information Literacy Study (ICILS) 2013 investigated ICT use in schools and digital competences in 21 education systems including 60,000 eighth graders across the world. They reported that in general 54% of students use computers at least once a week at school mainly with software like word processors, presentation software, and computer-based information resources for preparing reports, essays, or presentations, working with other students, and completing worksheets or exercises (Fraillon, Ainley, Schulz, Friedman, & Gebhardt, 2014). Thus, the main use case for digital media in schools so far seems to be related to writing texts and creating presentations. However, the educational potential of digital devices like tablets is much greater because digital learning materials like digital textbooks or learning apps can be used in classrooms that might adequately support learning processes (cf. Clark & Luckin, 2013).

In line with this notion, the Federal Ministry of Education and Research of Germany just announced to spend five billion euros on digital equipment such as Wi-Fi and hardware for schools all over Germany. When presenting the Digital Pact German Education Minister Johanna Wanka said: “Good education in the 21st century includes IT knowledge and confident handling of technology and of risks of digital communication, as well as learning through the many new possibilities of digital media“. At the same time, Johanna Wanka asked for pedagogical concepts regarding technology use in teaching and teacher training in media didactics to be provided by the states of Germany, which are responsible for school policy (Bundesministerium für Bildung und Forschung, 2016). These demands and the Digital Pact are in line with a report on media education in German schools with recommended actions by the Initiative D21 association (Initiative D21, 2014). They proposed a model for teaching digital media competencies at schools. They identified three basic fields of action: (a) embedding the teaching of digital media competencies in the curriculum, (b) media pedagogical teacher education during studies as well as in pre-service and in-service training, and (c) infrastructure such as hard- and software, digital learning materials, and administration/support of digital learning environments. Only if targeted measures related to all three fields interlock, meaningful teaching of digital media competencies at schools that on the one hand aim at learning with digital media but on the other hand also at learning about digital media (e.g., risks of digital communication) are possible (cf. Initiative D21, 2014). The ‘Digital Pact’ mainly focuses on providing the infrastructure of German schools with regard to hardware, while the states are supposed to adapt their curricula and teacher education accordingly and develop digital learning materials.

Along with the requirement of integrating digital instructional materials into education one of the related pivotal questions is: How does digital instructional content need to be designed in order to foster meaningful learning? A closer look at learning apps and digital textbooks at least in the field of science reveals that these digital materials typically include multimedia, which refers to the presentation of texts and pictures. Consequently, among other things, evidence from research on multimedia instructional design can provide useful information on the design of digital learning materials that foster meaningful learning.

Multimedia materials contain either spoken or written text and static or dynamic visualizations such as simulations, videos, and animations. For example, digital science textbooks present models of scientific phenomena on the microscopic level by means of text accompanied by dynamic visualizations like simulations or sequences of static pictures. Importantly, in a large number of studies multimedia has been found to be more beneficial for learning than text alone (cf. *multimedia principle*, Mayer, 2009; Mayer, 2014a). However, meaningful learning with multimedia is assumed to occur only if corresponding information from texts and pictures is integrated into a coherent integrated mental model containing information from text and picture (e.g., Mayer, 2014b). Hence, learning with multimedia can be challenging because learners need to process not only one external representation (only text), but a combination of at least two different external representations (text and picture), and relate information from these representations to each other (e.g., Renkl & Scheiter, 2015). In order to support students in this effort, research recommended instructional support measures for multimedia learning (Mayer, 2014a).

One of the recommended instructional support measures for multimedia materials is to explicitly highlight correspondences between representations by means of signals such as using the same colors for corresponding elements in text and picture (color coding, see Figure 1), which are supposed to support the integration process and hence meaningful learning with multimedia (*signaling principle*; van Gog, 2014). In the remainder of this thesis, this type of signals will be referred to as *multimedia integration signals (MIS)* to distinguish them from signals that are used to highlight important information within text only (text signals such as words printed in bold face or italics; cf. Lemarié, Lorch, Eyrolle, & Virbel, 2008).

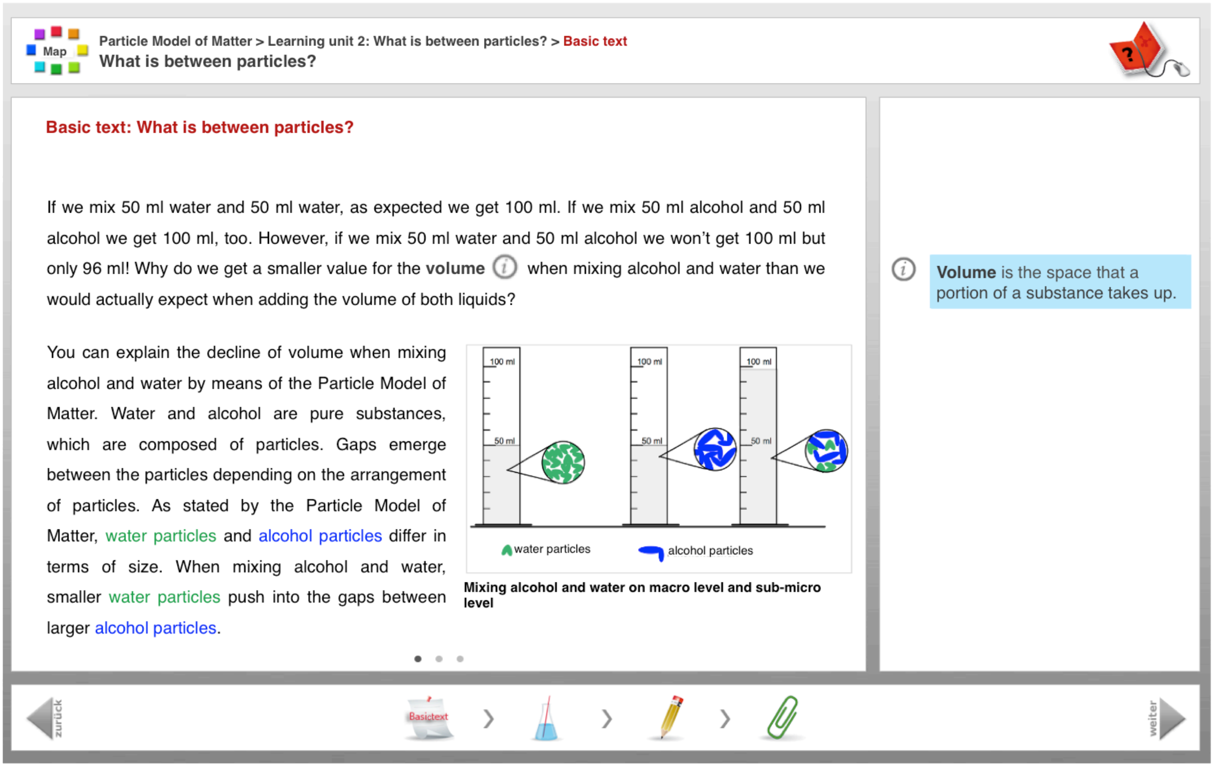


Figure 1. Example of the MIS color coding on a page of a digital textbook for chemistry education. The terms water particles and alcohol particles are displayed in the same color as in the related picture on the right-hand side.

However, although MIS are expected to support multimedia learning a comprehensive meta-analysis is lacking. Thus, it is an open question how large the multimedia signaling effect is and whether it is affected by boundary conditions. This question is corroborated by evidence suggesting that instructional techniques such as multimedia instructional design measures might not be effective for learning in general. The *expertise reversal effect* (ERE; Kalyuga, 2014; Kalyuga, Ayres, Chandler, & Sweller, 2003) states that the effectiveness of instructional support depends on the domain-specific prior knowledge of learners. Based on the ERE, learners with low domain-specific prior knowledge (LPK) should profit from instructional support whereas learners with high domain-specific prior knowledge (HPK) are supposed not to profit or are even hindered in learning. Importantly, the latter scenario would be a reason for concern if it were found to be the case related to the effectiveness of MIS. Nowadays, all students in a school class are mostly provided with the same learning materials such as textbooks. However, classes are expected to be among other things heterogeneous with respect to domain-specific prior knowledge related to a particular topic (e.g., Slavin, 1987). Against the backdrop of the ERE, educators and publishers of learning materials would face an ethical conflict when deliberately providing a certain type of instructional support to all learners in a class if HPK students might be hindered in learning. Thus, research is needed to clarify under which conditions an instructional support measure such as multimedia signaling is effective for learning. Moreover, the digitalization of education can be an opportunity to address potential individual differences between learners by providing digital learning material that adapts to the individual needs of each student at each time during his or her learning process by means of learner-tailored instructions (cf. Kalyuga, 2007).

The present dissertation seeks to shed light on one aspect of multimedia instructional design: the effectiveness of MIS for learning related to learners' level of domain-specific prior knowledge. In order to systematically investigate the validity of the signaling principle in multimedia learning, first, a comprehensive meta-analysis on multimedia learning studies was conducted (Study 1). The meta-analysis compared performance of a group learning from signaled multimedia material with that of a control group. In doing so, it was investigated whether there is a significant positive effect of MIS in multimedia learning, and if so, how large this effect is. Secondly, the meta-analysis aimed at assessing for whom and under which conditions (e.g., different levels of domain-specific prior knowledge) MIS yield positive effects (chapter 6). In order to more thoroughly investigate the influence of domain-specific prior knowledge on the multimedia signaling effect in a more ecologically valid context than used in studies included in the meta-analysis, secondly an experimental field study was conducted with eighth graders in schools

(Study 2). They learned with a digital textbook about a topic from the curriculum containing either mainly text signals or additional MIS (chapter 7). Third, processes underlying a potential ERE related to multimedia signaling were investigated by means of eye tracking in a lab study with students in grade seven to nine in secondary higher education (Study 3). They learned with part of the digital textbook used in Study 2, which again contained either mainly text signals or additional MIS. During learning their eye movements were recorded (chapter 8). Overall, results contribute to answers to one of the key questions related to the use of digital devices in schools, namely, how digital instructional content needs to be designed in order to foster meaningful learning.

To begin with, the theoretical background related to multimedia learning, the signaling effect in multimedia learning, the influence of domain-specific prior knowledge and potential underlying processes and process measures will be described. Then an overview and the five overall research questions related to the present thesis will be reported in chapter 5. Results of the three studies will be discussed within the related chapters and summarized and discussed generally in the light of the five research questions in a comprehensive discussion. Moreover, practical implications as well as strength and limitations of the present dissertation will be outlined.

2. Learning with Multimedia: The Importance of Integration

The term multimedia refers to the simultaneous presence of verbal and pictorial information – that is (written or spoken) text and (static or dynamic) visualizations. When both formats are used together in instructional material this is referred to as a multimedia instructional message (Mayer, 2014b). Evidence strongly suggests that people learn more deeply from verbal and pictorial information than from verbal information alone (*multimedia principle*; Butcher, 2014; Mayer, 2009; Mayer, 2014a). An influential theory that describes the underlying processes of multimedia learning is the *cognitive theory of multimedia learning* (CTML; Mayer, 2009; Mayer, 2014b) (Figure 2).

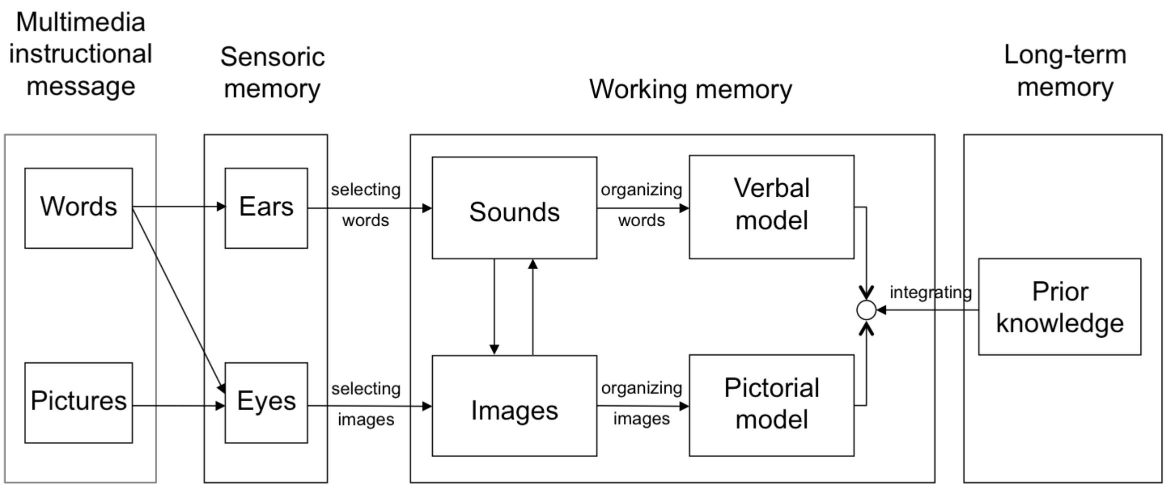


Figure 2. Cognitive theory of multimedia learning (adapted from Mayer, 2014b, p. 52).

The CTML is based on three assumptions related to the human cognitive system (Mayer, 2014b). First, according to the dual-channel assumption, which is related to Paivio's dual-coding theory (Paivio, 1991), there are two different channels for information processing for (a) visual and (b) auditory/verbal representations. Secondly, these channels are limited with respect to their capacity for processing and transferring information (limited-capacity assumption). The third assumption is that learners actively process information by selecting, organizing and integrating information into coherent mental representations (cf. Mayer, 2014b). The CTML defines these processes in more detail: Selection of (a) verbal and (b) pictorial information, organization of (c) verbal information into a coherent verbal mental model and of (d) pictorial information into a coherent pictorial mental model in working memory, and (e) integration of both models into a coherent mental model by referencing to prior knowledge in long-term memory (integrated mental model construction, cf. Johnson-Laird, 1983). Thus, according to the CTML verbal and pictorial information is selected and organized in separate mode-specific models in working memory that are then integrated with each other and with prior knowledge in a downstream process step. The integration process is “perhaps the most crucial step in multimedia learning“ (Mayer, 2014b, p. 57) since it is assumed to be necessary for meaningful learning to occur (Mayer, 1997; Mayer, 2008).

An alternative theory was proposed by Schnotz (2014): the *integrative model of text and picture comprehension* (ITPC) (Figure 3).

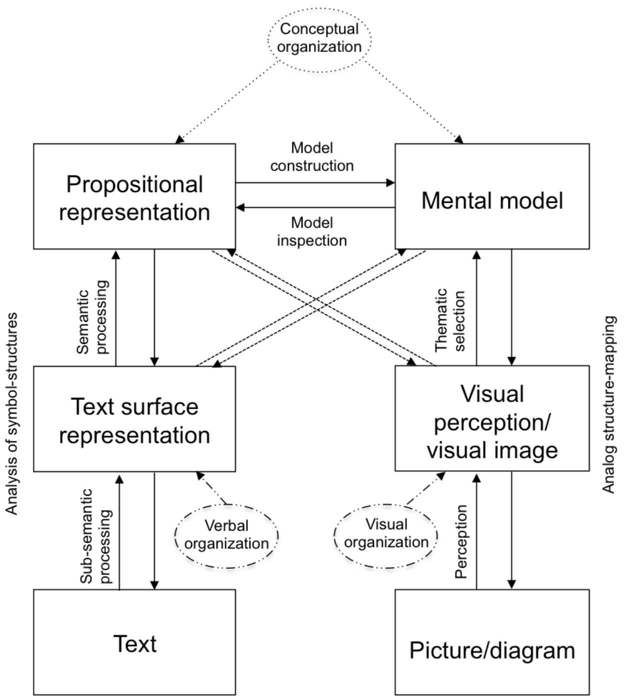


Figure 3. Theoretical framework for the integrative model of text and picture comprehension (Schnotz & Bannert, 2003, p. 145).

The ITCP is similar to the CTML regarding the dual channel assumption for verbal and pictorial information in that an auditory and a visual register initially process and transmit verbal and pictorial information. However, in contrast to the CTML the ITPC proposes interactions between verbal and pictorial information processing directly at the time this information is transferred to subsystems in working memory (indicated by arrows pointing in the opposite direction, Figure 3). The ITPC proposes that students have to establish coherence regarding the information within each representation (cf. *intra-representational coherence formation*; Seufert, 2003) as well as between verbal and pictorial representations (cf. *inter-representational coherence formation*; Seufert, 2003) each at the surface level and the level of semantic deep structures (Schnotz et al., 2014). Surface structure mapping is the process of connecting elements of verbal (e.g., words) and pictorial representations (e.g., shapes) whereas semantic deep structure mapping includes the establishment of connections between conceptual structures and characteristics of these structures included in the mental model (e.g., simple/complex relations between elements) (cf. Schnotz et al., 2014). Therefore, in the ITPC model all representations in working memory interact with each other, whereas the CTML proposes that the verbal and pictorial model do not interact prior to the final step of integration aimed at building a coherent mental model.

According to both theories, the integration of verbal and pictorial information is crucial for the creation of a coherent mental model that underlies meaningful learning with multimedia. To be more specific, Mayer (2014b) states that the presence of a coherent mental model is particularly reflected in deep measures of learning such as comprehension and transfer performance.

Empirical evidence supports this notion in that intensive integrative processing of verbal and pictorial information has been found to be clearly linked to better learning outcomes (e.g., Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Mason, Tornatora, & Pluchino, 2013; Seufert, 2003). Bodemer et al. (2004) compared three groups learning about statistical concepts with (a) spatially integrated text and pictures, (b) spatially separated text and picture, or (c) a version that required learners to actively map text to elements in the picture. Results revealed that learners who actively integrated verbal and pictorial information by means of mapping text to pictorial elements showed better learning performance than learners learning with the other versions. Thus, the construction of an integrated format containing verbal and pictorial information fostered learning, which is an indicator for the importance of the integration process for learning (Bodemer et al., 2004). Mason, Tornatora, and

Pluchino (2013) asked fourth graders to learn with an illustrated science text about the characteristics and a phenomenon related to air while their eye movements were recorded. The authors aimed at measuring integrative processing as reflected by more transitions between text and picture and longer fixations of the picture during re-reading the text as well as also longer fixations of the text during re-inspecting the picture. Results revealed that more integrative processing was related to the highest scores for factual and transfer knowledge. Furthermore, the least integrative processing behavior was related to the lowest scores for learning outcomes. Hence, results obtained by Mason, Tornatora, and Pluchino (2013) underline the crucial role of integration for successful multimedia learning.

However, learners often fail to establish coherent mental models by integrating verbal and pictorial information because the cognitive demands resulting from this process are too high (Renkl & Scheiter, 2015). Conversely, learners rely more strongly on information provided by text rather than picture (e.g., Hannus & Hyönä, 1999; Hegarty & Just, 1993; Schmidt-Weigand, Kohnert, & Glowalla, 2010). In addition, as shown by Mason, Tornatora, and Pluchino (2013) some learners show only infrequent attempts to integrate information from text and picture (cf. low integrators).

Therefore, learners may need extra support in identifying and mapping related elements in texts and pictures by means of instructional techniques in order to enable meaningful multimedia learning (Mayer & Moreno, 2003; Seufert, 2003). Research has provided evidence for the effectiveness of different instructional support measures for multimedia material that aim at supporting learners in their effort to integrate verbal and pictorial information into a coherent integrated mental model. The split-attention principle for example suggests to physically and temporally integrate related texts and pictures, which is supposed to lead to better learning than separate formats (Ayres & Sweller, 2014). Hence, verbal information that is necessary to understand a visualization should be located close to the visualization. In case of dynamic visualizations (e.g., animations) the related auditory verbal information should be presented in a timely manner related to the occurrence of related visual elements. This support measure should prevent learners from having to split their attention between multiple verbal and pictorial sources of relevant information, which might result in an increase of cognitive load (Ayres & Sweller, 2014). Another example of a multimedia design measure is the modality principle (Low & Sweller, 2014). The principle suggests that working memory load is reduced due to the presentation of corresponding information simultaneously in visual and auditory mode rather than in visual mode only. For example, students learning with an animation with visually presented narration need to switch back and forth between reading the narration and inspecting the animation. The information from both sources need to be processed in the visual channel (Paivio, 1991). In contrast, when animations are accompanied by spoken narration students can attend to the information simultaneously by using the visual *and* the auditory/verbal channel (Paivio, 1991). Thus, using visual stimuli accompanied by spoken narration is supposed to facilitate text-picture integration (Low & Sweller, 2014).

But even if verbal and pictorial information are presented in an integrated manner and in both visual and auditory mode, learners might still have difficulties in identifying corresponding elements in texts and pictures relevant for the integration into a coherent mental model. An instructional support measure that aims at supporting learners in this effort is signaling corresponding verbal and pictorial information by means of discursive or visual highlights (Van Gog, 2014). The signaling effect for multimedia learning will be presented in detail in the next chapter.

3. The Signaling Effect

The present chapter focuses signaling text-picture correspondences as an instructional measure to support multimedia learning. First, a broad definition of signaling in the context of learning with text as well as learning with multimedia will be given. Second, assumptions about how multimedia signaling works will be reported along with empirical evidence regarding learning outcomes and eye tracking measures. In a third subchapter, potential material-based boundary conditions regarding the multimedia signaling effect such as the pictorial format of visualizations will be described.

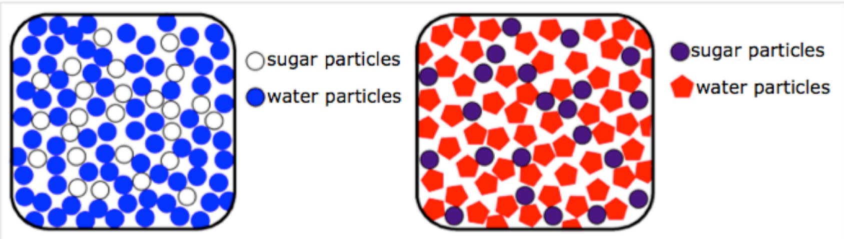
3.1 What is Signaling?

Signaling is basically highlighting of relevant information with the aim to foster comprehension of the materials. Importantly, signals in general serve as instructional elements that can be implemented and removed from materials without altering their contents (Lorch, 1989). One has to distinguish between signals that serve to support (a) text comprehension and (b) learning with multimedia. In the following, both types of signals will be reported.

The notion to support students in comprehension of contents by highlighting the organization of materials was initially implemented in text comprehension research. Text signals such as headings, preview/summary sentences, paragraphs, or bold face were used to support text comprehension (e.g., Loman & Mayer, 1983; Lorch, 1989; Lorch & Lorch, 1995). *Text signaling* is supposed to support learners particularly in selecting and organizing verbal and pictorial information into mode-specific mental models. These processes are assumed to be reflected by a positive effect of text signals particularly on recall performance (Lorch, 1989; Lorch & Lorch, 1995; Mautone & Mayer, 2001). In the text signaling theory *SARA* (*Signal Available Relevant Accessible Information*) Lemarié et al. (2008) characterized text signals along two dimensions: information functions and realization properties. Information functions are information about how to process materials that a signal communicates to the reader. For example, a heading may give information about the structure of the text by demarcating and identifying the function and topic of a particular part of the text. Realization properties describe the actual appearance of a signal as being either discursive or rather visual. Hence, an information function can be communicated by a discursive or rather visual text signal. For example, the information function of emphasizing relevant text elements can be realized discursively by beginning a sentence with "It is important to note…" or visually by highlighting the sentence by means of bold face. Both types of signals convey importance to the reader; however, they are realized in two different ways (cf. Lemarié et al., 2008).

As stated in the preceding chapter, in the case of multimedia learning it is not sufficient to establish mode-specific models separately from the text and the picture to learn successfully. Importantly, learners have to integrate verbal and pictorial information into a coherent mental representation in order to learn successfully. *Multimedia signaling* serves to highlight relevant corresponding information in texts and pictures aiming at supporting multimedia learning (signaling principle; van Gog, 2014). Within this thesis, signals or cues (these terms are used synonymously; cf. de Koning, Tabbers, Rikers, & Paas, 2009) that serve to specifically support the integration process of verbal and pictorial information by highlighting text-picture correspondences are referred to as multimedia integration signals (MIS). Since a comprehensive coherent mental representation is supposed to reflect more elaborate knowledge (Mayer, 2014b), the effectiveness of MIS should be reflected primarily by improved comprehension and transfer performance (mapping of knowledge to a different context) (cf. Mautone & Mayer, 2001).

Examples for MIS are deictic references in the text referring to elements in a picture (see example in Figure 4: "In the visualization on the left you can see…"), and color coding of corresponding elements in text and picture (see example in Figure 5). A variation of color coding for dynamic visualizations are highlights such as colored labels or spotlights on elements in the picture presented synchronously with the occurrence of the related term in a spoken narration (cf. Boucheix & Guignard, 2005; Ozcelik, Arslan-Ari, & Cagiltay, 2010; Tabbers, Martens, & Merriënboer, 2004). Finally, words from the text can be used as labels in the picture that help to identify which term in the text relates to which pictorial element in the picture. Hence, MIS can be located either in the text (like a deictic reference), in the illustration (like a spotlight, e.g., a red circle around an element in the picture that is referred to in a narration) or in both types of representations like for example color coding of corresponding aspects.

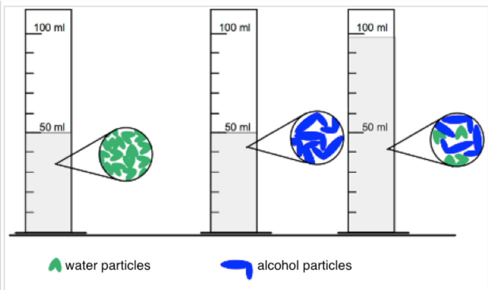


Different visualizations of a sugar solution on sub-micro level

Both visualizations depict exactly the same: a sugar solution. In the visualization on the left you can see water particles depicted as blue circles and sugar particles as white circles. The visualization on the right however depicts water particles as red pentagons and sugar particles as purple circles. The colors and shapes of the particles serve only for a better distinction. **The particles themselves do not have a color and the Particle Model of Matter gives no information about the shape.**

Figure 4. Example of the MIS deictic reference in the text: “In the visualization on the left you can see…”.

You can explain the decline of volume when mixing alcohol and water by means of the Particle Model of Matter. Water and alcohol are pure substances, which are composed of particles. Gaps emerge between the particles depending on the arrangement of particles. As stated by the Particle Model of Matter, water particles and alcohol particles differ in terms of size. When mixing alcohol and water, smaller water particles push into the gaps between larger alcohol particles.



Mixing alcohol and water on macro level and sub-micro level

Figure 5. Example of the MIS color coding (highlighting corresponding elements in text and picture in the same color).

Similar as for text signals, the function of MIS can be conveyed discursively or visually to the learner. For example, text-picture integration can be supported by the discursive MIS deictic reference referring from the text to the picture with “As you can see in the picture on the right-hand side element x is…” or visually by means of color coding.

3.2 How does Multimedia Signaling Work?

There are a number of studies showing that MIS may improve meaningful learning from multimedia reflected by comprehension and transfer performance (e.g., Mautone & Mayer, 2001; Ozcelik et al., 2010; Scheiter & Eitel, 2015). For instance, Ozcelik et al. (2010) presented learners with narration and a labeled illustration about how a turbofan jet engine works. When a term was mentioned in the auditory narration the corresponding label in the illustration of the turbofan jet engine became red in order to signal the relation between narration and illustration. Results showed that learners with color coded materials outperformed learners receiving no signals with regard to matching and transfer performance. In a study by Mautone and Mayer (2001, Exp. 3) students learned about how airplanes achieve lift with a narrated animation either with or without MIS. The group learning with MIS included in the material outperformed the group learning without MIS with regard to transfer performance; however, these groups did not differ with respect to recall performance.

The underlying cognitive processes of the effectiveness of MIS related to learning outcomes were investigated by means of eye tracking methodology. The basic underlying assumption related to recording eye movements during learning is that visual attention devoted to materials provides information about concurrent cognitive processes (*eye-mind assumption*; Just & Carpenter, 1980). Evidence suggests two different hypotheses related to visual attention during multimedia learning with MIS: (a) *guiding-attention hypothesis* (Ozcelik et al., 2010), and (b) *unnecessary visual-search hypothesis* (Jamet, 2014; Ozcelik et al., 2010).

The guiding-attention hypothesis states that MIS guide attention to highlighted information, which increases the visual attention and thus cognitive processing of this particular information. Eye tracking parameters that reflect an increase in attention to signaled information are an overall longer fixation (time) of these elements as well as more fixations (fixation count) on these elements in general. Accordingly, also the average fixation duration (fixation time divided by fixation counts) should increase on signaled elements. Numerous studies showed the guiding function of signaling on attention (Boucheix & Lowe, 2010; de Koning, Tabbers, Rikers, & Paas, 2010a; Kriz & Hegarty, 2007; Ozcelik et al., 2010; Scheiter & Eitel, 2015).

The visual-search hypothesis suggests that multimedia signaling reduces visual search, which becomes evident by an earlier fixation of highlighted elements. However, although some studies confirmed the visual-search hypothesis (Ozcelik, Karakus, Kursun, & Cagiltay, 2009; Ozcelik et al., 2010; Scheiter & Eitel, 2015), de Koning et al. (2010a) did not find effects of signaling on visual search. One of the reasons for the divergent findings might be the types of signals used. Ozcelik et al. (2010) used color to highlight elements in an illustration when they were mentioned in a spoken narration. The signals used by de Koning et al. (2010a) did not highlight specific elements in the animation but rather highlighted the area of the display containing relevant elements. De Koning et al. (2010a) concluded that especially LPK learners were probably not sufficiently guided by these types of signals. Thus, their visual search was not reduced in this study (De Koning et al., 2010a).

Evidence for the guiding-attention and unnecessary visual-search hypotheses was corroborated by a mediation analysis conducted by Scheiter and Eitel (2015). This analysis was motivated by the observation that eye tracking and learning outcome measures in previous studies were analyzed separately only yielding positive effects of signaling for both types of measures (e.g., Boucheix & Lowe, 2010; Ozcelik et al., 2009; Ozcelik et al., 2010). This way of analysis left open whether changes in eye movements would be suited to explain differences in learning outcomes. Moreover, there were also studies revealing positive signaling effects only for eye tracking measures while lacking effects on learning outcomes (e.g., de Koning et al., 2010a; Kriz & Hegarty, 2007). This pattern sheds doubt on changes in visual attention being responsible for differences in learning outcomes. Therefore, Scheiter and Eitel (2015) tested a mediation hypothesis regarding visual attention measures. They implemented several MIS into instructional material about the functioning of the human circulatory system, which was comprised of texts and diagrams. Their signaling condition included deictic references, corresponding labels, and color coding of related elements. Learners in the signaling condition showed better performance in a text-diagram integration task. Moreover, they showed that fixating signaled information more frequently and earlier during learning explained better performance in the text-diagram integration task for the group learning with MIS. This pattern of results corroborated the guiding-attention (more fixations) and unnecessary visual-search (earlier fixation) hypotheses (cf. Ozcelik et al., 2010).

A measure that is supposed to be more sensitive for detecting the main function of MIS namely to support the integration of verbal and pictorial information into a coherent mental representation is the number of transitions (also saccades) between text and picture. Transitions are the number of shifts between fixations of the text and the picture summed up to a total number of shifts between texts and pictures (cf. Mason, Tornatora, & Pluchino, 2013; Scheiter & Eitel, 2015). Johnson and Mayer (2012) suggested that a greater number of transitions reflects more intensive attempts to integrate verbal and pictorial information into a coherent integrated mental model, which is corroborated by studies showing that the number of transitions are related to better learning outcomes (Mason, Tornatora, & Pluchino, 2013; O'Keefe, Letourneau, Homer, Schwartz, & Plass, 2014). Consequently, MIS should increase the number of transitions, because these signals are supposed to mainly foster the integration process of verbal and pictorial information.

After having defined MIS and the way they are assumed to work, potential boundary conditions of the multimedia signaling effect will be reported in the next chapters. In doing so, I will distinguish between boundary conditions related to the materials (e.g., pictorial format of visualizations and pacing of the materials) and the domain-specific prior knowledge related to the learners. The effectiveness of multimedia signaling in general and the way it is influenced by potential material-based boundary conditions as well as the domain-specific prior knowledge will be subject of the meta-analysis in chapter 6 (Study 1). The results of the meta-analysis revealed that domain-specific prior knowledge plays an important role regarding the multimedia signaling effect. Thus, after having reported several material-based boundary conditions in the following, chapter 4 will focus on the learner-based boundary condition domain-specific prior knowledge exclusively.

3.3 Material-based Boundary Conditions of the Multimedia Signaling Effect

Against the backdrop of theories such as CTML (Mayer, 2014b) and SARA (Lemarié et al., 2008), MIS can be assumed to be more or less beneficial for multimedia learning, depending on the design of instructional materials, and the experimental procedure. Referring to the existing literature on multimedia signaling in learning situations, four potential material-based boundary conditions were derived that will be described in the present chapter: (a) pacing of the materials, (b) pictorial format, (c) multimedia mapping requirements, and (d) distinctiveness of MIS. These boundary conditions were considered as potential moderators of the multimedia signaling effect in Study 1 (chapter 6).

Pacing of the materials. Digital learning material can be presented in either system-paced or self-paced formats. A presentation paced by the system provides no options to the learner to control information delivery (e.g., start, stop, pause, go forward and backward), whereas a self-paced presentation enables interactivity. System-paced learning makes it necessary for learners to attend to relevant information at the right time, since otherwise they might miss important information. On the other hand, students who learn in a self-paced manner can control what information to attend to at their own pace. Also, they are able to go back in the material and restudy the given information, thereby clarifying possible misunderstandings or gaps in their knowledge. As a consequence, user pacing usually leads to increases in learning time (Boucheix & Schneider, 2009; Kriz & Hegarty, 2007; Tabbers & de Koeijer, 2010), which might also be associated with improvements in performance (Boucheix & Schneider, 2009; Mayer & Chandler, 2001; Tabbers & de Koeijer, 2010). Moreover, Schmidt-Weigand et al. (2010) contrasted different system-controlled presentation speeds for learning with multimedia and found that when given more time for learning, students invested this time in intensifying their viewing of the animation and integrating it with text, as revealed by their eye movements.

Against the backdrop of these studies, MIS might be more beneficial under system control, in which students are more pressured and are likely to not conduct the right cognitive processes within the allotted time. MIS ensure that learners attend to the relevant information at the right time and that they are able to quickly identify corresponding text-picture elements without engaging in extensive visual search, as has been evidenced in various eye tracking studies (e.g., Jamet, 2014; Ozcelik et al., 2010; Scheiter & Eitel, 2015). Consequently, MIS should enable learners to successfully integrate texts and pictures especially when efficient processing is a necessity, as is the case in system-paced instruction. On the other hand, if they have or can take more time (i.e., in the case of self-paced learning or slower presentation speeds) they are more likely to integrate texts and pictures even without guidance from signals, as suggested by Schmidt-Weigand et al. (2010). Hence, MIS should have a smaller or no effect in the case of self-paced learning.

An analogous pattern of results was found for the modality effect (Ginns, 2005; Tabbers et al., 2004). According to the modality effect, in general students learn better with spoken text and (static/dynamic) pictures than with written texts and pictures. This is because listening to spoken text allows attending to texts and pictures simultaneously, whereas in the case of written text learners need to switch back and forth between reading the text and inspecting the picture (cf. Low & Sweller, 2014). Identifying correspondences between texts and pictures may thus be particularly difficult in the case of written text when there is little time to process the multimedia materials, thereby revealing a stronger modality effect, whereas text modality should be less important when there are no time constraints. In line with this reasoning, Ginns (2005) used pacing of the material as a moderator in his meta-analysis on the modality effect and showed that the modality effect was larger for system-paced in contrast to self-paced learning materials. Tabbers et al. (2004) even found a reverse modality effect for self-paced presentation of instructional material.

Pictorial format. Based on the literature, opposing assumptions can be derived as to whether multimedia signaling works better for text with static pictures or for text with dynamic visualizations (e.g., animation, video).

On the one hand, Kühl, Scheiter, and Gerjets (2012) hypothesized that signaling is more effective for dynamic rather than static visualizations, because visual complexity is higher in dynamic visualizations than in static visualizations. Additional temporal relations associated with the movement of elements need to be processed. Because MIS such as spotlights, color coding and zooming emphasize important aspects and organize dynamic stimuli, they support learners in processing transient information. Therefore, MIS were expected to facilitate the coherence formation processes especially during learning from dynamic stimuli. However, the authors did not find an interaction effect between the type of visualization (static versus dynamic) and signaling and could thus not confirm their hypothesis.

On the other hand, de Koning et al. (2009) argue that signaling might be less helpful for dynamic than for static presentations, because the salience of signals might decline in the context of moving or flashing elements within animations. These moving or flashing elements already require a great deal of visual attention. In this competition for attention, learners might therefore not notice the presence of signals. In their review of 13 studies, de Koning et al. (2009) found mixed evidence for the effectiveness of signaling in dynamic visualizations and concluded that the type of signals used determined its effectiveness for learning outcomes. They stated that signaling measures that are effective with static pictorial instructions do not necessarily improve learning from instructional animations. Against the backdrop of the work by Kühl et al. (2012) and de Koning et al. (2009), one can thus assume signaling to have no or even harmful effects in the case of dynamic visualizations, whereas positive signaling effects have been well documented for static pictures (e.g., Jamet, 2014; Ozcelik et al., 2010; Scheiter & Eitel, 2015). In contrast, Höffler and Leutner (2007) found no moderating effect of signaling in a meta-analysis on static pictures and instructional animations.

However, since the review by de Koning et al. (2009) and the meta-analysis by Höffler and Leutner (2007) were published, there have been attempts to design signals that are optimized for use with dynamic visualizations (e.g., Boucheix & Lowe, 2010). Hence, one can expect signaling to have more pronounced effects in dynamic rather than static visualizations when including these newer studies. The signals used in previous studies all have in common that they highlight individual elements involved in the process at a given point in time. However, they do not emphasize how changes regarding these elements are interlinked and contribute to the causal chain of events, which is a major learning goal in the comprehension of events. To counteract this limitation, Boucheix and Lowe (2010) developed spreading color cues, which highlighted how changes are propagated through a series of events within an animation. In other words, the color cues moved through an animation starting for example at element X of a causal system (e.g., a piano mechanism) and spread further to the next element Y that was necessary to understand the underlying process of the mechanism. They showed that spreading cues aided the comprehension of a mechanical system. Accordingly, it might be that more recent studies have used improved signals when studying learning from dynamic visualizations, and that these might be more apt to help learners manage the complexity arising from the transience of these visualizations. Thus, the pattern of results might have changed in favor of stronger effects of signaling for dynamic rather than static visualizations, as had initially been postulated by Kühl et al. (2012).

Multimedia mapping requirements. In order to make sense from a multimedia instructional message learners have to relate corresponding elements within the verbal and pictorial representation. Instructional material may contain few or many mapping requirements depending on the number of elements included in verbal and pictorial representations that need to be mapped to form a coherent mental representation. Moreover, mapping requirements depend on the amount of visual search required for identifying corresponding elements in (non-signaled) multimedia materials. That is, a picture comprised of many visual elements necessarily contains more elements that are irrelevant in the context of identifying one particular text-picture correspondence, as compared with a picture comprised of only few visual elements. Thus, a complex picture contains more distractors, thereby rendering integration of text and picture more difficult.

An example for material containing few multimedia mapping requirements was used in a study by Mason, Pluchino, and Tornatora (2013). Their instructional material was about how the suction cup of a sink plunger works. They used text and a labeled versus non-labeled illustration. The illustration showed three states of the sink plunger that were explained in the text. Therefore, three content elements in the text and illustration had to be mapped by learners (i.e., one text-picture correspondence for each state of the sink plunger). Only two visual elements served as distractors for identifying each of the text-picture correspondences (that is, those elements relating to the two remaining states). An example of multimedia material containing relatively many mapping requirements was used by Florax and Ploetzner (2010). A text and illustration about information processing in the human nervous system was presented either signaled by means of labels or non-signaled. The text described 21 steps of information processing at a non-activated synapse, an excitatory synapse, and an inhibitory synapse, which were also depicted in the illustration. In order to create a coherent mental representation, learners had to map and integrate all 21 process steps, and identify the relevant visual element for each piece of text, with 20 other visual elements being irrelevant in the given context.

Forming a coherent mental representation from a multimedia message that requires extensive visual search for correspondences and that contains many elements that need to be mapped is cognitively more demanding than from a multimedia message including only little visual search and few correspondences. Thus, MIS might be more useful for learners learning from material containing many mapping requirements than for learners learning with material containing only few mapping requirements.

Evidence in line with this assumption was reported by Jeung, Chandler, and Sweller (1997). They manipulated visual search requirements in geometry instructions delivered by auditory narration and a diagram or diagram only, using signaling in the form of flashing elements. When the induced visual search was high, signaling was beneficial for learning outcomes, whereas when visual search requirements were low, signaling did not improve learning outcomes.

Distinctiveness of MIS. Lemarié et al. (2008) postulated that the distinctiveness of signals affects whether a reader accesses signaled information or not. Thus, the accessibility of signaled information may depend on whether a signal conveys its information discursively or visually. They argue that visual signals (e.g., bold face) are more salient than discursive signals (beginning a sentence with “It is important to note that…”) and might therefore make the signaled information more easily accessible to readers in contrast to discursive signals.

MIS can also be presented either visually, for example by means of color coding, or discursively, for instance by means of deictic references. Based on the SARA theory by Lemarié et al. (2008) the assumption can be derived that because discursive signals are less salient than visual signals they might also be less effective for learning, because they can be easily overlooked. Visual MIS such as color coding or zooming change the visual appearance of a multimedia instructional message by making relations between verbal and pictorial information salient, whereas discursive signals such as deictic references and corresponding labels are far less salient. Therefore, the argument made by Lemarié et al. (2008) can also be applied to MIS.

4. Domain-specific Prior Knowledge as a Boundary Condition of the Multimedia Signaling Effect

In the present chapter theory and evidence related to the influence of domain-specific prior knowledge on the effectiveness of instructional techniques in general and specifically on the effectiveness of multimedia signaling will be reported. First, the phenomenon that novice learners profit from an instructional technique whereas expert learners do not - the expertise reversal effect (ERE; Kalyuga et al., 2003) - will be described. Related evidence will be reported. Second, explanatory approaches underlying EREs will be described for instructional techniques in general and specifically related to the effectiveness of MIS. In the third subchapter, measurement approaches regarding cognitive load and visual attention that may provide insight into the processes underlying EREs will be reported.

4.1 The Phenomenon of the Expertise Reversal Effect

The finding that effects of instructional design depend on characteristics of learners is well known from research regarding aptitude-treatment interactions (ATI; Cronbach & Snow, 1977). Evidence from ATI research suggest that an instructional measure (treatment) does not necessarily support learning in general but rather fosters learning performance of students with a particular aptitude, which is a learner characteristic that is predictive for learning success such as interest, attitudes, personal traits, or cognitive ability (cf. Bracht, 1970; Cronbach & Snow, 1977; Shute & Gluck, 1996). In line with this reasoning, Cronbach and Snow (1977) concluded:

> Aptitude x Treatment interactions exist. To assert the opposite is to assert that whichever educational procedure is best for Johnny is best for everyone else in Johnny's school. Even the most commonplace adaption of instruction, such as choosing different books for more and less capable readers of a given age, rests on the assumption of ATI that it seems foolish to challenge. (p. 492)

A variant of the ATI is the ERE (Kalyuga et al., 2003; Schnotz, 2010). The ERE states that the effectiveness of an instructional technique depends on learners' domain-specific prior knowledge (henceforth referred to as prior knowledge). More specifically, EREs reveal that LPK learners profit from an instructional technique whereas HPK learners do not profit with regard to their learning performance.

Numerous studies have revealed EREs related to various instructional techniques (e.g., Homer & Plass, 2009; Kalyuga, Chandler, & Sweller, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996; Nückles, Hübner, Dümer, & Renkl, 2010; Oksa, Kalyuga, & Chandler, 2010). Moreover, Kalyuga (2007) reviewed empirical findings of EREs obtained with different instructional techniques such as labeling, multimedia presentations, or worked examples. What becomes evident from this review is that one should distinguish between situations in which a support measure does not have any beneficial effect for HPK learners (partial reversal) and situations in which it even hampers learning (full reversal), which is important from a practical as well as a theoretical perspective (cf. Kalyuga & Renkl, 2010). Overall, Kalyuga (2007) reported 48 experiments in his review with effect sizes regarding the instructional manipulation for both novice and expert learners as well as the differences between these effect sizes. The majority of the listed experiments showed a full reversal effect (39 out of 48) as indicated by a positive effect size for novices and a negative effect size for experts. Therefore, the review hints towards a full rather than a partial reversal effect for the investigated instructional design features.

When considering the results of Kalyuga (2007), it has to be kept in mind that partial EREs may always also be due to a lack of power regarding the statistical analysis to reveal a disordinal interaction. Moreover, knowledge acquisition is a continuous process that can probably not accurately be reflected by only the two extremes, namely low and high prior knowledge. The state in between those extremes should also be taken into consideration when investigating the effectiveness of instructional techniques: a medium prior knowledge level (MPK). These learners already have established a knowledge base and schemas, although they are not as extensive and automated as for HPK students. Seufert (2003) included participants with differing prior knowledge levels into a signaling study and categorized them into three prior knowledge levels (LPK, MPK, and HPK). She hypothesized that only MPK learners would profit from signaling because signals activate existing knowledge and therefore help MPK students to extend their knowledge base. The results revealed that MPK learners profited from the given signals with respect to learning outcomes, whereas LPK learners did not. Seufert (2003) argued that the provided signals were probably too difficult to use for LPK learners. Therefore, the effectiveness for signals was obtained only for MPK students. HPK students, however, did not profit from the given support, suggesting a partial reversal at the intersection of MPK and HPK. This result underlines that the categorization of the prior knowledge of learners might influence the interpretation of the effectiveness of an instructional technique.

From a practical perspective, partial reversals imply that an instructional support measure will alleviate differences between LPK and HPK learners by only aiding the LPK learners. A full reversal, on the other hand, implies that implementing the same support measure for all learners prevents HPK learners from exhibiting their full potential. Thus, from a normative standpoint full reversals impose an ethical conflict. From a theoretical perspective, distinguishing between partial and full reversals is important because they may be associated with different underlying cognitive processes.

4.2 Explanatory Approaches for the Expertise Reversal Effect

As stated in the preceding chapter the phenomenon of EREs in general contains a positive effect of an instructional support measure on learning outcomes for LPK learners whereas for HPK learners evidence revealed two different situations: (a) HPK learners are either not affected by instructional support (partial ERE), or (b) they are even hindered in learning (full ERE).

Explanations for the ERE are mostly framed against the backdrop of cognitive load theory (CLT; Chandler & Sweller, 1991), which will hence be introduced in the following. The theory makes assumptions on how the human cognitive system and the format of instructions influence learning (Kalyuga et al., 2003). The CLT proposes that human working memory capacity, which is limited, is overloaded during learning due to load induced by the format of instruction, and/or the task itself and its contents. Accordingly, three types of cognitive load are distinguished: (a) intrinsic cognitive load (ICL), (b) extraneous cognitive load (ECL), and (c) germane cognitive load (GCL) (Sweller, van Merriënboer, & Paas, 1998). ICL is imposed by a learner's prior knowledge and the contents of the material itself and cannot be changed by instructional design. ECL, on the other hand, is considered unnecessary load that results from an inadequate design of the contents. Finally, GCL is related to the cognitive load learners experience when processing materials and constructing schemas. It reflects positive load that arises from active and elaborate processing of materials. HPK learners are supposed to have automated knowledge structures in long-term memory: schemas (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Thus, their intrinsic load should be reduced in contrast to LPK learners because they automatically activate schemas during learning that might help them to incorporate or align the to be learned information into their existing knowledge structures (cf. Chi, Feltovich, & Glaser, 1981). Conversely, LPK learners have no or just few incomplete schemas available that do not support them in establishing schemas based on to be learned information. Therefore, LPK learners have to use their working memory resources to a greater extent than HPK learners in order to process information. Kalyuga et al. (2003) stated that for LPK learners instructional support measures might step in and substitute missing schemas, which is supposed to reduce working memory load (i.e., ECL) and thus aids LPK learners to process new information and construct schemas. Therefore, LPK learners should profit from instructional support regarding their learning outcomes.

There are different explanations of how HPK learners respond to instructional support. On the one hand, HPK learners may remain unaffected by instructional support, which – taken together with the benefits for LPK learners – would yield a partial reversal. On the other hand, HPK learners in contrast to LPK learners may even suffer from instructional support, thereby yielding a full reversal.

Explanatory approaches for a partial reversal rely on the cognitive load theory's assumption stating that HPK learners do not need further instructional support because they have automated schemas at hand that support them in learning processes. Hence, their schemas allow them to show a high level of performance regardless of whether instructional support is present. In line with this reasoning, Mayer and Sims (1994) suggested the *ability-as-compensator hypothesis*. This hypothesis was initially related to the spatial ability of learners rather than their prior knowledge. It predicts that high-spatial ability learners should be well able to learn with multimedia materials with and without instructional support. Their high spatial ability is assumed to compensate for poor instruction. Conversely, learners with low-spatial ability are supposed to require instructional support because they have to use their working memory resources to a greater extent than high-spatial ability learners. Hence, low-spatial ability learners should perform better when instructional support is present than when it is not (Mayer & Sims, 1994). The results obtained by Mayer and Sims (1994) corroborated the ability-as-compensator hypothesis. Moreover, the authors concluded that a similar pattern of result could also be expected when considering the prior knowledge of learners. If we consider the reasoning behind the ability-as-compensator hypothesis in the light of EREs, this would speak in favor of a partial ERE. HPK learners are assumed to compensate for missing instructional support by means of schemas in long-term memory that guide them during learning. Thus, HPK learners should not be affected by instructional support regarding their learning outcomes.

Applying this reasoning to MIS, HPK learners are expected to be able to identify text-pictures correspondences by applying their background knowledge. Accordingly, they can establish a coherent integrated mental model without receiving further guidance similar to LPK learners who receive additional support. Hence, expert learners might compensate for missing guiding information since they already have established schemas that guide them during learning.

Explanatory approaches for a full reversal focus on the question of how HPK students process the instructional support. According to the first explanation, HPK learners might refrain from elaborating the multimedia materials once instructional support is present, which in turn leads to less learning. Similar effects have been observed in text comprehension research, where LPK readers benefit from coherent texts, whereas HPK readers show better comprehension when reading less coherent texts (McNamara et al., 1996). A lack of coherence forces HPK learners to overcome coherence gaps by applying their prior knowledge, which leads to more active processing and elaboration of relations within the text. Applying this line of reasoning to the use of MIS, un-signaled multimedia instruction can be considered an incoherent format of instruction that induces gap-filling inferences in HPK learners. In contrast, MIS will suppress these inferences, thereby hampering HPK learners' performance. The finding by McNamara et al. (1996) and its application to the signaling context can be interpreted in the light of the CLT. Since HPK learners refrain from deeper processing when MIS are present, they are assumed to experience less GCL. In turn, less GCL is assumed to decrease learning.

The second explanation is prominent in the context of the CLT. According to this explanation, HPK learners are unable to ignore instructional support, even though the information provided by the instruction and their schema in long-term memory are redundant or at least partly overlapping. Kalyuga et al. (2003) assumed that HPK learners relate both sources of information or even try to integrate them with each other. This processing of potentially redundant information in turn increases their ECL and hence hinders learning resulting in a full ERE. This interpretation is also supported by subjective mental load measures (Kalyuga & Renkl, 2010). Applying this explanation to learning with MIS suggests that MIS induce unnecessary processing of information, thereby leading to a decline in performance. With MIS present, learners are more likely to process both the text and the picture, even though HPK learners could learn just as effectively with only one external representation such as the picture. A similar explanation has been put forward for ERE regarding the split-attention effect (Kalyuga et al., 1998). Here, instructional formats where the text is not integrated into the picture were shown to be more effective for HPK learners, whereas LPK learners benefitted from instructional formats in which the text was physically integrated into the picture. According to Kalyuga et al. (1998) an integrated format that encourages processing of both text and picture (like MIS do) enforces HPK learners to process redundant information, even though one representational format would be sufficient for them.

However, there are at least two problems with the redundancy explanation. First, as Schnotz (2010) points out there is an inconsistency of this notion with the CLT itself. According to the CLT, learners with a high level of expertise have established cognitive schemas that are quite well automated. The automation of schemas is supposed to lead to a decline in cognitive load. Accordingly, expert learners should have an overall lower basic cognitive load level than novice learners (cf. Chi, 2006). Schnotz (2010) therefore questions why an instruction that does apparently not overload novice learners (since it fosters learning) should overload learners with expertise.

Second, the redundancy explanation implies that HPK learners are unable to ignore information that is not helpful for them. However, the information-reduction hypothesis by Haider and Frensch (1999) suggests that with increasing expertise people become better at ignoring information that is unnecessary for task performance. With regard to the signaling effect, Scheiter and Eitel (2015) showed that learners who were presented with mismatched signals (i.e., signals that highlighted alleged text-picture correspondences where there were none) only initially attended to these signals but ignored them once they recognized that they did not provide helpful information for learning. Accordingly, learning outcomes were also not affected by these mismatched signals. These results were obtained with LPK learners. But if LPK learners, who experience a quite high level of cognitive load are already able to ignore misleading and hence unnecessary information, HPK learners with an overall lower cognitive load level should even be more likely to do so.

To conclude, the effectiveness of MIS might be moderated by learners' prior knowledge. However, at present it is not possible to decide whether MIS yield a partial or full reversal for HPK learners and, if a full reversal occurs, why it does. Furthermore, the existing explanatory approaches related to EREs are problematic because they speculate about underlying processes without having clear evidence. Thus, in Study 2 and 3 cognitive load measures were assessed to address this limitation.

4.3 Insight into Expertise Reversal Effects: Measurement approaches

According to the preceding chapters, explanatory approaches for EREs are based on the CLT (e.g., Kalyuga et al., 2003). Thus, cognitive load measures might provide indications regarding the underlying cognitive processes for EREs.

As reported by de Jong (2010) and Paas, Tuovinen, Tabbers, and van Gerven (2003), research has used different methods in order to assess (types of) cognitive load. Among the reported measures are (a) self-ratings through questionnaires, and (b) physiological measures such galvanic skin response, heart rate variability, neuro-imaging techniques, and pupil diameter (cf. de Jong, 2010; Schnotz & Kürschner, 2007). In the present thesis, pupil diameter was assessed as a cognitive load measure beyond using self-rating. Both measures will be presented in detail hereinafter. In addition, the distribution of visual attention might also shed light on underlying processing differences of multimedia related to expertise reversals of the signaling effect. Thus, the use of (c) measures reflecting visual attention distribution will be described in detail in chapter 4.3.2.

4.3.1 Measuring Cognitive Load

Self-ratings. Self-ratings are a very frequently used method to assess cognitive load types (De Jong, 2010; Schnotz, & Kürschner, 2007). The basic assumptions underlying the use of subjective ratings is that learners are able to contemplate their own cognitive processes and report their mental effort and their perceived difficulty during learning (Paas et al., 2003; Schnotz & Kürschner, 2007).

One of the most widely used self-rating items in cognitive load research was developed by Paas (1992). He asked participants to rate their perceived mental effort on a 9-point rating scale from 1 (very, very low mental effort) to 9 (very, very high mental effort) several times during training and testing. Paas (1992) distinguished two concepts related to cognitive load: mental load and mental effort, whereby he classifies mental effort as an index of cognitive load in general. According to Paas (1992) “mental load is imposed by instructional parameters (e.g., task structure, sequence of information” (p.429), which seems to be closely related to the concept of

ECL imposed by the design of materials. For mental effort he stated that “mental effort refers to the amount of capacity that is allocated to the instructional demands” (p. 429), which shows parallels to the concept of GCL. De Jong (2010) also addressed this issue of the definition of cognitive load components and comes to a similar assessment of mapping mental load and mental effort to the cognitive load components ECL and GCL based on conclusions by Kirschner (2002) and Sweller et al. (1998).

However, although subjective ratings are frequently used in cognitive load research (Paas et al., 2003), de Jong (2010) points out several research issues related to the measurement of (types of) cognitive load. Hence, the outcomes of studies regarding cognitive load might depend for instance on the particular question asked and the timing of the questionnaire (e.g., during learning or testing). With regard to the latter issue, some studies have assessed cognitive load during or after learning, whereas others measured cognitive load during the test phase (cf. van Gog & Paas, 2008). Van Gog and Paas (2008) suggest that cognitive load measurements during the test phase reflect the quality of learning and the resulting mental representation, respectively. That is, students who learned more due to receiving better instructional support should experience less cognitive load when retrieving their knowledge in the test phase than students who learned less in the learning phase. Moreover, according to de Jong (2010) it is an open question whether participants are able to specify an average of their cognitive load or whether an average calculated from several cognitive load ratings is a better estimate for the subjectively experienced cognitive load. Schmeck, Opfermann, van Gog, Paas, and Leutner (2014) found that an overall rating of cognitive load given after problem solving was higher than the average of ratings given during problem solving for each problem individually. The single rating corresponded best to the ratings of the most complex problems, suggesting that subjects used the peak of their experienced cognitive load as an anchor to base their overall evaluation upon (cf. Paas et al., 2003). Finally, Schnotz and Kürschner (2007) pointed out that another issue of subjective cognitive load ratings is that the individual framework of reference for ratings is not stable, because it might vary due to motivational or emotional changes during learning. This issue can decrease the reliability of this measurement approach in general (cf. Schnotz & Kürschner, 2007).

To sum up, subjective self-ratings are a commonly used method to assess (types of) cognitive load and can easily be administered in studies; however, measurement problems exist.

Pupil diameter. The basic assumption for the relation between pupil diameter and cognitive load is that the pupil dilates with increasing load at least in young adults (van Gerven, Paas, van Merriënboer, & Schmidt, 2004). The pupil is a round hole located in the central area of the eye. It controls how much light strikes the retina, which is located in the posterior part of the eye. For this purpose, eye muscles dilate and constrict the pupil depending on different factors such as luminance of the surrounding (Attar, Schneps, & Pomplun, 2016; De Groot & Gebhard, 1952; Nunnally, Knott, Duchnowski, & Parker, 1967) or mental activity (e.g., Beatty, 1982). The diameter of the pupil is measured as the length of a straight line that crosses the center of the pupil and ends at the outer boundary of the pupil.

Early work by Hess and Polt (1964) and Kahneman and Beatty (1966) revealed homogeneous pattern of results in that pupil diameter was positively related to task difficulty, which was interpreted as a proxy of memory load, mental activity, and effort. Kahneman and Beatty (1966) presented participants with either strings of digits with differing length ranging from three to seven digits, a string of four high-frequency monosyllabic nouns and a string of four digits that required the addition of one by the participants. The strings were each first presented and then subjects were asked to respond immediately to the task either by adding the digits or recalling the nouns. The results revealed that the pupil diameter was significantly larger for (a) more complex tasks like the presentation and recall of seven digits in contrast to easier tasks with fewer digits and (b) the addition tasks in contrast to recall tasks with four digits/words. Similar results were obtained for example for link selection processes during text reading (Scharinger, Kammerer, & Gerjets, 2015), spatial visual search tasks (Porter, Troscianko, & Gilchrist, 2007) or memory search tasks (Van Gerven et al., 2004).

However, there are also issues related to the use of pupil diameter as a measure for cognitive load. As mentioned above, the pupil dilates and constricts also in response to changes in luminance (Attar et al., 2016; De Groot & Gebhard, 1952; Nunnally et al., 1967). This issue can be addressed by keeping the luminance of a stimulus presented to learners stable. Moreover, individual baselines related to the pupil diameter may vary between participants (Schnotz & Kürschner, 2007). Nevertheless, Paas et al. (2003) concluded that this measure is a "highly sensitive instrument for tracking fluctuating levels of cognitive load" (p. 66).

What has become evident from the literature on pupil diameter as a measure for cognitive load is that a consistent definition of the construct of cognitive load correlating with pupil diameter seems to be lacking. Thus, in summary, a conclusion made by Scharinger et al. (2015) might best describe the sensitivity of the pupil diameter in that "pupil dilation may be seen as a rather overall load measure, including aspects of effort, motivation, arousal, and emotion" (p. 3).

4.3.2 Measuring Visual Attention Distribution

As already mentioned, the basic assumption about tracking the movement of the eyes during a task is that visual attention devoted to materials provides information about concurrent cognitive processes (eye-mind assumption; Just & Carpenter, 1980). Eye tracking allows recording the movement of the eyes on a stimuli, like a textbook page, thereby providing data about the distribution of visual attention during the processing of the information. Eye tracking methodology is frequently used in reading research (for a review, see Rayner, 1998, 2009). In addition, research on multimedia instructional design increasingly uses eye tracking for a variety of research questions (Scheiter & van Gog, 2009; van Gog & Scheiter, 2010). The signaling effect in multimedia learning is one of the fields of application for eye tracking. Importantly, the majority of signaling studies in the field of multimedia learning including eye tracking methodology used samples consisting of LPK learners. Therefore, the results of multimedia signaling studies regarding eye tracking parameters (guiding-attention- and unnecessary visual-search hypothesis; e.g., Jamet, 2014; Ozcelik et al., 2010; Scheiter & Eitel, 2015) can only be interpreted for LPK learners. If we consider the explanatory approaches for EREs (cf. chapter 4.2), HPK learners presumably process material including MIS differently in contrast to LPK learners. Moreover, their visual attention distribution during learning might differ because HPK learners have schemas at hand that guide them during learning.

Van Gog, van Kester, Nievelstein, Giesbers, and Paas (2009) addressed the latter issue by suggesting to use eye tracking methodology to gain insights into the underlying processes of the ERE. In line with this notion, Gegenfurtner, Lehtinen, and Säljö (2011) conducted a meta-analysis on eye tracking research related to expertise differences particularly in the comprehension of visualizations in different domains such as biology (e.g., Jarodzka, Scheiter, Gerjets, & van Gog, 2010) or radiology (e.g., Kundel, Nodine, Conant, & Weinstein, 2007). The meta-analysis revealed differences in visual attention distribution on visualizations for experts in contrast to novices. Experts had shorter fixation durations, which is in line with the *theory of long-term working memory* by Ericsson and Kintsch (1995) indicating that experts encode and retrieve information faster than novices. Moreover, in line with the information-reduction hypothesis by Haider and Frensch (1999) experts had more fixations on task-relevant areas of the visualization and shorter times to first fixating relevant information compared to novices, which suggests that experts are able to ignore task-irrelevant information (Gegenfurtner et al., 2011). The meta-analysis thus provides evidence that eye tracking measures are suitable for detecting differences in visual attention distribution between experts and novices at least for processing visualizations.

Against the backdrop of the meta-analysis by Gegenfurtner et al. (2011), eye tracking measures might also reveal differences in visual attention distribution for learners with different prior knowledge levels learning with multimedia materials. This assumption is corroborated by results obtained by Mason, Tornatora, and Pluchino (2013) in a study with fourth graders learning with science multimedia materials. The results revealed that prior knowledge was positively correlated with the number of transitions between texts and pictures and the fixation time on the picture during re-reading the text during multimedia learning. They concluded that HPK learners might have more attention left than LPK learners and therefore show a much more strategic processing in integrating verbal and pictorial information (Mason, Tornatora, &

Pluchino, 2013). Moreover, they showed that visual behavior was significantly associated with immediate recall, transfer and factual knowledge performance as well as performance in a delayed factual knowledge test (determined by Kruskal–Wallis tests), in that the more integrative pattern of eye movements, which is presumably related to text-picture integration processes (i.e., longer fixation on picture and more transitions) was related to highest scores in learning outcomes and vice versa (Mason, Tornatora, & Pluchino, 2013).

Moreover, in a study by Schwonke, Berthold, and Renkl (2009) participants learned about probability theory with a multimedia learning environment containing texts and diagrams. Participants were either provided with information about the function of displayed diagrams or not. Schwonke, Berthold, and Renkl (2009) aggregated different eye tracking parameters like fixation time and the number of transitions between texts and pictures into one measure reflecting visual attention on the different types of representations like for example on the displayed diagrams included in the multimedia materials. They used the aggregated visual attention measure as a mediator and prior knowledge as a moderator variable to conduct a moderated mediation analysis to analyze whether visual attention moderated by prior knowledge would explain the effect of informing students about the function of diagrams on learning outcomes. The effect of the intervention on learning outcomes was, however, not moderated by prior knowledge, thus revealing no ERE. However, the intervention increased learning outcomes significantly. The visual attention on diagrams decreased in the group without information about the function of diagrams with increasing prior knowledge, which is in line with the result obtained by Gegenfurtner et al. (2011) in that experts showed shorter fixation durations in contrast to novices. The influence of prior knowledge disappeared in the group that received information about the function of the diagrams. In this group visual attention remained stable across different prior knowledge levels. Thus, the authors interpret that their intervention prevented HPK learners from devoting too little attention to representations within the learning material (Schwonke, Berthold, & Renkl, 2009). They also tested the relationship between visual attention and learning outcomes and the influence of prior knowledge. Results revealed that the effect of visual attention on diagrams on learning outcomes was moderated by prior knowledge in that HPK learners tended to learn more the more visual attention they devoted to diagrams. Conversely, LPK learners tended to perform worse the more visual attention they devoted to diagrams, albeit not significantly for both HPK and LPK learners. Finally, a moderated mediation analyses showed that visual attention moderated by prior knowledge could explain better performance in a post-test.

Evidence reported in the latter section points towards eye tracking as a powerful tool to gain insight in processing differences between learners with different levels of prior knowledge for learning with multimedia materials. What remains open is how exactly prior knowledge affects the relation between the instruction, visual attention distribution, and learning outcomes. Mason, Tornatora, and Pluchino (2013) showed that depending on prior knowledge learners visually processed materials differently. At the same time, visual attention distribution was related to learning outcomes. However, Mason, Tornatora, and Pluchino (2013) did not test these relations by means of a moderated mediation analysis (cf. Preacher, Rucker, & Hayes, 2007). A moderated mediation analysis would on the one hand allow to investigate whether (a) the effect of the instruction on visual attention and (b) the effect of visual attention on learning outcomes is influenced by prior knowledge. On the other hand, a moderated mediation analysis reveals whether these particular effects explain the effect of the instruction on learning outcomes depending on prior knowledge. Hence, in the study by Mason, Tornatora, and Pluchino (2013) it remained unclear whether prior knowledge only affected the way learners visually processed multimedia materials or whether prior knowledge also influenced how visual processing was related to the cognitive outcomes, namely, learning outcomes. Conversely, Schwonke, Berthold, and Renkl (2009) investigated these relations by means of a moderated mediation analysis. The authors showed that prior knowledge influenced the way learners visually attended to the materials as well as learning outcomes.

Against the backdrop of these considerations, it is at present not possible to decide how prior knowledge exactly influences the effectiveness of MIS for learning on a process level. Although Scheiter and Eitel (2015) investigated the effectiveness of MIS by means of a mediation analyses, their results were obtained for LPK learners only. Neither Mason, Tornatora, and Pluchino (2013) nor Schwonke, Berthold, and Renkl (2009) manipulated the presence or absence of MIS in their multimedia learning studies. Thus, the interpretation of their results related to the context of the present thesis is only limited. However, taken the results by Gegenfurtner et al. (2011), Mason, Tornatora, and Pluchino (2013), and Schwonke, Berthold, and Renkl (2009) together prior knowledge seems to influence the way learners visually process materials. Regarding the second potential influence of prior knowledge on cognitive processing of looked at information the empirical basis is weak. Only Schwonke, Berthold, and Renkl (2009) showed that prior knowledge also influenced how information that learners looked at was cognitively processed, thereby explaining differences in learning outcomes.

Thus, the question what visual attention processes underlie a potential expertise reversal of the multimedia signaling effect will be addressed in Study 3 (chapter 8) of the present thesis by assessing eye tracking measures and using a moderated mediation analysis method.

5. Overview and Research Questions

The goal of the present dissertation is to investigate the effectiveness of multimedia signaling for learning. Against the backdrop of the theoretical considerations presented in the preceding chapters, the following research questions were derived:

1. Are MIS an effective design measure to foster learning with multimedia (Research Question 1)? It was expected that MIS facilitate the integration process of verbal and pictorial information. Thus, MIS should support the establishment of coherent mental representations, which in turn should be reflected by better learning outcomes. In particular, MIS were expected to improve comprehension and transfer performance because these measures reflect deep learning (cf. Mayer, 2014b).
2. Which material-based boundary conditions moderate the multimedia signaling effect (Research Question 2)? It was assumed that four material-based boundary conditions might influence the effectiveness of MIS: (a) pacing of the materials, (b) pictorial format, (c) multimedia mapping requirements, and (d) distinctiveness of MIS.
3. Does an expertise reversal effect occur related to the effectiveness of MIS (Research Question 3)? Against the backdrop of the ERE (Kalyuga et al., 2003) and the ability-as-compensator hypothesis (Mayer & Sims, 1994) the assumption was derived that the learner-based boundary condition prior knowledge affects the effectiveness of MIS.
4. Does the influence of prior knowledge lead to a partial or full expertise reversal of the multimedia signaling effect (Research Question 4)? In case prior knowledge influences the effectiveness of MIS it was expected that either a partial or full ERE would be obtained. LPK learners were expected to profit from MIS whereas HPK learners should either not be affected (partial ERE) or even be hindered in learning (full ERE).
5. How can a potential partial or full expertise reversal of the multimedia signaling effect be explained (Research Question 5)? Related to this research question it was assumed that depending on prior knowledge signals cause

learners to visually process materials and experience cognitive load differently. This in turn was expected to influence learning outcomes.

To sum up, in the following chapters three studies including a variety of methodological approaches will be reported. Research Questions 1 to 4 were addressed in a comprehensive meta-analysis (Study 1). Because this meta-analysis revealed prior knowledge to be a more important boundary condition than material-bound boundary conditions, two empirical studies focusing on prior knowledge as a moderator of the MIS effect were carried out in addition. In an ecologically valid field study (Study 2) and a laboratory experimental eye tracking study (Study 3) Research Questions 1 as well as 3 to 5 were addressed.

6. Study 1: Meta-Analysis on Multimedia Signaling[1]

The goal of the present study was to systematically review literature regarding the effectiveness of MIS for learning by means of a meta-analysis (Research Question 1). Moreover, potential material-based boundary conditions and the learner-based boundary condition prior knowledge related to the effectiveness of MIS were investigated (Research Questions 2 to 4). To the best of my knowledge, only two reviews about the signaling effect in multimedia learning have been published so far that provide a more systematic account of findings related to the signaling effect in multimedia learning. In 2009, Mayer re-analyzed five studies from his own lab to determine the effectiveness of multimedia signaling by comparing the transfer test performance of groups that learned from signaled- and groups that learned from non-signaled multimedia material. He found a positive multimedia signaling effect resulting in an overall medium effect size (d = .52). It is important to keep in mind when interpreting this as confirmatory evidence for the multimedia signaling principle that, because they were all conducted in the same lab, all studies used very similar materials, testing procedures, and subject samples. This similarity among the studies is likely to cause an overestimation of the size of the signaling effect and is not particularly suited to unravel possible boundary conditions because of the homogenous study conditions under which the signaling effect was investigated in this case. Moreover, several empirical studies, including those by other authors and those conducted since 2009 (e.g., Jamet, 2014; Ozcelik et al., 2010; Scheiter & Eitel, 2015), are not included in the review of Mayer (2009).

De Koning et al. (2009) also reviewed the effects of signaling, but only when learning from instructional animations. They reviewed 13 studies and proposed a framework that classifies different functions of signaling. Although important, instructional animations are only one of several types of multimedia instructions. Another very common type of multimedia instruction is to present text with static pictures, which is thus also considered in the present meta-analysis.

---

[1] Richter, J., Scheiter, K., & Eitel, A. (2016) Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19-36. doi:10.1016/j.edurev.2015.12.003.

To conclude, the signaling principle for multimedia learning lacks a comprehensive review including relevant studies from different labs. The current meta-analysis aimed at systematically investigating the validity of the signaling principle in multimedia learning studies that compared performance of a group learning from signaled multimedia material with that of a control group. Moreover, potential boundary conditions for the signaling effect were investigated in order to determine for whom and under which conditions multimedia signaling fosters or rather hampers learning. The material-based boundary conditions reported in chapter 3.3 and the learner-based boundary condition prior knowledge reported in chapter 4 will be considered as moderator variables in the current meta-analysis.

6.1 Hypotheses

Theory and research suggest that there is a beneficial effect of signaling in multimedia learning, but that it might be subject to certain boundary conditions. To shed light on the signaling effect, a comprehensive meta-analytic review was conducted in order to determine the overall size of the signaling effect (for MIS) along with its potential moderators. The following hypotheses were derived:

1. Overall effect: Because MIS can facilitate integration of information from different media, an overall positive effect on comprehension outcomes was hypothesized (Hypothesis 1).

   Regarding Hypothesis 1, it was furthermore taken into account that studies differ in what they consider an adequate control group. Some studies used rather weak control groups including no MIS at all, whereas other studies implemented basic MIS in the control group including corresponding labels in verbal and pictorial information. The prior studies used control groups with learning material that would be unlikely to be used in real educational contexts because of its poor design. This might be problematic, since the effects of the instructional intervention are potentially maximized by the control group design. Referring to Schwonke, Renkl et al. (2009), effects should thus be tested not only in comparison to 'lousy' control conditions but also in comparison to fair control conditions in terms of rather ecologically valid materials to test the robustness of an effect. It was assumed that the overall signaling effect in studies using rather weak control conditions (i.e., without MIS) would be larger than in studies using basic MIS such as corresponding labels in text and picture in their control group.

2. Domain-specific prior knowledge: Referring to the ability-as-compensator assumption and the ERE, instructional support in the form of MIS is assumed to be more beneficial for LPK than for HPK learners. HPK learners might achieve integration without MIS by using schemas or might even be hindered in learning due to unnecessary processing of redundant information whereas LPK learners require extra support (Hypothesis 2).
3. Pacing of the materials: The signaling effect should be larger for system-paced than for self-paced instructions, because in the latter learners can compensate difficulties in integrating information to a greater extent, for instance, by allocating more study time (Hypothesis 3).
4. Pictorial format: It was expected that the signaling effect will be moderated by the format of pictorial information (static vs. dynamic visualizations) (Hypothesis 4). As there is mixed evidence for the influence of the pictorial format on the signaling effect and different theoretical approaches, no directed hypothesis was postulated.
5. Multimedia mapping requirements: Signals should be more effective for learning from material with many mapping requirements than from material that poses only few mapping requirements (Hypothesis 5).
6. Distinctiveness of MIS: Discursive MIS are expected to be less effective for learning than visual MIS because they are less salient (Hypothesis 6).

6.2 Method

This review incorporates studies that used multimedia learning material (dynamic and/or static pictorial and verbal information) and signals highlighting correspondences between multiple external representations (MIS). To this end, various steps common to meta-analyses were carried out: (a) data collection by means of literature search, (b) definition and application of inclusion criteria to filter relevant studies and coding of study characteristics, (c) calculation of effect sizes, (d) conducting a basic meta-analysis, (e) conducting moderator analysis and estimating the publication bias (Field & Gillett, 2010). These steps are described in the following chapters.

6.2.1 Data Collection and Inclusion Criteria

The literature search was conducted by using combinations of the keywords “signal”, “signal(l)ing”, “cue”, “cueing”, “multimedia learning” and “learning” and “text and pictures” separated by the Boolean operators “AND”/ “OR”. The databases ERIC, PsychINFO and ScienceDirect, as well as EARLI Books of Abstracts (2005 – 2013) and AERA Proceedings (2010-2014, only available online) and reference lists in relevant articles were used to find relevant studies. To counteract publication bias, researchers were contacted to ask them for further published or unpublished studies that meet the inclusion criteria (via mailing lists of the German Psychological Society [DGPS] and EARLI/JURE social media presence and mailing to [first] authors of included studies).

A total of 1,060 articles were identified by search results in databases and proceedings, from scanning reference lists of relevant articles and by replies from first authors to mailings based on the aforementioned search criteria. However, 1002 articles had to be excluded due to different aspects: (a) no use of multimedia learning material or signaling, (b) the language of the article (not in English or German), (c) data was published multiple times, or (d) publication was not accessible.

The set of potentially relevant articles (*N* = 58) was then scanned based on the following inclusion criteria:

1. Verbal information was provided in either written or spoken format.
2. Pictorial information was provided in either static or dynamic format.
3. Signals that are aimed at supporting integration of verbal and pictorial information (MIS) were implemented.
4. MIS were either implemented in the verbal information or in pictorial information or in both.
5. A control group learning with material including basic MIS (strong control group) or no MIS (weak control group) was used.

6. The study reported sufficient quantitative data in order to be able to calculate effect sizes.

After a careful review of the potentially relevant articles based on all inclusion criteria, a set of 28 articles remained[2]. Most of the excluded articles did not meet all of the inclusion criteria or were not accessible even after contacting the author[s].

6.2.2 Study Features

Feature characteristics of the identified studies were extracted and coded (see Table 1). The type of learning outcome measure was defined as transfer or comprehension performance only, because (a) transfer is the major variable of interest in multimedia research, and because (b) successful integration is assumed to be especially beneficial for transfer performance (cf. Mayer, 2014a). Moreover, choosing only one dependent variable (transfer/comprehension; henceforth called comprehension) allowed us to avoid dependencies between effect sizes introduced by multiple outcome measures, thereby contributing to the independence of effect sizes, which is among other things important for the validity of a meta-analysis (cf. Scammacca, Roberts, & Stuebing, 2014).

The prior knowledge of the learners was classified as being either on a low/medium level or on a high level. In the majority of cases, the classification of low versus high prior knowledge made by the authors of the original studies was adopted. If there was no such classification available (26% of studies), the percentage of the mean of correct answers in relation to the maximum score that could potentially be reached in the prior knowledge test was used to code the prior knowledge level. The mean of prior knowledge test results was calculated by weighting it by the corresponding sample size given in the studies. If participants scored on average below 60% on the prior knowledge test, they were classified as low to medium prior knowledge learners (LPK). Participants who scored above 60% would have been classified as a HPK level, but none of the learners in the studies that lacked a prior knowledge classification by the authors scored this high.

[2] Only three out of six effect sizes used in the signaling review by Mayer (2009) were included in the present meta-analysis since the three experiments included in Stull and Mayer (2007) did not meet the inclusion criteria of using MIS in the experimental conditions.

Table 1

Study feature characteristics

Study features	Characteristic
First author	[…]
Year of the publication	[…]
Type of publication	journal article, conference paper, PhD thesis
Sample size	[…]
Sample characteristic	primary and middle school (up to 9th), high school (up to 12th), university, vocational training
Learning domain	biology, chemistry, computer science, educational psychology, math, physics/mechanics
Pictorial format	static, dynamic
Learning outcome	transfer and comprehension measures only
Domain-specific prior knowledge	low, high
Multimedia mapping requirements	few, many
Pacing of the materials	self-paced, system-paced
Distinctiveness of MIS	discursive, visual, mixed discursive and visual
Additional instructional support	absent, present
Type of instructional support if present	[…]
Type of control group	weak, strong

Note. The symbol […] means that the actual terms/values were used as study feature characteristics.

The pacing of the materials was categorized as either system-paced or self-paced. A system-paced learning material presentation was coded when learners had no option to interact with the presentation. When learners could start, pause, and stop (control) the presentation, the presentation of the learning material was categorized as self-paced.

Two experts (co-authors of the published paper) were asked to rate the mapping requirements of the learning materials with respect to the integration process of verbal and pictorial information. Sample materials provided in the papers were rated according to the number of text-picture correspondences that had to be identified by the learner and the number of visual elements that could potentially serve as distractors during this identification. Based on screening the whole set of studies, these ratings were then classified as representing either few or many mapping requirements. Cases of disagreement were resolved by discussion. I am aware of the fact that this procedure is not only rather subjective, but also heavily relies on the sample materials provided by the studies' authors; thus, respective findings should be treated with caution.

Because the focus of some of the studies was not solely on the signaling manipulation, instructional supports like prompts or step-by-step presentation were additionally coded (absent vs. present).

The type of control group was also rated as to whether it included no MIS at all (weak control group) or whether it implemented a basic MIS (strong control group).

6.2.3 Computation of Effect Sizes and Analysis

For the effect size calculation and meta-analysis an approach proposed by Field and Gillett (2010) was chosen. The correlation coefficient r was used as an effect size since it is well understood in the field and it is flexible in that it can be calculated for any combination of dichotomous and quantitative variables (McGrath & Meyer, 2006). Interpretative benchmarks for the effect size r as suggested by Cohen (1992) were used: a small effect corresponds to $r = .10$, a medium effect is denoted by $r = .30$, and a large effect corresponds to $r = .50$ or larger.

Since the validity of a meta-analysis result depends strongly on the independence of the included effect sizes, dependencies were avoided by choosing only one outcome measure (only comprehension performance). Moreover, there was no danger of having dependencies due to multiple group comparisons, since there were no within-subject designs among the included effect sizes (Scammacca et al., 2014). Hence, only independent effect sizes were included: (a) all studies used between-subjects designs and (b) control groups were only used once for effect size calculations within a study.

The effects sizes of most pair-wise comparisons were calculated by means of the software environment for statistical computing R using the compute.es package

(Del Re, 2014; R Development Core Team, 2008). Within the compute.es package the mes function is defined, which transforms d into r using the formulas suggested by Rosenthal (1991). The mes function converts raw mean scores and standard deviations of the experimental and control group into effect sizes accompanied by the confidence interval, variance and p-value. The computed effect sizes were then compiled into an SPSS data file (IBM Corp., 2013) containing all coded variables of each pair-wise comparison. Detailed information about the adjustments of group combinations for the effect size calculation of the included studies can be found in Table 2.

Outliers representing extreme values regarding their effect sizes were excluded from the meta-analysis following Tukey's approach (Hoaglin, Mosteller, & Tukey, 1983). To this end, two effect sizes (r = -.33 in Jeung et al. [1997, Exp. 1] and r = .80 in Moreno, Reisslein, and Ozogul [2010]) were removed from the data resulting in 27 remaining articles.

The meta-analysis was conducted using the Hedges and Vevea random-effects method developed by Hedges and colleagues (Hedges & Olkin, 1985; Hedges & Vevea, 1998) using SPSS scripts provided by Field and Gillett (2010). A random-effects model was preferred to a fixed-effects model because the conditions under which signaling effects were tested across studies cannot be assumed to be completely identical (e.g., differences in samples, materials etc. between studies). Hence, effect sizes can be assumed to vary randomly across studies rather than being fixed (Hunter & Schmidt, 2000). In addition, Field and Gillett (2010) conclude that for social science data the standard model applied should be conceptualized as a random-effects model. According to Monte Carlo simulation results derived by Field (2005) the Hedges and colleagues' method shows higher proportions of confidence intervals containing the true effect sizes in contrast to another popular random-effects method by Hunter and Schmidt (2004). Therefore, the Hedges and Vevea random-effects method was applied in this meta-analysis.

Moderator analysis and publication bias estimation were conducted using SPSS scripts provided by Field and Gillett (2010). For the moderator analysis contrast weights were introduced to compare groups. Different approaches for the publication bias estimation were tested. A popular measure for this estimation is *fail-*

safe N, Rosenthal (1979). The *fail-safe N* measure indicates the estimated number of unpublished studies that would be necessary to be included in the meta-analysis in order to turn an estimated significant population effect size into a non-significant effect size. However, the *fail-safe N* measure is heavily discussed and alternative procedures have been introduced for the estimation of the publication bias (Vevea & Woods, 2005). Accordingly, following a recommendation by Field and Gillett (2010), the Begg and Mazumdar (1994) rank correlation was additionally used to test for publication bias. The rank correlation test quantifies the association between the effect sizes and their sampling variance. The smaller the correlation, the more independent the effect sizes are from the sample sizes of the studies, and hence, the more unlikely is a publication bias. Begg and Mazumdar (1994) state that the test has moderate to large power for the number of effect sizes used in this meta-analysis.

6.3 Results

6.3.1 Sample

Twenty-seven articles yielding k = 45 pair-wise comparisons were included in the meta-analysis (see Table 2). Pair-wise comparisons included the comprehension performance of multimedia learning material with and without MIS. The studies were published between 1997 and 2015 as journal articles (85.2%), conference papers (7.4%) or PhD theses (7.4%). A total of 2,464 subjects participated in the studies with a mean sample size of N = 55 (ranging from N = 16 to N = 158). The total number of participants in the experimental group was 1,285.

Most of the participants were students in tertiary education (e.g., universities, 60.0%) followed by primary/middle school (33.3%), high school (4.4%), and vocational training (2.2%). The topics used in the studies were mostly from science domains: physics/mechanics (42.2%), biology (28.9%), math (11.1%), educational psychology (6.7%), and computer science (2.2%). In addition, 8.9% of the studies used learning material from both biology and physics domain.

Table 2

Overview on study features of 27 articles yielding 45 pair-wise comparisons

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
Arslan-Ari, 2013	university	55	0.15	transfer	biology	low/medium	self-paced	dynamic	many	visual	strong	no
	university	44	-0.26	transfer	biology	high	self-paced	dynamic	many	visual	strong	no
Baetge & Seufert, 2010	university	59	0.12	comprehen-sion	computer science	low/medium	self-paced	static	few	mixed discursive and visual	weak	no
Berthold & Renkl, 2009	high school	43	-0.05	problem-solving	math	low/medium	self-paced	static	many	visual	weak	no
	high school	42	0.31	problem-solving	math	low/medium	self-paced	static	many	visual	weak	scaffolding self-explanation prompts
Boucheix & Guignard, 2005	primary/ middle school	22	0.21	transfer	mechanics	low/medium	self-paced	static	many	mixed discursive and visual	strong	no
	primary/ middle school	20	0.08	transfer	mechanics	low/medium	self-paced	dynamic	many	mixed discursive and visual	strong	no
	primary/ middle school	41	0.15	transfer	mechanics	low/medium	system-paced[a]	static	many	mixed discursive and visual	strong	no
	primary/ middle school	40	0.11	transfer	mechanics	low/medium	system-paced[a]	dynamic	many	mixed discursive and visual	strong	no
de Koning, Tabbers, Rikers, & Paas, 2010b	university	38	0.51	transfer	biology	low/medium	system-paced	dynamic	few	visual	strong	no

[a] The pacing factor levels fast and slow were averaged into one system-paced level by calculating the means weighted by their sample sizes and by calculating pooled variances for the static and animated visualization group.

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
Florax & Ploetzner, 2010	university	66	0.10	comprehen-sion	biology	low/medium	self-paced	static	many	discursive	weak	no
	university	66	0.24	comprehen-sion	biology	low/medium	self-paced	static	many	discursive	weak	yes, segmented text
Harp & Mayer, 1998, Exp. 3[b]	university	48	0.16	transfer	physics	low/medium	self-paced	static	few	discursive	strong	no
Jamet, Gavota, & Quaireau, 2008	university	102	0.04	transfer	biology	low/medium	system-paced	static[c]	few	visual	strong	no
Jamet, 2014	university	32	0.19	transfer	social science	low/medium	system-paced	static	few	visual	weak	no
Jeung, Chandler, & Sweller, 1997, Exp. 1[d]	primary/ middle school	20	-0.02	transfer (problem 1)	math	low/medium	self-paced	static	few	visual	strong	no
Jeung, Chandler, & Sweller, 1997, Exp. 2	primary/ middle school	20	0.07	transfer (problem 1)	math	low/medium	self-paced	static	many	visual	strong	no
Jeung, Chandler, & Sweller, 1997, Exp. 3	primary/ middle school	20	0.07	transfer (problem 1)	math	low/medium	self-paced	static	few	visual	strong	no

[b] Additionally included the text signals preview sentences and enumeration.
[c] The display mode factor levels static and sequential displays were averaged by calculating the mean weighted by the corresponding sample sizes and by calculating the pooled variance.
[d] One pair-wise comparison was excluded from experiment one since it was identified as an extreme value.

(continued on next page)

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
Johnson, Butcher, Ozogul, & Reisslein, 2013[e]	university	98	0.11	far transfer	physics	low/medium	self-paced	static	few	discursive	strong	no
Johnson, Ozogul, Moreno, & Reisslein, 2013[f]	primary/ middle school	158	0.20	problem-solving	physics	low/medium	self-paced	static	many	visual	strong	no
	primary/ middle school	139	-0.14	problem-solving	physics	high	self-paced	static	many	visual	strong	no
Johnson, Butcher, Ozogul, & Reisslein, 2014	university	107	-0.08	far transfer	physics	low/medium	self-paced	static	few	discursive	strong	no
Johnson, Ozogul, & Reisslein, 2014[f]	primary/ middle school	127	0.34	problem-solving	physics	low/medium	self-paced	static	many	visual	strong	no
		123	0.09	problem-solving	physics	high	self-paced	static	many	visual	strong	no

[e] Total values of the factor representation type (abstract and contextualized representation) used.

[f] Signaling factor levels APA signaling and arrow signaling were averaged by calculating the mean weighted by the corresponding sample sizes and by calculating the pooled variance.

(continued on next page)

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
Kalyuga, Chandler, & Sweller, 1999, Exp. 2	vocational training	16	0.59	transfer (multiple choice task)	mechanics	low/medium	self-paced	static	many	visual	strong	no
Kühl, Scheiter, & Gerjets, 2012	university	100	0.31	transfer	physics	low/medium	system-paced	static[g]	few	visual	weak	no
	university	50	0.50	transfer	physics	low/medium	system-paced	dynamic	few	visual	weak	no
Lin, 2011	university	42	0.20	comprehen-sion	biology	low/medium	self-paced	dynamic	few	visual	weak	no
	university	84	0.17	comprehen-sion	biology	low/medium	self-paced	dynamic	few	visual	weak	yes, prompts[h]
Mason, Pluchino, & Tornatora, 2013	primary/ middle school	36	0.28	transfer	physics	low/medium	self-paced	static	few	discursive	weak	no
Mautone & Mayer, 2001, Exp. 3[i]	university	46	0.38	transfer	physics	low/medium	system-paced	dynamic	few	visual	weak	no

[g] The visualization type factor levels static-sequential and static-simultaneous were averaged by calculating the mean weighted by the corresponding sample sizes and by calculating the pooled variance.

[h] The prompting factor levels predictive and reflective prompts were averaged by calculating the mean weighted by the corresponding sample sizes and by calculating the pooled variance.

[i] Additionally included text signals like additional headings, paragraphs and preview summary.

(continued on next page)

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
McTigue, 2009[J]	primary/ middle school	40	-0.23	comprehen-sion	biology/ physics	low/medium	self-paced	static	many	discursive	strong	yes, parts diagram
	primary/ middle school	40	0.07	comprehen-sion	biology/ physics	low/medium	self-paced	static	many	discursive	strong	yes, steps diagram
	primary/ middle school	40	0.16	comprehen-sion	biology/ physics	low/medium	self-paced	static	few	discursive	strong	yes, parts & steps diagram
Ozcelik, Karakus, Kursun, & Cagiltay, 2009	university	52	0.41	transfer	biology	low/medium	self-paced	static	many	visual	strong	no
Ozcelik, Arslan-Ari, & Cagiltay, 2010	university	40	0.33	transfer	physics	low/medium	system-paced	static	few	visual	strong	no
Paik & Schraw, 2013	university	33	0.30	transfer	physics	n/a	system-paced	static	many	visual	strong	yes, arrows & step-by-step presentation
	university	32	0.11	transfer	physics	n/a	system-paced	dynamic	many	visual	strong	yes, arrows & step-by-step presentation

[J] For each learning material the effect size r was calculated and averaged by computing the mean effect size with the Meta_basic_r SPSS script provided by Field and Gillett (2010).

(continued on next page)

Study	Sample characteristic	Sample size *n* of pair-wise comparison	Effect size *r*	Performance measure	Learning material domain	Domain-specific prior knowledge	Pacing of the materials	Pictorial format	Multimedia mapping requirements	Distinctiveness of MIS	Type of control group	Additional instructional support
Scheiter & Eitel, 2015, Exp. 1[k]	university	55	0.35	text-diagram integration	biology	low/medium	self-paced	static	many	mixed discursive and visual	weak	no
Scheiter & Eitel, 2015, Exp. 2[k,l]	university	53	0.27	text-diagram integration	biology	low/medium	self-paced	static	many	mixed discursive and visual	weak	no
Seufert, 2003	university	17	0.29	comprehen-sion	biology	low/medium	self-paced	static	many	discursive	strong	no
	university	17	-0.07	comprehen-sion	biology	high	self-paced	static	many	discursive	strong	no
Tabbers, Martens, & Merriënboer, 2004	university	56	0.05[m]	transfer	educational psychology	low/medium	self-paced	static	few	visual	weak	no
	university	55	0.18[n]	transfer	educational psychology	low/medium	self-paced	static	few	visual	weak	no
Van Oostendorp, Beijersbergen, & Solaimani, 2008	university	30	0.12	transfer	biology, physics[o]	low/medium	self-paced	dynamic	many	visual	strong	no

[k] Additionally included text signal paragraphing.

[l] Condition with unmatched signals was excluded since it was not focus of the meta-analysis.

[m] The instructional text was presented in written form.

[n] The instructional text was presented as a spoken narration.

[o] Transfer performance of both learning materials was averaged by calculating the mean weighted by the corresponding sample sizes and by calculating the pooled variance of the animation and animation with focus conditions.

6.3.2 Overall Signaling Effect

The distribution of the derived effect sizes (see Figure 6) indicates that there were 38 out of 45 positive effect sizes (84.4%), suggesting that signaled multimedia material was more effective in fostering comprehension performance than non-signaled material. Taken together, there was a small-to-medium overall signaling effect in favor of signaled compared to non-signaled multimedia learning material: r = .17, 95% CI [.11, .22][3]. The estimated variance in the population as given by the Fisher-transformed correlation was $\tau = 0.014$.

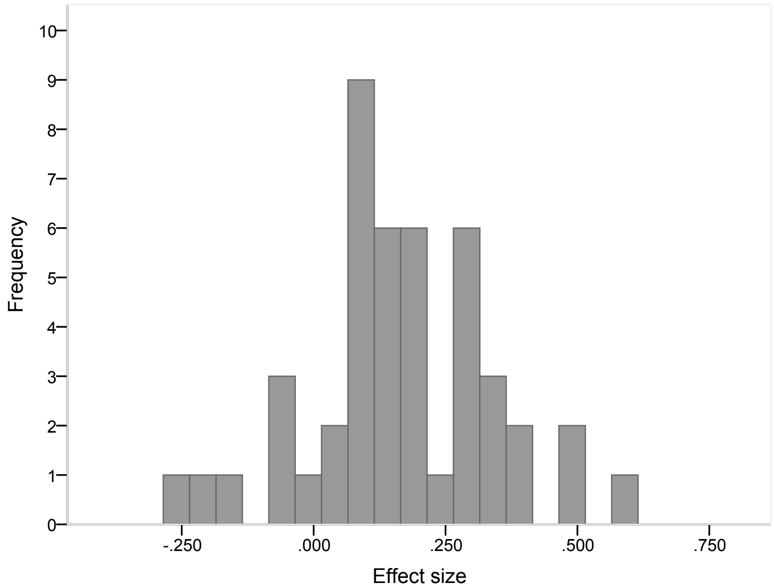


Figure 6. Distribution of effect sizes *r* included into the meta-analysis.

Since only published studies were included into the meta-analysis, it was necessary to estimate the publication bias. The *fail-safe N* in this meta-analysis was 937, meaning that there would need to be at least 937 (unpublished) studies with a non-significant signaling effect to render the overall signaling effect identified in the present meta-analysis non-significant. According to a guideline by Rosenthal (1991) the result of the present meta-analysis seems to be robust to the publication bias since the *fail-safe N* exceeds 5 x *k* +10 (5 x 45 + 10 = 235 < 937). In addition, the Begg and Mazumdar (1994) rank correlation was non-significant, $\tau(N=45) = 0.03$, p = .791, indicating that a publication bias was unlikely to be present in the data.

[3] Without correcting for extreme values the overall signaling effect was larger: r = .19, 95% CI [.10, .27].

To test whether the signaling effect would differ depending on the type of control group used in the studies, a moderator analysis was run using only type of control group as moderator. The result showed that the type of control group marginally significantly moderated the signaling effect: $\chi^2(1)= 3.66$, $p = .056$. The signaling effect tended to be larger for studies using weak control groups, $r = .23$, 95% CI [.16, .29] compared with stronger control groups, $r = .13$, 95% CI [.05, .20], although this difference was not significant.

The set of effect sizes was overall heterogeneous ($Q = 75.86$, $df = 44$, $p = .002$). Therefore, in the next step moderator analyses were conducted addressing those moderators of the signaling effects that were derived from theory and empirical research[4].

6.3.3 Impact of Moderator Variables

The hypothesized moderator variables were entered into a moderator analysis. The signaling effect was significantly moderated only by the prior knowledge of the learners (low/medium vs. high), $\chi^2(1)= 12.27$, $p < .001$, revealing that learners with a low to medium level of prior knowledge, $r = .19$, profited more from MIS than learners with a high level of prior knowledge. There was even a negative effect of MIS for the HPK learners, $r = -.08$, which was, however, not significantly different from zero, $z = 0.93$, $p = .352$.

A moderating effect was also hypothesized for the pacing of the materials (self-paced vs. system-paced), $\chi^2(1)= 1.59$, $p = .208$, the pictorial format (static vs. dynamic), $\chi^2(1)= 0.02$, $p = .884$, the mapping requirements (few vs. many), $\chi^2(1)= 0.57$, $p = .451$, and the distinctiveness of MIS (discursive vs. visual), $\chi^2(2)= 2.94$, $p = .230$. These assumptions could not be confirmed. Additionally, separate moderator analyses were conducted for each boundary condition to have an isolated view on their influence on the signaling effect. The results show that again the prior knowledge of learners significantly moderates the signaling effect, $\chi^2(1)= 11.15$, $p = .001$, as did the pacing of the materials, $\chi^2(1)= 4.92$, $p = .027$. MIS were more effective when system-paced materials were used, $r = .27$, in contrast to self-paced

---

[4] Two effect sizes had to be excluded for the moderator analysis since the information about the prior knowledge level of the participants in the study by Paik and Schraw (2013) was not given.

materials, *r* = .13. However, it was chose to focus on the results of the moderator analysis that included all moderator variables simultaneously, as it was chose to follow a rather conservative approach in general. Table 3 reveals the effect sizes and confidence intervals for all hypothesized moderator variable levels.

Table 3

Effect sizes and confidence intervals for moderator categories

Moderator variable and level	Number of effect sizes *k*	Number of participants *n*	Effect size *r*	95% CI for *r*
Domain-specific prior knowledge[a,b]				
Low/medium	39	2,076	0.19	[0.14, 0.24]
High	4	323	-0.08	[-0.24, 0.09]
Pacing of materials				
Self-paced	34	1,910	0.13	[0.07, 0.19]
System-paced	11	554	0.27	[0.16, 0.37]
Pictorial format				
Static	34	1,983	0.15	[0.10, 0.21]
Dynamic	11	481	0.20	[0.06, 0.33]
Multimedia mapping requirements				
Few	19	1,073	0.19	[0.11, 0.27]
Many	26	1,391	0.15	[0.07, 0.22]
Distinctiveness of MIS				
Discursive	11	575	0.08	[-0.01, 0.17]
Visual	27	1,599	0.19	[0.11, 0.27]
Mixed visual and discursive	7	290	0.20	[0.08, 0.31]

[a] Variable moderates the overall estimated signaling effect in the population.
[b] Since the level of prior knowledge could not be determined for a study by Paik and Schraw (2013), two related effect sizes had to be excluded the from the analysis.

One concern when conducting moderator analyses in meta-analyses is that there are confounds between the various moderators so that found effects cannot be unambiguously interpreted as being due to one particular factor. For instance, in the present case it could be that studies with LPK learners are at the same time studies using weak control groups, making it impossible to disentangle the effects of those two factors. Therefore, it was tested whether there were significant relationships between each of the hypothesized moderator variables by means of Pearsons's chi-square test (Pearson, 1900), and Fisher's exact test (Fisher, 1922) when there were too few cases in one category. The results are shown in Table 4. A significant result was obtained for the relationship between pacing of materials and multimedia mapping requirements, suggesting that studies with self-/system-paced materials were at the same time studies with few/many mapping requirements. In addition, the relationship between type of control group and multimedia mapping requirements turned out to be marginally significant, suggesting that studies with weak/strong control groups were at the same time studies with few/many mapping requirements. Importantly, the moderator variable prior knowledge and type of control group were independent from each other, thereby indicating that the finding of a larger signaling effect for studies with LPK learners is not an artifact of these studies having weak control groups at the same time.

Table 4

Chi-square test results for boundary conditions and type of control condition

		Domain-specific prior knowledge (*df* = 1)	Pacing of materials (*df* = 1)	Pictorial format (*df* = 1)	Multimedia mapping requirements (*df* = 1)	Distinctive-ness of signals (*df* = 2)	Type of control group (*df* = 1)
Domain-specific prior knowledge	χ^2	-	1.17	0.01	3.49	0.93	2.61
Pacing of materials	χ^2	1.17	-	2.86	5.21*	3.91	0.26
Pictorial Format	χ^2	0.01	2.86	-	0.18	4.52	0.04
Multimedia mapping requirements	χ^2	3.49	5.21*	0.18	-	3.16	3.47$^{(*)}$
Distinctive-ness of signals	χ^2	0.93	3.91	4.52	3.16	-	0.64
Type of control group	χ^2	2.61	0.26	0.04	3.47$^{(*)}$	0.64	-

* $p < .05$. $^{(*)}$ $p < .10$.

6.4 Summary & Discussion

The present meta-analysis sought to determine the overall size of the signaling effect in multimedia learning along with its potential boundary conditions. Following the CTML (Mayer, 2014b), a positive effect of MIS across studies was expected (Hypothesis 1), because highlighting correspondences between multiple media (i.e., text and picture) should support integration of information into a coherent mental representation, which in turn should foster comprehension. The results revealed a small-to-medium overall signaling effect in multimedia learning, $r = .17$, suggesting that across a multitude of studies MIS indeed fostered comprehension by facilitating integration. Thus, results yield support for multimedia signaling as an effective design principle to optimize multimedia instructions (cf. signaling principle; van Gog, 2014). Moreover, the present results suggest that there is a positive signaling effect, which is not just an artefact of weak control conditions ("lousy"; cf. Schwonke, Renkl et al., 2009). Even though the signaling effect was found to be stronger with weak than with strong control groups ($r = .23$ vs. $r = .13$), a significant positive effect of signaling remained even with strong control groups, suggesting that the effect is true and not just an artefact of comparing signaling with instructional conditions that would be unlikely to be found in educational practice due to their poor design.

Referring to the ability-as-compensator hypothesis (Mayer & Sims, 1994) and the ERE (Kalyuga et al., 2003), instructional support in the form of MIS was assumed to be more beneficial for LPK than for HPK learners. Supporting Hypothesis 2, prior knowledge turned out to moderate the signaling effect significantly. LPK learners profited more from MIS ($r = .19$) than HPK learners ($r = -.08$). This suggests that MIS as a means to facilitate integration is beneficial especially for LPK learners, who would not be able to integrate information adequately without this type of instructional support. By contrast, for HPK learners signaling tended to show a negative, albeit non-significant, effect size for comprehension outcomes on the descriptive level, suggesting that they were able to integrate information adequately without needing signals. In fact, three of the four multimedia studies investigating signaling effects with HPK learners (Arslan-Ari, 2013; Johnson, Ozogul, Moreno, & Reisslein, 2013; Johnson, Ozogul, & Reisslein, 2014; Seufert, 2003) found negative effects of signaling, and the remaining one was positive but very small.

Because very few of the studies investigated signaling with HPK students, results must be treated with some caution. However, the non-significant effect size suggests that MIS were not helpful for learners with HPK. Referring to the ERE

(Kalyuga et al., 2003), additional instructional support in the form of signaling might even interfere with pre-existing mental representations. For instance, MIS indicate which elements correspond to each other in a text and in a picture. Thus, when processing this element in the text, a visual MIS seeks to guide attention towards the corresponding element in the picture (e.g., by printing the elements in the same color). Such recommendations for an optimal processing sequence may interfere with HPK learners' habitual processing of this instructional material, and hence, these learners might be confused or distracted by the MIS, potentially leading to worse performance.

Nevertheless, one could also argue that if learners realize that they do not profit from MIS, they should be able to ignore them, and therefore should not demonstrate weaker overall performance. Such results have been found in a study by Scheiter and Eitel (2015) who presented LPK learners with mismatched signals (i.e., signals that highlighted alleged text-picture correspondences where there were none). Learners only initially attended to these mismatched signals but ignored them once they recognized that they did not provide helpful information for learning. Accordingly, learning outcomes were also not affected by these mismatched signals (Scheiter & Eitel, 2015). Although Scheiter and Eitel (2015) used an LPK sample, results might also be applicable for learners with HPK. It can be assumed that, similar to mismatched MIS for LPK learners, MIS that interfere with already present mental representations of HPK learners are ignored by them during the course of learning and should therefore not hinder learning. However, this is only a tentative conclusion because there are only few studies in the field investigating the specific interaction between the effectiveness of MIS for learning outcomes and the moderating role of prior knowledge. Moreover, there are even fewer studies investigating effects of MIS and prior knowledge on a process level. Future research should thus explicitly address the differences between LPK and HPK learners in processing MIS with regard to learning outcomes and visual attention parameters to gain insight into processing differences.

The other hypothesized moderator variables - pacing of the materials, pictorial format, multimedia mapping requirements and distinctiveness of MIS - turned out to not moderate the signaling effect. Therefore, Hypotheses 3 to 6 had to be rejected.

However, a look at separate meta-analyses for the different levels of each moderator variable indicated deviations between effect sizes (see Table 3). Importantly, these results are not interpreted as confirmatory evidence but rather as indications for the potential influence of boundary conditions on the multimedia signaling effect. Such indications do provide important insights regarding directions for future studies that aim at more systematically addressing potential moderations among factors within one study design, compared to what can be achieved with a meta-analytic review.

Regarding the pacing of the materials, there was small-to-medium effect size for signaling for system-paced instruction (r = .27) in contrast to a small effect for self-paced learning (r = .13). A separate moderator analysis that included only the pacing of materials as a variable showed a significant moderation of the signaling effect. Because the strict timing of system-paced presentations forces learners to attend to relevant information at the right time, it was initially expected that signaling would be more effective here than during self-paced learning (cf. Ginns, 2005; Tabbers et al., 2004). Apparently, while the pattern of results is in the right direction and the separate moderator analysis reveals a moderation, it is not (yet) strong enough to reveal a significant moderation when it is included simultaneously in a more conservative moderator analysis along with all other hypothesized boundary conditions.

Regarding the influence of the distinctiveness of MIS, the effect size was very small and not significant for studies that used discursive MIS only (r = .08), whereas the usage of visual MIS (r = .19) or both discursive and visual MIS (r = .20) led to small-to-medium effect sizes. The relative size of these effects is thus in line with our initial assumptions. This result suggests that MIS differ in terms of their distinctiveness and how accessible they are for the learner (cf. realization property; Lemarié et al., 2008). Future research should address the distinctiveness of MIS more explicitly, using eye tracking methodology. This methodology lends itself nicely to testing the assumption that varying salience among discursive and visual signals causes differences in visual attention that in turn might explain differences in learning outcomes (cf. Jamet, 2014; Ozcelik et al., 2010; Scheiter & Eitel, 2015).

Furthermore, the results revealed only small differences in the effect sizes for different pictorial formats (static vs. dynamic) and multimedia mapping requirements (few vs. many). In both cases, the differences in effect sizes are too small to allow for any further interpretation. Consequently, based on the present results, it can be concluded that the signaling effect is the same for static and dynamic visualizations, which is in line with previous findings from Kühl et al. (2012) and Höffler and Leutner (2007). Importantly, there is a small-to-medium significant effect for both types of pictorial formats, which stands in contrast to earlier conclusions by de Koning et al. (2009), who found harmful effects of signaling in dynamic visualizations. Potentially, these earlier negative results have been compensated by more effective means of signaling in dynamic visualization that were developed more recently (Boucheix & Lowe, 2010).

At the same time, multimedia mapping requirements do not seem to matter for the signaling effect, which is somewhat surprising given that this variable seems to be so closely related to what signaling is assumed to foster, namely, the identification of correspondences. Different from our initial assumptions, multimedia integration signaling was equally beneficial in studies that used multimedia material including only few mapping requirements (r = .19) and for those requiring many mappings (r = .15). However, there might be methodological reasons for this finding, as discussed in the following chapter.

6.5 Limitations and Conclusions

In general, a meta-analysis is limited by the inclusion criteria that are applied and by the empirical set resulting from these criteria. Our results therefore can only be interpreted within the frame of MIS, multimedia learning material and comprehension performance. The present results confirm the beneficial effects of signaling in multimedia learning across a broad range of studies, which were selected and analyzed in a systematic fashion. Moreover, they specify the situations under which signaling works best, that is, for learners with low levels of prior knowledge.

One limitation of meta-analyses is that they can always be only as good as the studies on which they are based; moreover, their soundness heavily relies on the information documented by authors. This problem becomes particularly evident when trying to code material-related aspects as study features such as mapping requirements in the present meta-analyses. The included studies often reported only very little information about the material used and examples given might not always have reflected the characteristic of all of the multimedia material. Therefore, it might be that the expert rating of mapping requirements did not fully reflect the actual mapping requirements imposed by the materials. Thus, the rating of mapping requirements may not only have been subjective but also to some degree speculative, since it relied on too little information, which may explain why this moderator did not show any effects. I thus urge authors to provide more comprehensive descriptions of their materials and samples in the future and to use the opportunity to upload materials onto the journals' online repositories, since this would immensely improve the validity of generalizing statements.

Because the present meta-analysis is mostly based on lab experiments, the high internal validity of the single studies warrants an unambiguous interpretation in favor of the signaling effect. The downside of this approach is that results may not be generalized in a one-to-one fashion to situations in the field. In the present case, the overall positive effect of signaling results from the fact that the majority of studies tested LPK learners rather than HPK learners. Testing LPK students makes sense from an experimental point of view, because knowledge gains from learning with multimedia can be better investigated when there is much improvement possible in knowledge levels. However, one has to be careful in generalizing the study results from mainly LPK students, because in "real" learning situations in schools, universities or in informal settings, prior knowledge levels are not always low, and hence, MIS may not always be effective. Therefore, prior to deciding whether to implement MIS in learning environments, teachers or instructional designers should be aware of the learners' existing knowledge level.

Apart from prior knowledge, there were no further significant moderators of the signaling effect. However, future research should have yet another look into the role of the distinctiveness of MIS and pacing, which revealed substantial effect size differences whereby the latter moreover was indeed a significant moderator in a separate moderator analysis. The existing number of studies falling into the different categories of these moderators might still be too low to allow for any firm conclusions regarding their irrelevance for the signaling effect.

Despite these limitations, the results of the meta-analysis unambiguously support signaling as a design principle for multimedia learning, particularly for learners with low prior knowledge.

To sum up, the current meta-analyses can make statements only regarding the instructional effectiveness of MIS mainly based on lab studies; it leaves open the question if multimedia signaling is also effective under ecologically valid conditions and how multimedia signaling works for learners with different prior knowledge. Consequently, in the following studies included in the present dissertation the effectiveness of MIS was tested under ecologically valid conditions in schools (Study 2) as well as by means of eye tracking methodology in a lab study (Study 3) both with secondary higher education students with differing levels of prior knowledge, ecologically valid learning material conformed to the curriculum related to introductory chemistry education and a strong control condition.

7. Study 2: The Effects of Multimedia Signaling in an Ecological Valid Context[5]

The goal of the present study was to test the moderating role of prior knowledge on the signaling effect for multimedia learning in an ecologically valid context with a strong control group (Research Questions 1 as well as 3 and 4). Moreover, ecologically valid learning material conformed to the curriculum was used. Thereby, limitations of the set of studies included in the meta-analysis (Richter, Scheiter, & Eitel, 2016) were addressed (see chapter 6.5). Students learned a topic of introductory chemistry education with an eBook in their regular chemistry class. Two versions of the eBook were distributed randomly in multiple classrooms: (a) a MIS- version with text signals and labels in the picture corresponding to the text and (b) an MIS+ version with additional MIS like color coding and deictic references (example see Figure 7). Students learned with the eBook in two learning sessions lasting 90 minutes each. Before the learning sessions their prior knowledge was assessed; their cognitive load and learning outcomes were measured immediately after the learning sessions. Subjective cognitive load ratings were assessed in order to shed light on the underlying processes related to cognitive load for EREs (Research Question 5).

7.1 Hypotheses

Against the backdrop of the meta-analysis (see chapter 6) and related research concerning other instructional support measures (cf. Kalyuga, 2007), an ERE with respect to learning outcomes was expected: LPK learners but not HPK learners should show better learning outcomes when learning with additional multimedia integration signals (MIS+) compared with a strong control condition including text signals such as headings, preview sentences and bold face as well as corresponding labels in texts and pictures (MIS-) (Hypothesis 1). Because MIS are

[5] Richter, J., Scheiter, K., & Eitel, A. (revise and resubmit). Signaling text-picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology.*

assumed to aid LPK learners in constructing a meaningful mental model from text and picture, their effect should not be limited to recall but also be evident in measures of comprehension. In line with results obtained by Kalyuga (2007) a full reversal for HPK learners was expected in that additional MIS should lead to lower learning outcomes for them. Students with a medium level of prior knowledge (MPK) should perform equally well in both signaling conditions. The multimedia signaling effect was expected to be in line with the ERE especially for comprehension performance, since MIS aim at establishing text-picture correspondences, and this process underlies deeper learning with multimedia (cf. Mayer, 2014a; Richter et al., 2016). However, also recall measures and a measure of misconceptions related to the content of the materials are reported in order to provide a comprehensive picture of evidence.

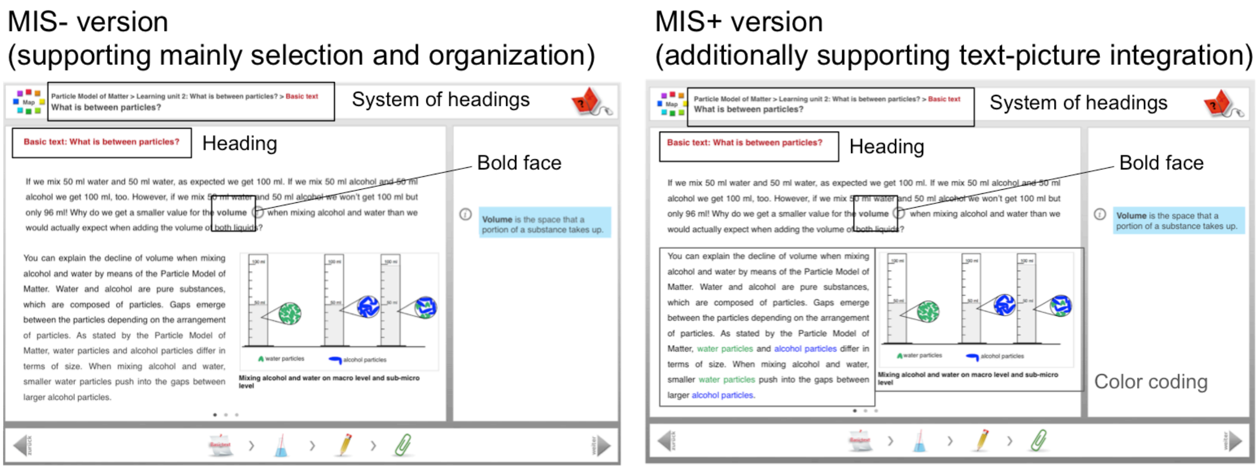


Figure 7. Example of an eBook page on the left with the text signals headings and bold face (MIS- condition) and on the right additionally with the multimedia integration signal color coding (MIS+ condition).

Different measures of cognitive load were administered to approach the question of whether potential detrimental effects of multimedia signaling for HPK learners would be due to a reduction of GCL or an increase in ECL. With respect to ECL, we hypothesized that MIS (MIS+ condition) support LPK learners by reducing the need for them to search for text-picture correspondences by themselves, which would be too taxing for them. Thus, LPK learners were expected to experience less ECL in the MIS+ condition than in the MIS- condition. For MPK students no differences were expected between the signaling conditions because, on the one hand, they have a certain level of prior knowledge that helps them to process material without further instructional integration support (MIS-). On the other hand, they might also be supported by MIS, at places where prior knowledge is still lacking (MIS+).

Finally, HPK students were assumed to experience more ECL when learning with the MIS+ material due to the need of processing redundant information (Kalyuga et al., 1998). HPK learners might be able to identify text-picture correspondences on their own by using their prior knowledge. Hence, the MIS- condition, which supports only the selection and organization of verbal (and pictorial) information into mode-specific models should reduce ECL for HPK learners, since they do not have to process redundant information associated with the explicit highlighting of text-picture correspondences in the MIS+ condition (Hypothesis 2).

With respect to GCL, LPK students were expected to experience more GLC in the MIS+ condition than in the MIS- condition, because MIS were assumed to help them elaborate text-picture correspondences. The additional support of the integration process in the MIS+ condition is supposed to induce elaboration processes of information from texts and pictures. Text signals were supposed to not trigger these elaboration processes to the same extent as MIS, because the prior signals are supposed to mainly support the selection and organization of text and picture information. For MPK students we again expected no difference between the signaling conditions due to the same reasoning as for the ECL measure. For HPK students, it was expected that GCL would be higher when learning with the MIS- than with MIS+ version. When guiding information for the integration of verbal and pictorial information is absent, HPK students should engage in deeper processing of the materials by applying their prior knowledge to resolve gaps and identify text-picture correspondences on their own. When MIS are present, HPK learners might refrain from deeper processing because the salience of text-picture correspondences does not stimulate them to engage in elaboration processes (cf. McNamara et al., 1996) (Hypothesis 3).

Testing the hypotheses for cognitive load includes a major obstacle related to the measurement of (different sources of) cognitive load (De Jong, 2010; Schnotz & Kürschner, 2007). In the present study learners were asked either for the difficulty experienced during learning with the materials (successfully used by Cierniak, Scheiter, & Gerjets, 2009; and adapted from Kalyuga et al., 1998) or for the mental effort they had voluntarily invested in the learning task (a slightly modified version of the very popular item initially introduced by Paas, 1992). The first item is supposed to measure ECL because it is related to the difficulty in learning with the materials, whereas the second item presumably addresses GCL, which should be corroborated by either negative (ECL) or positive (GCL) correlations with learning outcomes. In general, attempts to separately measure ECL and GCL through subjective ratings have been successful in some studies (e.g., Cierniak et al., 2009; DeLeeuw & Mayer, 2008), but have failed in others using the same measures as in Cierniak et al. (2009) (Kühl, Scheiter, Gerjets, & Edelmann, 2011). Furthermore, several issues related of subjective cognitive load ratings are reported in the literature (for detailed information see chapter 4.3.1). That is to say, I was fully aware that various measurement problems regarding cognitive load exist (cf. chapter 4.3.1).

Because in the present study cognitive load ratings could be assessed only at after the posttest for practical reasons, they might reflect not only load experienced during learning, but also during knowledge retrieval (Van Gog & Paas, 2008); moreover, their absolute levels should not be interpreted as an average of cognitive load experienced during learning (Schmeck et al., 2014). Because of these measurement problems, the question of whether a full reversal would be due to an increase in ECL or a decrease in GCL was treated as secondary only. Instead the study focused on the effects of signaling and the moderating role of prior knowledge on learning outcomes.

7.2 Method

7.2.1 Participants and Design

One hundred eighty-three students in 7 classes of 3 schools in 2 administrative districts from a southern federal German state participated in the study. Thirty-one students were excluded because they did not attend either a learning or a testing session. Because the aim was to investigate the effectiveness of MIS, it was necessary to make sure that students had opened the majority of eBook pages that contained signals. Thus, log file data was checked that showed the eBook use and only students who had opened all signaled eBook pages except the last one were included. The last signal was on the very last page of the last learning unit, which some students did not reach during the assigned time. This approach resulted in the exclusion of another 22 students from the analyses. Another 3 students had to be excluded due to refusal of work, technical problems, or insufficient knowledge of the German language.

To this end, 56 students were excluded from data analysis, which resulted in overall 127 students. The students (M_{age} = 14.36 years; age range 14 - 17 years; 47 female) were in eighth grade in secondary higher education (Gymnasium). The aim was to have students with heterogeneous prior knowledge in the study by including classes with different prior knowledge levels. Based on information provided by the chemistry teachers, 3 classes already had acquired either fundamental or broader knowledge about the eBook topic in class, whereas 4 classes learned about the topic for the first time in the context of the current study. Students were randomly assigned to the signaling condition within each class resulting in 62 students learning with the MIS- and 65 students learning with the MIS+ eBook version.

7.2.2 Instructional Materials

Students learned with the digital textbook *eChemBook* individually on a Lenovo ThinkPad T530 with a 15.6-inch monitor. The eChemBook was designed within a project funded by the German Research Foundation in cooperation with researchers from science education as well as a textbook publisher company and a manufacturer of interactive whiteboards, based on evidence from science education, text comprehension, and multimedia research. For example, multimedia design principles (Mayer, 2014b), design guidelines for instructional text (e.g., Hartley, 2004) and evidence from science education were considered for the design of the eChemBook. For example, this evidence recommended to explicitly address students' misconceptions about scientific phenomena (Mason, 2001).

The topic that the students were instructed to learn was the Particle Model of Matter, consisting of 6 learning units distributed across 58 pages (5,911 words, 37 static and 17 dynamic pictorial representations [video, animation, simulation]). This topic is part of the German curriculum in introductory chemistry education. In each learning session participants studied 3 learning units, which resulted in 26 pages in session 1 (learning units 1-3: “Introduction to the Particle Model of Matter”, “What is between Particles?” and “Diffusion”) and 32 pages in session 2 (learning units 4-6: “Pressure”, “Aggregate States” and “Dissolution”). Only for learning session 1, students were instructed to work through the introductory learning unit first. After that, students could decide on their own about the order of the units within the learning sessions. Each learning unit of the eChemBook consists of a motivational teaser, a basic text containing all relevant information, a related experiment with instructions, a video displaying the experiment and tasks to describe and interpret the result of the experiment, interactive drawing and drag-and-drop exercises, and a short summary. Due to time limitations and the aim to keep the setting between classes as constant as possible, students did not conduct the experiments in class but could work on the experiments by using the videos in the eBook.

In order to be able to compare an eBook version with MIS to a fair control condition, the eChemBook was modified with regard to the implementation of MIS:

1. The MIS- version included common text signals such as headings, preview sentences, and bold face for relevant terminology, but also corresponding labels between text and picture as a basic MIS. With regard to Schwonke, Renkl et al. (2009), who argue in favor of strong control conditions that reflect rather ecologically valid settings, we made sure that students in the MIS-condition could also unambiguously identify elements in the picture that were mentioned in the text and were also provided with basic text signals that are most commonly implemented in educational materials.
2. The MIS+ version additionally contained the MIS color coding, deictic references, and additional corresponding labels. In total 24 additional signals were implemented in the MIS+ eBook version. The number of implemented signals in the 6 learning units of the topic varied between no additional signals in Units 4 to 10 signals in Unit 6. They were only added where they could provide additional information about how to process text-picture relations. That means, all text-picture combinations that were not already signaled sufficiently in the MIS- condition were provided with an additional MIS like color coding, deictic reference, or additional corresponding labels in the MIS+ condition.

Thus, the signaling conditions differed in the signals used (mainly text signals versus text signals plus MIS) and with that also in their signaling functions. Whereas text signals were aimed at supporting the selection and organization process, MIS additionally supported the integration of verbal and pictorial information.

7.2.3 Measures

Pre-test. Before students learned with the eBook, they answered a paper-based pre-test that assessed their prior knowledge, as well as the control variables reading comprehension, domain- specific academic self-concept and interest. The control variables were assessed only to make sure that the two signaling groups were equivalent with respect to these measures. For the assessment of the prior knowledge of students, 15 verification items were partly constructed by the authors and partly adapted from 2 different German language sources (ZPG-Chemie, 2011; Hollstein, 2001). Two items had to be removed from the measure due to negative corrected item-total correlations. Cronbach's α for the prior knowledge measure was rather low with .42, which can be attributed to the conceptualization of prior knowledge by different misconceptions related to the subtopics. Nine different misconceptions related to the Particle Model of Matter were identified from literature (e.g., Aydeniz & Kotowski, 2012; Nakhleh, 1992; Yezierski & Birk, 2006) and used in the prior knowledge test (see Appendix A). Our approach to base the prior knowledge test on common misconceptions is in line to the approach followed by the American Association for the Advancement of Science (AAAS) in their Project 2061 (e.g., Herrmann-Abell & DeBoer, 2011). They developed and tested about 700 items relating to different science topics from the U.S. curriculum based on common ideas and misconceptions students might have related to a topic.

Per misconception, 1 to 2 items were used in the pre-test. Examples of misconceptions about the Particle Model of Matter are that particles have a color and that particles expand when they are heated. However, the Particle Model of Matter does not make any statements about the color of particles and the correct scientific understanding is that particles move faster when heated. Particles never change in size or shape. Verification items related to these example misconceptions are: "Individual sulfur particles are yellow." or "When gas particles are heated, they expand.” Students indicated whether they thought an item was true. When an item was identified correctly as being true, it was scored with 1 point, resulting in a maximum score of 13 points. The sum of points was transformed into percentage correct for easier interpretation.

In order to assess reading comprehension, a standardized German reading test (LGVT 6-12; Schneider, Schlagmüller, & Ennemoser, 2007) was used. Students read a text (3.5 pages, 1,727 words) within a time limit of 4 minutes. The text included 23 brackets that each contained 3 filler words. Students were instructed to underline the word that fit best in the context of the text. Reading comprehension performance was determined as the number of correctly chosen filler words.

The domain-specific academic self-concept was assessed by means of 5 items developed by Schanze (2002). An adjusted version of these items was used that were positively phrased, introduced by Grüß-Niehaus (2010). Sample items are: “I am simply talented for chemistry.” or “I can solve even difficult tasks in chemistry.” and were rated on a 4-point Likert-type scale ranging from “I do not agree at all” to “I completely agree”. Cronbach’s α for this test was .86.

The domain-specific interest was assessed by means of 5 items by Wilde, Bätz, Kovaleva, and Urhahne (2009), which were adapted to the science context. The items were for example: “I like reading science texts” or “I am interested in learning new things in science”. The items were also answered on a 4-point Likert scale ranging between “I do not agree at all” and “I completely agree” (Cronbach’s α = .88).

Post-test. In the paper-based post-test, students’ knowledge about the Particle Model of Matter was assessed. The item selection and development was conducted with respect to the contents of the eChemBook chapter and the related misconceptions that students might have regarding the contents. The knowledge test consisted of 15 verification items, 29 multiple-choice items and 8 drawing items covering all contents of the eBook topic. The items were partly constructed by the authors and partly adapted, and where necessary they were translated into German from different sources (Hollstein, 2001; Mulford & Robinson, 2002; Petermann, Friedrich, & Oetken, 2009; Yezierski & Birk, 2006; ZPG-Chemie, 2011). There were two categories of items: recall and comprehension items. Recall items queried for knowledge that was directly included in the text or visualizations of the eBook. Comprehension items queried for relations between concepts that required transfer of concepts included in the learning material to new contexts, which were not explicitly mentioned in the eBook. The comprehension performance of students was for example assessed by a question about whether particles can melt under certain conditions, like in gaseous substances. However, in the eBook no direct information about melting particles was included. Students had to relate information of different learning units of the eBook to come to the conclusion that particles never melt no matter what the current aggregate state is.

The verification items were equal to those used in the pretest ($n = 15$, Cronbach's $\alpha = .46$). The multiple-choice test assessed recall ($n = 19$, Cronbach's $\alpha = .79$) and comprehension ($n = 10$ items, Cronbach's $\alpha = .61$) performance. Students were instructed to decide which of the 5 options was correct and to select only 1 option per item. Correct multiple-choice items were scored with 1 point, resulting in a maximum of 19 points for recall and 10 points for comprehension performance.

To answer the drawing items, students were asked to draw phenomena at the level of model-based explanations and explain their drawings verbally. Answers were coded based on an author-developed coding schema by 2 raters. The interrater-reliability was Cohen's kappa = .84. Drawing items queried for recall (2 items, 16 possible points, Cronbach's $\alpha = .53$) and comprehension (6 items, 44 possible points, Cronbach's $\alpha = .70$) performance of students. Students earned points for correctly drawn and explained pivotal concepts such as using the same shapes and colors for particles that form matter. An example of a comprehension item was related to the learning units on aggregate states and diffusion. In the eBook, aggregate states were explained by means of the matter water. For the explanation of diffusion on the particle level, an example of perfume sprayed into a room was used. A related drawing comprehension item was: "One drop of bromine is trickled on the bottom of a cylinder. Then the cylinder is sealed. After a short time, the bromine spread in the cylinder while the drop is not visible anymore. Draw the two states of bromine at the model level". Students normally do not have knowledge about bromine at that stage of chemistry education. Nevertheless, they could answer this item by transferring their knowledge about aggregate states and diffusion to the new context. For data analysis, multiple-choice and drawing item scores of recall and comprehension performance were *z*-standardized separately and then the mean of the two item types (multiple-choice and drawing items) was calculated for comprehension and recall performance, in order to combine them into one recall and one comprehension outcome measure. Correspondingly, scores from the misconception measure (verification items) were also *z*-standardized to be better comparable to comprehension and recall performance.

ECL and GCL were each measured by one item. ECL was assessed by the question ”How difficult was it for you to understand the contents?”, which was adapted from Kalyuga et al. (1998) and used by Cierniak et al. (2009). The GLC question referred to the invested mental effort and read “How much effort did you invest to understand the contents?”, which was adapted from Paas (1992). Students subjectively rated these items on a 9-point rating scale (with 1 = *not difficult at all/no effort at all* to 9 = *extremely difficult/very much effort*).

Time-on-task. The overall time-on-task for both learning sessions was calculated based on log file data. In the log files, each opened eBook page was recorded together with the duration of the presentation (in milliseconds - until the students opened another eBook page). The sum of all durations for each eBook page was used to calculate the time-on-task in minutes for each student.

7.2.4 Procedure

Overall, 4 sessions took place: (a) pre-test session (45 minutes), (b) learning session 1 (90 minutes with a 5-minute break), (c) learning session 2 (90 minutes with a 5-minute break), and (d) post-test session (45 minutes). All sessions took place in chemistry lessons in schools but were managed by an experimenter. Within each learning session, students worked through the learning units in a self-paced manner. In the pre-test session, students were briefly introduced to the eChemBook project and filled in the pre-test. In addition, they were randomly assigned to a signaling condition. In learning session 1, they learned Units 1 to 3 of the eBook independently and at their own pace. In the following learning session, they learned Units 4 to 6. In the post-test session they were asked to fill in the post-test by first answering the verification items concerning misconceptions and then responding to the multiple-choice and drawing items, which were presented in an interspersed fashion. Finally, students answered the two cognitive load items. Since the study was conducted during regular chemistry lessons, the breaks between sessions differed between classes and ranged from 5 to 7 days between the learning sessions. In 3 of the 7 classes the post-test was conducted directly after the second learning session. The delay in the other classes ranged between 1 and 2 days.

7.2.5 Data Analyses

To ensure that students with varying levels of prior knowledge participated in the study, classes with either no knowledge or fundamental to broader knowledge about the eBook topic were included in the sample, based on information provided by the chemistry teachers. However, even if the content had already been taught in class, students were still expected to have varying levels of knowledge. Moreover, in classes where chemistry teachers stated not having taught the topic, parts of the contents might have been already known from physics or a cross-domain science subject. Therefore, rather than using the actual class membership as a quasi-experimental factor in our analyses, 3 prior knowledge groups were determined based on the 33rd and 67th percentile of the prior knowledge measure in order to classify students as low (LPK), medium (MPK), high prior knowledge (HPK) learners (cf. Seufert, 2003).

To test whether the aforementioned hypothesized pattern of results accurately described the data, effect coding was used for the analysis (Abelson & Prentice, 1997). This analysis allows parsimonious testing of whether a focal contrast, which reflects hypothesized relative group differences, fits the data or whether several alternative, orthogonal contrasts (residual contrasts) may have a better fit. Orthogonality of contrasts means that contrasts are not correlated with each other and thus each reflect a distinct prediction of patterns (Rosenthal, Rosnow, & Rubin, 2000). This procedure allowed to make conclusions about the dominant pattern in the data while ruling out an alternative pattern. Since there were k = 6 groups that needed to be compared (LPK, MPK, and HPK students learning either from a MIS- or MIS+ version of the eBook), in total 5 orthogonal contrasts (k-1) were defined (cf. Rosenthal et al., 2000).

As a result, 2 subsets of contrasts were entered into a regression model: (a) one focal contrast reflecting the hypothesized relative group differences and (b) a set of four residual contrasts describing alternative patterns. The hypothesized pattern of group differences describes the observed data best when the focal contrast significantly explains the data, whereas the set of residual contrasts do not. When the set of residual contrasts was significant, separate post-hoc tests for each residual contrast were used to reveal which of the residual contrasts significantly explained variance and thus contributed to the overall significant set of residual contrasts. Contrast analysis was chosen to directly test the hypotheses in order to prevent alpha-error inflation and a loss of statistical power through multiple group comparisons by means of post-hoc tests (Furr & Rosenthal, 2003; Hager, 2002; Judd, 2000; Rosenthal & Rosnow, 1985; Rosenthal et al., 2000). Moreover, the APA Guidelines for the use of statistical methods in psychology journals also recommend to use contrasts analysis instead of ANOVA and pair-wise multiple-comparison tests in order to test a specific hypothesis (Wilkinson & the Task Force on Statistical Inference, 1999).

It was hypothesized that the learning outcomes of MPK learners would not differ depending on the signaling condition. Therefore, both the MPK-MIS- group and the MPK-MIS+ group were coded with 0. LPK learners, however, were expected to profit from MIS, albeit with potentially lower overall performance scores than MPK and HPK learners due to their low prior knowledge. LPK-MIS- was therefore coded with -2 and LPK-MIS+ with -1. In the hypothesis, moreover a full ERE was postulated, which is why the HPK-MIS- group was coded with +2 and HPK-MIS+ with +1. This coding also implies that HPK learners were expected to perform better overall than MPK, and MPK learners better than LPK learners, which reflects the assumption that prior knowledge is positively correlated with learning outcomes (contrasts are listed in Table 5).

Table 5

Focal and residual contrasts

	LPK		MPK		HPK	
	MIS-	MIS+	MIS-	MIS+	MIS-	MIS+
Focal Contrast: Learning outcomes	-2	-1	0	0	2	1
Residual Contrast 1	0	0	1	-1	0	0
Residual Contrast 2	-1	1	0	0	-1	1
Residual Contrast 3	1	1	-2	-2	1	1
Residual Contrast 4	2	-4	0	0	-2	4
Focal Contrast: Extraneous cognitive load	1	-1	0	0	-1	1
Residual Contrast 1	0	0	1	-1	0	0
Residual Contrast 2	-1	1	0	0	-1	1
Residual Contrast 3	0	0	-1	-1	1	1
Residual Contrast 4	2	2	-1	-1	-1	-1
Focal Contrast: Germane cognitive load	-1	1	0	0	1	-1
Residual Contrast 1	0	0	-1	1	0	0
Residual Contrast 2	-1	1	0	0	-1	1
Residual Contrast 3	1	1	0	0	-1	-1
Residual Contrast 4	-1	-1	2	2	-1	-1

For the analysis of the cognitive load measures, also contrast analyses were used to test the related hypotheses directly. In order to test the ECL hypothesis, the LPK-MIS- group was coded with 1 and the LPK-MIS+ group with -1. Since no differences for MPK learners were expected, both signaling groups were coded with 0. It was expected that the HPK-MIS- group would experience lower ECL than the HPK-MIS+ group. Therefore, the MIS- group was coded with -1 whereas the MIS+ group was coded with 1. The focal contrast for the GCL hypothesis was coded with -1 for the LPK-MIS- group and 1 for the LPK-MIS+ group. Again both signaling groups for MPK learners were coded with 0 because no differences were expected. For the HPK-MIS- group a higher germane load was expected than for the HPK-MIS+ group, which is reflected by the coding of 1 (MIS-) and -1 (MIS+) (contrasts are listed in Table 1).

Regarding time-on-task there was no directed hypothesis. Therefore, time-on-task was analyzed exploratively by means of analysis of variance (ANOVA) followed by Bonferroni-corrected pairwise comparisons.

7.3 Results

Students were categorized based on their prior knowledge test scores into three groups based on percentiles: LPK, MPK, and HPK learners. Students who scored below the 33rd percentile (61.54% correct) of the prior knowledge test score were classified as LPK learners, students who scored between the 33rd and the 67th percentile (76.92% correct) were classified as MPK learners, and students who scored above the 67th percentile were classified as HPK learners.

7.3.1 Control Variables

It was tested whether there were differences between the signaling conditions with regard to the prior knowledge and the control variables reading comprehension, domain-specific academic self-concept and interest by means of ANOVAs. Despite randomization, students differed between the signaling conditions with respect to domain-specific interest, $F(1,125) = 5.45$, $p = .021$, $\eta_p^2 = .04$ (for means and standard deviations see Table 6).

Table 6

Means and standard deviations (in parentheses) for control variables as a function of signaling condition and prior knowledge group

	LPK		MPK		HPK	
	MIS- n = 29	MIS+ n = 19	MIS- n = 17	MIS+ n = 31	MIS- n = 16	MIS+ n = 15
Reading comprehension (correctly identified words)	12.45 (5.65)	12.63 (4.91)	12.82 (4.19)	13.06 (4.91)	13.75 (4.92)	13.40 (5.71)
Domain-specific academic self-concept	1.32 (0.57)	1.21 (0.59)	1.49 (0.66)	1.34 (0.51)	1.87 (0.71)	1.49 (0.49)
Domain-specific interest	1.81 (0.69)	1.57 (0.74)	1.81 (0.63)	1.58 (0.72)	2.14 (0.63)	1.75 (0.58)
Domain-specific prior knowledge (% correct)	53.32 (8.46)	56.28 (6.31)	72.85 (3.96)	71.22 (3.42)	91.35 (6.20)	87.69 (3.90)

The domain-specific interest of students was more pronounced in the MIS- condition in contrast to the MIS+ condition. Because domain-specific interest might influence the processing of materials, the differences between signaling conditions were controlled by including the variable as a centralized covariate in the regression analyses. There were no differences between the two signaling conditions regarding prior knowledge, $F < 1$, reading comprehension, $F < 1$, or domain-specific academic self-concept, $F(1,125) = 2.59$, $p = .110$, $\eta_p^2 = .02$. Means and standard deviations of the dependent variables are reported in Table 7.

Table 7

Means and standard deviations (in parentheses) for dependent variables as a function of signaling condition and prior knowledge group

	LPK		MPK		HPK	
	MIS-	MIS+	MIS-	MIS+	MIS-	MIS+
	n = 29	n = 19	n = 17	n = 31	n = 16	n = 15
Misconceptions (z-standardized)	-0.42 (0.94)	-0.02 (1.04)	0.04 (1.09)	-0.16 (1.03)	0.65 (0.67)	0.45 (0.79)
Comprehension performance (z-standardized)	-0.29 (0.77)	-0.01 (0.74)	0.44 (0.80)	0.03 (0.74)	0.70 (0.34)	0.32 (1.01)
Recall performance (z-standardized)	-0.14 (0.61)	0.06 (0.66)	0.24 (0.89)	-0.03 (0.77)	0.63 (0.55)	0.28 (1.03)
Time-on-task (minutes)	152.85 (20.28)	148.13 (15.85)	145.90 (21.96)	144.50 (17.76)	135.78 (27.78)	155.70 (13.28)
Extraneous cognitive load	3.85 (1.88)	4.68 (1.38)	3.73 (1.85)	4.00 (1.69)	2.79 (1.42)	3.67 (1.45)
Germane cognitive load	5.97 (1.97)	6.21 (1.48)	5.94 (2.25)	5.54 (1.45)	4.86 (2.31)	6.40 (1.12)

7.3.2 Misconceptions

For the measure of misconceptions, the regression model was significant, adj. R^2 = .08, F(6,120) = 2.86, p = .012. The amount of variance explained by the predictor domain-specific interest was not significant. The focal contrast explained a significant amount of variance, ΔR^2 = .11, F(1,120) = 14.38, p < .001, whereas the second subset of residual contrasts did not, ΔR^2 = .02, F(1,120) = 2.78, p = .098 (see Table 8).

Table 8

B-, SE-, β-, t- and p-values for dependent variables misconceptions, comprehension and recall performance

Dependent variable	Predictors	*B*	*SE*	β	*t*	*p*
Misconceptions	Domain-specific interest	-0.21	0.13	-0.01	-0.16	.871
	Focal contrast	0.26	0.07	0.34	3.79	< .001
	Residual contrast 1	0.10	0.15	0.06	0.70	.488
	Residual contrast 2	0.05	0.11	0.04	0.43	.665
	Residual contrast 3	0.08	0.06	0.11	1.26	.212
	Residual contrast 4	-0.01	0.04	-0.02	-0.20	.844
Comprehension performance	Domain-specific interest	0.20	0.10	0.17	2.01	.047
	Focal contrast	0.22	0.05	0.34	4.01	< .001
	Residual contrast 1	0.18	0.11	0.14	1.62	.109
	Residual contrast 2	0.01	0.09	0.01	0.08	.934
	Residual contrast 3	-0.03	0.05	-0.05	-0.57	.571
	Residual contrast 4	-0.02	0.03	-0.05	-0.61	.547
Recall performance	Domain-specific interest	0.14	0.10	0.13	1.44	.154
	Focal contrast	0.16	0.05	0.27	3.05	.003
	Residual contrast 1	0.12	0.11	0.10	1.08	.282
	Residual contrast 2	-0.01	0.09	-0.01	-0.16	.874
	Residual contrast 3	0.03	0.05	0.05	0.61	.545
	Residual contrast 4	-0.02	0.03	-0.05	-0.59	.556

The related data, which correspond to a full reversal of the signaling effect for HPK learners, are displayed in Figure 8.

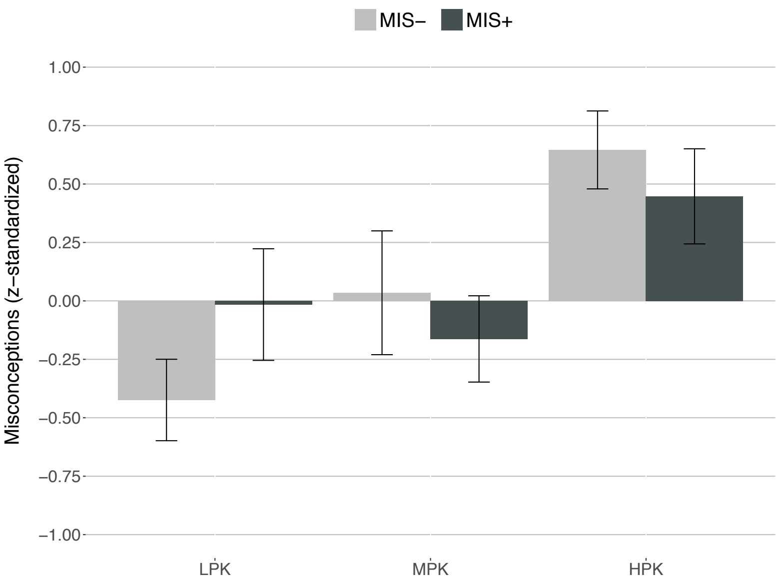


Figure 8. Performance on misconception measure as a function of prior knowledge level and signaling condition (error bars: +/- 1 SE).

7.3.3 Comprehension Performance

The regression model for comprehension performance was significant, adj. R^2 = .15, $F(6,120) = 4.66$, $p < .001$. The amount of variance explained by the predictor domain-specific interest was significant, $\Delta R^2 = .03$, $F(1,120) = 4.03$, $p = .047$. More interested learners showed better comprehension performance. Moreover, the focal contrast explained a significant amount of variance in the model, $\Delta R^2 = .12$, $F(1,120) = 16.05$, $p < .001$. For the residual contrasts, the explained variance was not significant, $\Delta R^2 = .02$, $F(1,120) = 3.07$, $p = .082$ (see Table 8). The results for comprehension performance, which again show a fully reversed signaling effect for HPK learners, are displayed in Figure 9.

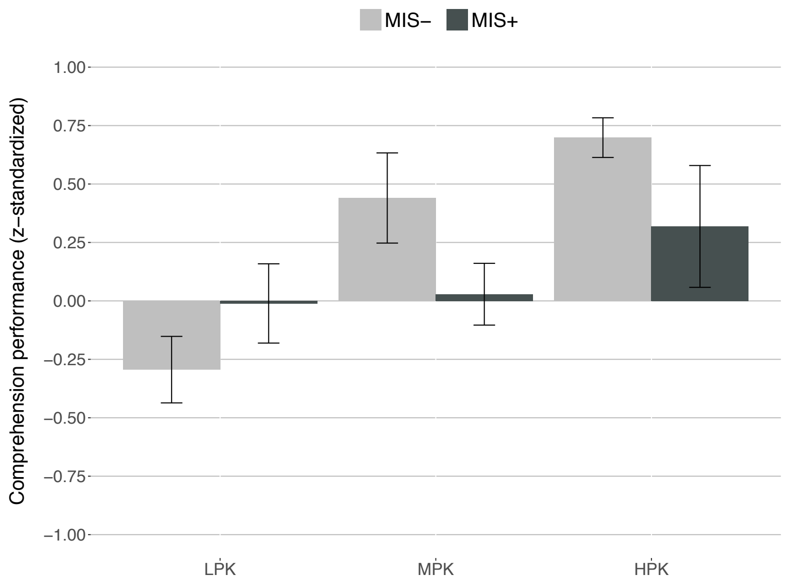


Figure 9. Comprehension performance as a function of prior knowledge level and signaling condition (error bars: +/- 1 SE).

7.3.4 Recall Performance

A significant regression model was obtained for recall performance, adj. R^2 = .07, $F(6,120) = 2.61$, $p = .021$. The first subset containing the domain-specific interest was not significant. The focal contrast explained a significant amount of variance, $\Delta R^2 = .07$, $F(1,120) = 9.29$, $p = .003$, whereas the residual contrasts did not, $\Delta R^2 = .02$, $F(1,120) = 2.27$, $p = .134$ (see Table 8). Related data are displayed in Figure 10 and indicate a fully reversed signaling effect for HPK learners.

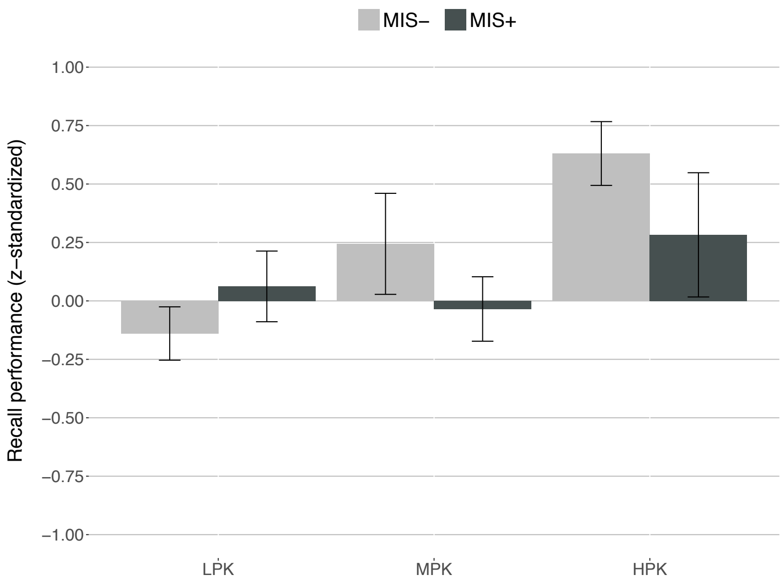


Figure 10. Recall performance as a function of prior knowledge level and signaling condition (error bars: +/- 1 SE).

7.3.5 Cognitive Load

Because some students failed to complete the cognitive load questionnaire completely, their missing values were replaced with the mean of the group that they belonged to.

The regression model for ECL was significant, adj. R^2 = .07, $F(6,120)$ = 2.48, p = .027. However, the focal contrast was not significant, ΔR^2 < .001, F < 1. Instead, the second subset of residual contrasts explained a significant amount of variance, ΔR^2 = .07, $F(1,120)$ = 8.91, p = .003. Separate omnibus tests for each of the residual contrasts in the set revealed that residual contrast 2 was marginally significant and residual contrast 4 was significant (see Table 5). Residual contrast 4 reflects that LPK learners experienced an overall higher ECL than MPK and HPK learners. According to residual contrast 2 a reverse pattern to what had been expected for LPK learners was present in the data, suggesting that LPK learners in the MIS+ condition perceived *more* ECL than in the MIS- condition despite the fact that they performed better in the MIS+ condition. Correlational analyses showed the expected negative correlation between ECL and learning outcomes for recall and comprehension measures; however, when looking at the correlations within each of the three levels of prior knowledge, the correlations were most pronounced for HPK learners, while they were not significant for LPK learners (see Table 9).

Table 9

Bivariate correlations between the three learning outcome variables and extraneous cognitive load by prior knowledge level

	Learning outcomes		
	Misconceptions	Comprehension performance	Recall performance
Extraneous cognitive load			
LPK (*n* = 48)	-.06	-.14	-.14
MPK (*n* = 48)	-.12	-.17	-.28$^{(*)}$
HPK (*n* = 31)	-.05	-.46**	-.41*
Overall (*n* = 127)	-.15$^{(*)}$	-.27**	-.30**

$^{(*)}$ *p* < .10. * *p* < .05. ** *p* < .01.

The regression model for GCL was not significant, adj. R^2 = .03, F(6,120) = 1.66, p = .137. At a descriptive level, in contrast to what had been expected, GCL was *higher* in the MIS+ condition compared with the MIS- condition for HPK learners. This observation was corroborated by the fact that the despite the overall non-significant model the second subset of residual contrasts explained a significant amount of variance, ΔR^2 = .04, F(1,120) = 5.43, p = .022, which was due to residual contrast 2 (see Table 10). Bivariate correlations between GCL and learning outcomes yielded no significant relations for the measure of misconceptions, r = .02, p = .867, recall performance, r = .11, p = .208, or comprehension performance, r = .03, p = .713.

Table 10

B-, SE-, β-, t- and p-values for dependent variables extraneous and germane cognitive load

Dependent variable	Predictors	*B*	*SE*	β	*t*	*p*
Extraneous cognitive load	Domain-specific interest	-0.37	0.22	-0.15	-1.69	.094
	Focal contrast	-0.00	0.19	-0.00	-0.01	.993
	Residual contrast 1	-0.09	0.25	-0.03	-0.36	.720
	Residual contrast 2	0.37	0.20	0.17	1.90	.060
	Residual contrast 3	-0.27	0.20	-0.13	-1.40	.164
	Residual contrast 4	0.22	0.10	0.19	2.16	.033
Germane cognitive load	Domain-specific interest	-0.32	0.24	-0.12	-1.34	.183
	Focal contrast	-0.31	0.21	-0.14	-1.50	.135
	Residual contrast 1	-0.24	0.27	-0.08	-0.89	.378
	Residual contrast 2	0.40	0.21	0.17	1.88	.063
	Residual contrast 3	0.19	0.21	0.08	0.90	.370
	Residual contrast 4	-0.05	0.11	-0.04	-0.47	.643

7.3.6 Time-on-task

It was tested whether there were differences between the signaling and prior knowledge groups with regard to time-on-task by means of a 2 (signaling) x 3 (prior knowledge) ANOVA. There were no main effects for either signaling, $F(1,121) = 1.59$, $p = .210$, $\eta_p^2 = .01$, or prior knowledge, $F < 1$, but a significant interaction, $F(2,121) = 3.98$, $p = .021$, $\eta_p^2 = .06$. The Bonferroni-corrected pairwise comparisons of signaling groups at each prior knowledge level revealed that only for HPK students there was a significant difference, $p = .006$, indicating that HPK students spent more time learning when they had the MIS+ material than those who had the MIS- material. The time-on-task was not significantly correlated with learning outcomes: measure of misconceptions, $r = .12$, $p = .179$, recall performance, $r = .04$, $p = .687$, and comprehension performance, $r = -.00$, $p = .987$.

7.4 Summary & Discussion

In a field study in schools with a high level of ecological validity and a strong control group, it was investigated how prior knowledge moderates the effectiveness of MIS. With the eChemBook a learning material of high ecological validity was used since it was designed by science education and multimedia researchers as well as partners from the educational industry and conformed to the actual curriculum used in schools.

Based on results of a meta-analysis (Richter et al., 2016) and the ERE (Kalyuga, 2007), it was hypothesized that LPK learners would profit from MIS with regard to learning outcomes, whereas HPK learners would perform better without these additional signals. MPK learners were expected not differ between signaling conditions (Hypothesis 1). Contrast analyses that directly tested this hypothesis revealed that the hypothesized pattern significantly explained variance for comprehension and recall measures as well as a measure that was designed to assess learners' misconceptions in the domain. Consequently, the results corroborate the main assumption that prior knowledge moderates the signaling effect in that signals help LPK learners, but hinder learning for more advanced students, thereby suggesting a full reversal of the signaling effect for HPK learners.

Form a cognitive load perspective, there are at least two possible explanations for this full reversal. First, HPK learners might be forced to process redundant information when studying MIS, which should lead to higher ECL (Hypothesis 2). Second, they might engage less in meaningful learning activities when studying MIS, which should lead to less GCL being invested (Hypothesis 3). Because of the problems associated with disentangling these two cognitive load components at an empirical level and with subjective ratings in general (cf. time of measurement, item types; de Jong, 2010; volatile individual frame of reference for subjective ratings; Schnotz & Kürschner, 2007), these explanations were investigated only as a secondary research question.

Hypothesis 3 related to GCL could not be confirmed. At a descriptive level, GCL was even higher for the HPK-MIS+ group in contrast to the HPK group learning with the MIS- material. However, the GCL measure was not correlated to learning outcomes, which is why it was assumed that it did not measure the construct of GCL properly.

With regard to ECL, no difference for MPK learners were found whereas both LPK and HPK students who learned from the MIS+ version tended to report a higher ECL than students in the corresponding MIS- groups. Thus, the results hint towards the hypothesized pattern for HPK students, but the pattern for LPK students is not in line with CLT, which partly confirms Hypothesis 2. Since instructional techniques like signaling should reduce ECL for learners with a low level of prior knowledge from a theoretical point of view (Sweller et al., 1998), the results might again point towards measurement problems. Correlational analyses revealed that the ECL measure was negatively correlated with comprehension and recall performance only for HPK learners. Hence, the direction of the correlation between ECL and learning outcomes for HPK learners is in line with the theoretical considerations of the CLT (Sweller et al., 1998). For LPK and MPK learners there was no relation between ECL ratings and learning outcomes. Although Sweller et al. (1998) concluded that learners should have no difficulties estimating their experienced cognitive load, the non-significant correlations between ECL ratings and learning outcomes for LPK and MPK might be due to difficulties for these groups in observing their cognitive processes and translating their experienced mental load into a rating. That is, prior knowledge may serve as a prerequisite to adequately judge one's learning. Similar findings have been observed in research on monitoring accuracy, where students are asked after learning to judge their ability to perform well in a subsequent assessment. Here it has been shown that students with less prior knowledge also provide more inaccurate judgments of their learning (cf. van Loon, de Bruin, van Gog, & van Merriënboer, 2013).

If I assumed that, in contrast to LPK learners, HPK learners were able to accurately judge their cognitive load, then their ECL ratings confirm the explanation regarding the full reversal, namely, that these learners are hindered in their learning by being forced to engage in unnecessary processing of potentially redundant information. This explanation is further corroborated by the results regarding time-on-task. Here the signaling groups differed only for HPK learners in that students in the MIS+ group spent more time with the eBook compared to students in the MIS- group.

This result can be interpreted as indicating that HPK students engaged in further processing of the materials, which, however, did not contribute to their learning. At the same time, it suggests that different from what would be expected from previous studies (Haider & Frensch, 1999; Scheiter & Eitel, 2015) even HPK learners are unable to ignore information that is not helpful for their learning. However, because of the unexpected results regarding the ECL ratings of LPK learners and the lack of correlations between ECL and learning outcomes for LPK and MPK learners, the question of whether a full reversal for HPK learners is due to unnecessary processing and hence higher ECL warrants further study.

In particular, eye tracking may be used in the future to shed more light on the underlying processes and to more thoroughly investigate processing differences among learners with low, medium and high levels of prior knowledge. The lack of this data in the present study may be considered to be one of its limitations, but is a natural consequence of the wish to run a study under ecologically valid conditions. Moreover, verbal protocols or more objective physiological data on cognitive load could be obtained for example by means of measurement of pupil diameter or the heart rate variability of students during learning (De Jong, 2010; Schnotz & Kürschner, 2007). Thus, in Study 3 in the following chapter eye tracking methodology will be used to shed light on the nature of the expertise reversal of the multimedia signaling effect by means of visual attention distribution measures and pupil diameter of students as a measure of cognitive load in addition to subjective cognitive load ratings.

8. Study 3: Studying the Effects of Multimedia Signaling at a Process Level – Evidence from Eye Tracking[6]

Study 3 aimed at addressing the question whether prior knowledge affects the multimedia signaling effect and whether this influence leads to a partial or full ERE (Research Question 1 as well as 3 and 4). Thus, a similar sample as used in Study 2 with varying levels of prior knowledge and part of the eBook about the Particle Model of Matter used in Study 2 was used in the current study. However, the main focus of Study 3 was to shed light on underlying processes related to the expertise reversal of the multimedia signaling effect by means of eye tracking (Research Question 5). To assess cognitive load, again as in Study 2 subjective ratings of ECL and GCL were assessed, but in addition the objective general cognitive load measure pupil diameter was recorded during learning. The visual attention distribution of learners was operationalized through fixation times, the time to first fixating highlighted parts of the learning material, and transitions between texts and pictures (for theoretical background, see chapter 4.3.2).

8.1 Hypotheses

Based on the ERE (Kalyuga et al., 2003), the ability-as-compensator hypothesis (Mayer & Sims, 1994) and results obtained in Studies 1 and 2 it can be assumed that prior knowledge influences the relation between multimedia instructions with and without MIS and learning outcomes. The processes underlying such an expertise reversal of the multimedia signaling effect might be related to visual attention distribution and cognitive load (cf. theoretical background in chapters 4.2 and 4.3).

In order to thoroughly investigate these processes related to Research Question 5, moderated mediation analyses were conducted (Preacher et al., 2007). Two different models were derived from the existing literature and compared to each

[6] Richter, J., & Scheiter, K. (in preparation). *Studying the effects of multimedia signaling at a process level: Evidence from eye tracking.*

other. Both models are shown as path models in in Figure 11 and will be explained in the following. In both models, MIS are expected to have an influence on learning outcomes (*path c*) that is assumed to be moderated by the learners' prior knowledge as indicated by an arrow from prior knowledge onto *path c*. In general, the aim of the moderated mediation analysis is to test whether process measures related to visual attention and cognitive load (mediator variables, indicated by the box M) and the influence of prior knowledge (moderator variable, indicated by the box W) can explain the potential interaction between multimedia signaling and prior knowledge on *path c* (i.e., the moderation reflecting the expertise reversal effect). The two models differ in the exact way that prior knowledge is assumed to influence the mediators and learning outcomes in order to explain the expertise reversal effect.

In the following, both moderated mediation models will first be explained in general related to (a) visual attention and (b) cognitive load mediator variables. Then, the concrete hypotheses for each of the paths included in the models and mediator variables will be reported.

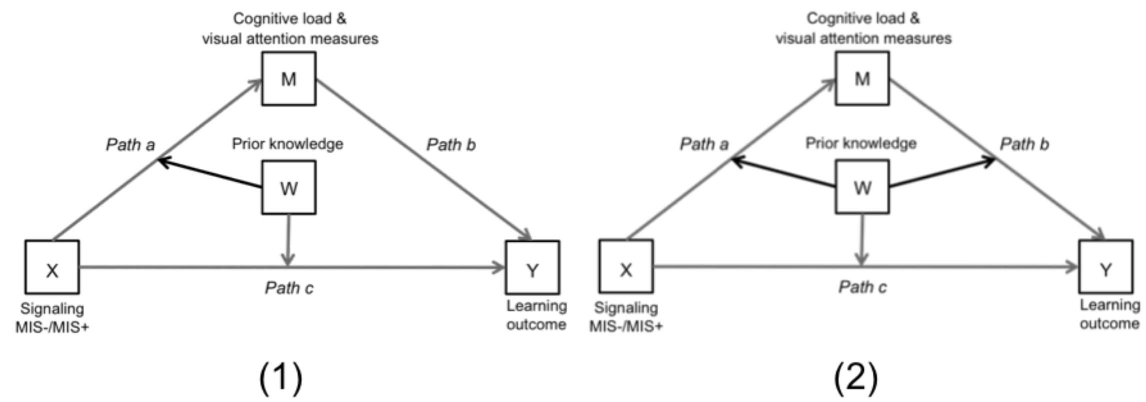


Figure 11. Hypothesized moderated mediation models with multiple mediator variables for visual attention and cognitive load measures and the moderator variable prior knowledge.

If we consider results obtained by Gegenfurtner et al. (2011), Mason, Tornatora, and Pluchino (2013), and Schwonke, Berthold, and Renkl (2009), we can assume that prior knowledge influences the way learners visually process materials. Thus, LPK learners were expected to distribute their attention during learning differently than HPK learners do. Hence, this assumption is reflected in both moderated mediation models (indicated by an arrow from prior knowledge onto *path a*).

The empirical basis regarding the question of whether and how prior knowledge influences the effect of visual attention distribution on learning outcomes (i.e., *path b*) is weak. To the best of my knowledge, only Schwonke, Berthold, and Renkl (2009) tested this interaction between visual attention and prior knowledge related to learning outcomes. They reported that the effect of visual attention on learning outcomes was indeed influenced by learners' prior knowledge. When interpreting this result in the context of the present study one has to keep in mind that Schwonke, Berthold, and Renkl (2009) manipulated the information learners got related to the functions of representations. Thus, one has to be cautious when transferring this result to the current context because different processes might be triggered by the instructional support measure MIS and providing information about the function of representations.

To conclude, at present it is not possible to decide whether prior knowledge only affects the effect of MIS on visual attention measures (*path a*) or whether prior knowledge additionally influences the effect of visual attention on learning outcomes (*path b*). Therefore, the postulated moderated mediation models differ with respect to the interaction between prior knowledge and visual attention measures related to learning outcomes. In line with the result of Schwonke, Berthold, and Renkl (2009), model 2 contained the interaction (indicated by an arrow from prior knowledge onto *path b*). Conversely, from a theoretical point of view it might also be that visual attention distribution directly affects learning outcomes. For example, the longer students fixated highlighted information the better their learning performance might be *independen*t of their prior knowledge. Thus, in model 1 a direct effect of visual attention on learning outcomes was postulated (indicated by the *path b* arrow from box M reflecting visual attention measures to box Y reflecting learning outcome measures as well as by the *lack* of an arrow from prior knowledge onto *path b*).

With regard to the potential mediating effect of cognitive load depending on prior knowledge the moderated mediation models are framed against theoretical assumptions based on the CLT (Chandler & Sweller, 1991). As stated in chapter 4.2, additional instructional support can substitute missing schemas in the long-term memory of LPK learners and, thus, reduce their working memory load during learning. HPK learners, however, have automated schemas at hand. Hence, their overall cognitive load during learning should be much lower than for LPK learners. Therefore, HPK learners might not need further instructional support provided by MIS. As hypothesized in Study 2 (chapter 7.1) from a theoretical point of view MIS should have a different effect for LPK and HPK learners related to ECL and GCL. For LPK learners it was expected that they experience less ECL and more GCL when learning with the MIS+ compared to the MIS- condition. Vice versa, HPK learners were expected to experience either more ECL or lower GCL due to MIS. The competing ECL and GCL assumptions for HPK learners were associated with different explanatory approaches for the occurrence of full EREs. Taken together, for the current study it was expected that the effect of multimedia signaling on cognitive load measures would be influenced by the prior knowledge of learners (indicated by the arrow from prior knowledge onto *path a*).

Similar as for the mediator variables related to visual attention, empirical evidence on the potential effect of cognitive load on learning outcomes and the influence on this effect by prior knowledge is weak. Generally speaking, the CLT assumes that meaningful learning should go along with low levels of ECL and high levels of GCL (Van Gog & Paas, 2008; Paas, van Gog, & Sweller, 2010). Thus, cognitive load is supposed to have an effect on learning outcomes, as indicated in model 1 (*path b* arrow from box M reflecting cognitive load measures to box Y reflecting learning outcome measures). The question whether this direct effect of cognitive load on learning outcomes is influenced by learners' prior knowledge cannot be answered at this point because related empirical evidence is lacking. This question was addressed in model 2 indicated by the arrow from prior knowledge onto *path b*.

Hypotheses related to both moderated mediation models are described for each of the included paths and mediator variables hereinafter.

1. Based on evidence from Study 2 (chapter 6), the hypothesis related to the effect of multimedia signaling on learning outcomes was a full ERE in that LPK learners were supposed to profit from MIS with respect to learning outcomes. Again, HPK learners were supposed to be hindered in learning, resulting in worse learning performance for learners in this group learning with signaled learning material. This assumption forms the basis for the investigation of potential underlying processes of the ERE. Therefore, this assumption is the same in both moderated mediation models (Hypothesis 1, *path c,* models 1 and 2).

2. With regard to Hypothesis 2a (*path a*, visual attention measures, models 1 and 2), it was expected that MIS increase fixation times on highlighted pictures at least for LPK students (guiding-attention hypothesis with LPK students; cf. Boucheix & Lowe, 2010; de Koning et al., 2010a; Kriz & Hegarty, 2007; Ozcelik et al., 2010; Scheiter & Eitel, 2015). The fixation times on pictures are especially relevant because students tend to rely more strongly on information provided by text rather than pictures (e.g., Hannus & Hyönä, 1999; Hegarty & Just, 1993) when no further instructional support is given during multimedia learning. Moreover, for LPK learners the time to first fixating highlighted pictures should decrease when MIS are present in the learning material (Ozcelik et al., 2009; Ozcelik et al., 2010; Scheiter & Eitel, 2015). Moreover, it was expected that MIS also positively affect the number of transitions because MIS are supposed to support especially the integration process of verbal and pictorial information for LPK learners (cf. Richter et al., 2016). Thus, for LPK learners it was expected that they make more transitions between texts and pictures when MIS are present than when they are not. Furthermore, as suggested by Mason, Tornatora, and Pluchino (2013) and Schwonke, Berthold, and Renkl (2009), prior knowledge of learners might moderate the relationship between signaling and visual attention measures. However, the results of the studies were contradictory with respect to the direction of the influence of prior knowledge on visual attention measures. Results by Mason,

Tornatora, and Pluchino (2013) revealed that visual attention increased the more prior knowledge learners had when learning with multimedia material. In contrast, Schwonke, Berthold, and Renkl (2009) reported that visual attention decreased with increasing prior knowledge in the group learning with multimedia material that did not contain further instructional support. Thus, it was hypothesized that prior knowledge affects *path a* (models 1 and 2) of the moderated mediation model, however, no directed hypothesis for the influence of prior knowledge on the relation between multimedia signaling and visual attention measures was postulated.

3. Similar to Study 1, (types of) cognitive load are expected to differ based of multimedia signaling and the influence of prior knowledge. Thus, in Hypothesis 2b (models 1 and 2) it was assumed that MIS affect the level of subjectively rated ECL and GCL as well as the pupil diameter of learners, depending on the prior knowledge of learners (moderation of prior knowledge). The assumptions for the influence of prior knowledge on the effect of multimedia signaling on subjectively rated ECL and GCL are the same as in Study 2 (see chapter 7.1). Regarding HPK learners, it was expected that MIS either increase ECL or decrease GCL, which was related to the two different explanatory approaches for full EREs. For LPK learners MIS were expected to decrease ECL and increase GCL. With respect to pupil diameter of students, no directed moderation hypothesis was postulated because evidence in the context of multimedia signaling is lacking (Hypothesis 2b, *path a*, cognitive load measures, models 1 and 2).

4. Based on results obtained by Schwonke, Berthold, and Renkl (2009), it was expected that visual attention distribution (fixation times, time to first fixation, and transitions) is related to learning outcomes depending on the prior knowledge of learners (model 2). However, no directed moderation hypothesis was postulated because Schwonke, Berthold, and Renkl (2009) did not investigate the effect of MIS (Hypothesis 3a, *path b*, visual attention measures, model 2). However, as stated above the empirical basis on the influence of prior knowledge on the effect of visual attention on learning outcomes is weak. Moreover, visual attention distribution in the study by Schwonke, Berthold, and

Renkl (2009) was affected by giving learners information about the functions of representations rather than multimedia signaling. Therefore, we have to be cautious when interpreting the results by Schwonke, Berthold, and Renkl (2009) in the light of the current study. As a result, in model 1 a direct effect of visual attention on learning outcomes without any influence of prior knowledge was hypothesized (Hypothesis 3a, *path b*, visual attention measures, model 1).

5. Similar to the preceding Hypothesis 3a, in Hypothesis 3b it was expected that cognitive load measures (GCL, ECL ratings and pupil diameter) would either be related to learning outcomes depending on the prior knowledge of learners (*path b*, model 2) or would directly affect learning outcomes (*path b*, model 1). Again, no directed moderation hypothesis for model 1 was postulated, because results obtained by Schwonke, Berthold, and Renkl (2009) were obtained with a different manipulation than multimedia signaling and moreover, only visual attention parameters were used as mediators in their moderated mediation model. Nevertheless, from a theoretical point of view it was expected that independent of prior knowledge (model 1) learning performance decreases with increasing ECL and increase with increasing GCL (e.g., Paas et al., 2010).

6. Finally, visual attention distribution and cognitive load measures were expected to mediate the effect of multimedia signaling on learning outcomes moderated by prior knowledge, whereby due to a lack of empirical evidence it was expected that prior knowledge would either influence *path a* only (model 1) or both *path a* and *path b* (model 2) (Hypothesis 4, moderated mediation, model 1/2).

8.2 Method

8.2.1 Participants and Design

Eighty-three students of secondary higher education (Gymnasium) from a southern federal German state participated in the study for payment (20€). Students

(M_{age} = 14.67 years; age range 13 - 17 years; 39 female) were in seventh, eighth, and ninth grade to ensure heterogeneity regarding their prior knowledge. Data from 11 students had to be excluded from the eye tracking analysis due to insufficient calibration quality. Participants with poor calibration values nevertheless completed the experimental procedure for ethical reasons. To this end, 72 participants were included in the eye tracking analysis resulting in mean calibration values on the *x*-axis, M = 0.41, SD = 0.34, and on *y*-axis, M = 0.41, SD = 0.21. The mean tracking ratio was 92.11% (SD = 9.54). Students were randomly assigned to one of the two signaling conditions resulting in 38 students learning with the MIS- and 34 students learning with the MIS+ learning material.

8.2.2 Instructional Materials

Students learned from two learning units of the digital textbook *eChemBook*, which was also used in Study 1 (chapter 7) about the Particle Model of Matter, which is part of the curriculum in introductory chemistry education in grade 8. In the first learning phase students were introduced to the model, its assumptions and boundaries (6 pages, approx. 1,050 words, 6 static pictorial representations). In the second learning phase they learned how diffusion can be explained by means of the Particle Model of Matter (7 pages, approx. 860 words, 5 static pictorial representations). The two learning units of the eChemBook were experimentally manipulated with regard to MIS similar to the manipulation in Study 1 (see chapter 7.2.2).

1. The MIS- version of the learning units contained basic text signals like headings, paragraphs, and bold typeface. In order to ensure a fair control condition (cf. Schwonke, Renkl et al., 2009), the MIS- version also contained the MIS corresponding labels in text and picture where appropriate.

2. In the MIS+ version of the learning units additional MIS were implemented. In the introductory learning unit, 1 corresponding label and 2 deictic references were implemented. In learning phase 2 about dissolution, 4 color coding signals and 2 deictic references were included.

8.2.3 Measures

Pre-test. Prior to learning, students answered a pre-test that assessed demographic data, their prior knowledge, reading comprehension as well as domain-specific self-concept, interest, and scientific understanding. The prior knowledge test consisted of 9 verification items that were partly self-constructed and partly adapted from Hollstein (2001) and ZPG-Chemie (2011), and 15 multiple-choice items that were also partly self-constructed and partly adapted and where necessary translated into German from different sources (Hollstein, 2001; Mulford & Robinson, 2002; Petermann et al., 2009; Yezierski & Birk, 2006; ZPG-Chemie, 2011). These items were directly related to common misconceptions related to the Particle Model of Matter (e.g., Aydeniz & Kotowski, 2012; Nakhleh, 1992; Yezierski & Birk, 2006). Students for example commonly think that particles have a color or that they change their shape depending on the aggregate state of the matter. However, particles do not have a color or change their shape based on the Particle Model of Matter. A related verification item was for example “When gas particles are heated, they expand.“. Students had to indicate whether they think this is true or false. For each item that was answered correctly as being true or false, it was scored with 1 point resulting in a maximum of 9 points. A multiple-choice item related to a common misconception was for example “Which of the statements is correct? Between particles that form matter is… nothing/ water if the matter is a liquid/ air/ steam if the matter is a gas/ dust and pollutants”. The correct answer was “nothing” which was scored with 1 point, whereas the choice of the other options resulted in 0 points. Overall, a maximum of 15 points could be reached by students. The sum of points was transformed into percentage correct for better interpretation. Cronbach’s α for the domain-specific verification items was very low with .12 whereas Cronbach’s α for the domain-specific multiple-choice items was .69. Therefore, only multiple choice items were used in the analyses to represent the domain-specific prior-knowledge (see Appendix B).

Reading comprehension was assessed by means of a standardized reading test (LGVT 6-12; Schneider et al., 2007) that asked students to read a text (3.5 pages, 1,727 words). The text contained 23 brackets that each contained 3 words.

When students came across a bracket they had to decide which of the 3 words included in the bracket fits best in the context of the text. The number of correctly chosen words in brackets was used to determine reading comprehension.

The domain-specific scientific understanding was assessed by 10 multiple-choice items that were partly self-constructed and partly adapted from test items for chemistry of the TIMMS Germany study for class 7 and 8 (Baumert et al., 1998). An example item was “What is an example of a chemical reaction? melting ice/ salt crystals grinded into powder/ burning wood/ water evaporates from a puddle”. Each correct item was scored with 1 point, which resulted in a maximum of 10 points that were also transformed into percentage correct for better interpretation. Cronbach’s α for these domain-specific scientific understanding items was .58.

The domain-specific self-concept was assessed by means of 5 items developed by Schanze (2002) that were adjusted by Grüß-Niehaus (2010). Items were for example: “I am simply talented for chemistry” or “I can solve even difficult tasks in chemistry” and were rated by students on a four-point Likert scale ranging between “I do not agree at all” and “I completely agree”. Cronbach’s α for this test was .83. The domain-specific interest was assessed by means of 5 items by Wilde et al. (2009) that were adapted to the science context. Example items were: “I like reading science texts” or “I am interested in learning new things in science”. The items were also answered on a four-point Likert scale ranging between “I do not agree at all” and “I completely agree” (Cronbach’s α = .82).

Post-test. The posttest assessed the misconceptions of the students related to the Particle Model of Matter, the recall performance of the dissolution process on the model level, their comprehension performance as well as mapping performance of terms to microscopic- or macroscopic level. The test for misconceptions consisted of 6 verification items (Cronbach’s α = .49) and 13 multiple-choice items (Cronbach’s α = .62). The items measuring misconceptions were partly similar to those from the pre-test and partly self-constructed or adapted and where necessary translated into German from different sources (Hollstein, 2001; Mulford & Robinson, 2002; Petermann et al., 2009; Yezierski & Birk, 2006; ZPG-Chemie, 2011). The scoring of the verification- and multiple-choice items was the same as in the pre-test. The verification items resulted in a maximum of 6 points and the multiple choice items into a maximum of 13 points that were each transformed into percentage correct for better interpretation. Verification and multiple-choice item scores were *z*-standardized separately and then the mean of the two item types (verification and multiple-choice items) was calculated in order to combine them into one misconceptions outcome measure for data analysis.

A self-constructed recall task (15 possible points, Cronbach's α = .60) was used to assess recall performance about the dissolution process based on the Particle Model of Matter. Students were asked to draw the dissolution process of sugar into water on the model level and explain verbally what happens in the different phases. In a comprehension task (9 possible points, Cronbach's α = .69) students were asked to use their knowledge about the Particle Model of Matter to explain the phenomenon that particles that form matter A can permeate through a membrane whereas particles that form matter B cannot. Students were again asked to draw their solution and explain it verbally. The solution is that particles that form different matter can also differ with respect to their size. This is one of the basic assumptions about the model in the introductory learning unit that students had to transfer to this situation. When particles are larger than the wholes in the membrane they cannot permeate. Both recall and comprehension task were rated based on an author-developed coding schema by 2 raters resulting in an overall interrater-reliability of Cohen's kappa = .84. The scores of both tasks were again transformed into percentage correct for better interpretation.

In a mapping task students were asked to assign 10 terms like temperature, color, or particle movement to either the macroscopic or the microscopic level. Each correctly assign item to one of the two levels was scored with 1 point resulting in a maximum of 10 possible points for this task (Cronbach's α = .74). The score was also transformed into percentage correct for better interpretation.

ECL and GCL were each measured by one item. ECL was assessed by the question "How difficult was it for you to understand the contents?", which was like in Study 1 (chapter 7) adapted from Kalyuga et al. (1998). The GLC item in the current study was changed with respect to the item used in Study 1, because it was probably subject of measurement errors. The item for GCL, therefore, did no longer directly relate to the mental effort spend during learning but rather indirectly measured effort by referring to the concentration during learning and read “How much did you concentrate during learning with the digital textbook?” (used by Cierniak et al., 2009). Students subjectively rated these ECL and GCL items on a 9-point rating scale with 1 (*not difficult at all/not concentrated at all*) to 9 (*extremely difficult/concentrated very intensively*).

Time-on-task. The time-on-task for each learning session was tracked and noted down by the experimenter for each learning session. For the data analysis, both time-on-task values were combined into a cumulated time-on-task in minutes for both learning sessions.

Eye tracking parameters. Eye Tracking data were analyzed with respect to four measures: (a) fixation times on pictures, (b) the number of transitions between texts and pictures and (c) the time to first fixating highlighted pictures as measures of visual attention, as well as (d) pupil diameter as a general measure of cognitive load. For the analysis of fixation times on pictures, transitions and the time to first fixation areas of interest (AOIs) were on the one hand defined globally around text and corresponding pictures on eBook pages that contained MIS. On the other hand, in order to enable more fine grained analysis, AOIs were also defined within particularly relevant parts of texts necessary for integration and also elements pivotal for integration within pictures. The eye tracking parameters were cumulated over all fixations and transitions a participant made during learning on or between certain AOIs. The time to first fixation was averaged across related parts of an eBook page in cases multiple parts were highlighted by means of MIS. Pupil diameter values were averaged across the total time a learner inspected an eBook page.

8.2.4 Apparatus

The learning material was presented using SMI Experiment Center™. Because two different SensoMotoric Instruments (SMI, Teltow, Germany) eye tracking devices were used in the study, part of the participants were presented with stimuli on a screen with a 1680x1050 pixel-resolution and part of the participants with stimuli on a screen with a 1920x1080 pixel-resolution. The refresh-rate was 250 Hz. The eye tracking data was analyzed with the software BeGaze 3.6^{TM} by using the default saccade velocity algorithm for detecting fixations and saccades.

8.2.5 Procedure

Participants were tested individually in sessions of approximately 90 minutes in the laboratory or in rooms provided by schools. Sessions in schools took place during afternoons outside the regular class time. Testing and learning were self-paced without any time limits. Participants either learned with the MIS- or the MIS+ version of the eBook consistently in both learning sessions. The brightness of the laptop screen was kept constant in order to prevent influences on pupil diameter of learners. At first, students filled in a paper-based questionnaire assessing demographic information and the domain-specific self-concept and interest. Participants without prior experience with the subject chemistry in school were asked to imagine how it would be to learn about chemistry. Thereafter, they filled in a digital questionnaire assessing prior knowledge about the Particle Model of Matter and the domain-specific scientific understanding followed by the first learning session at the eye tracking device. After the calibration of the eye tracking device students learned 6 pages about the introduction to the Particle Model of Matter, its assumptions and boundaries. They could navigate forward and backward by pressing the corresponding arrow keys on the keyboard. After the first learning session there was a break of 15 minutes where participants colored mandalas. The break was necessary for ethical reasons in order to prevent participants from strain on the eyes. After the break, the eye tracking device was calibrated again and participants learned the second learning unit containing 7 pages about diffusion on the microscopic level. Subsequently, participants filled in the post-test. The first items of the digital post-test were verification items followed by multiple choice items assessing misconceptions and the mapping task. Thereafter, the two paper-based drawing items assessing recall and comprehension performance were handed to the participants and participants rated their extraneous and germane cognitive load. Finally, participants were debriefed and thanked for their participation.

8.2.6 Data Analyses

In order to test the hypothesized moderated mediation models (see Figure 11), separate regression analyses for *path a, b*, and *c* were computed. In order to test whether the indirect effects on *path a* and *path b* (interactions with prior knowledge) are significant (Hypothesis 4), the joint significance test method was used (Judd, Yzerbyt, & Muller, 2014; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Muller, Judd, & Yzerbyt, 2005) together with bootstrapped 95% confidence intervals using 5000 bootstrap samples to determine the effect size. According to the joint significance test, a moderated mediation is significant if both hypothesized interaction effects related to *path a* and *path b* are significant.

In order to test Hypotheses 1 as well as 2a/b related to *path c* and *path a* of the moderated mediation models, separate regression analyses were conduced for each of the dependent variables related to learning outcomes (*path c*) and visual attention as well as cognitive load measures (*path a*) with the independent variable signaling and the potential moderator variable prior knowledge and the related interaction (signaling x prior knowledge) as predictors.

To test Hypotheses 3a/b related to *path b* of the moderated mediation models, it was decided to compute separate regression analyses for each of the mediator variables (mean centered visual attention and cognitive load measures) rather than computing one regression analysis with the potential mediator variables included at once. This decision was made because computing only one regression analysis with all mediator variables can cause multicollinearity problems due to the intercorrelations among the mediators. Visual attention measures as well as cognitive load measures are likely to correlate, which means they are not independent from each other. Consequently, some of the measures would have had to be excluded from the analysis, which would have led to a loss of information related to the effects on *path b*. Therefore, separate regression analyses for each of the mediator variables as independent variable, a learning outcome measure as dependent variable and prior knowledge as a potential moderator variable with the related interactions between prior knowledge and the independent variables were computed.

Nevertheless, in order to detect potential multicollinearity problems (Belsey, Kuh, & Welsch, 2004), the Variance Inflation Factor (VIF) was computed for each regression analysis (Allison, 1999; Mansfield & Helms, 1982). The VIF indicates how much the variance of the coefficient estimate is being inflated due to multicollinearity. The rule of thumb regarding the VIF is that a multicollinearity problem might be present in the data in cases when the VIF value is smaller than 0.1 and larger than 10 (Miles, 2005). Allison (1999), however, plead for a more conservative rule of thumb, namely that a VIF above 2.5 and a tolerance value smaller than .40 is a reason for concern that the collinearity assumption is not met (p. 142). For the regression model related to *path b* of the moderated mediation model (Figure 11) the more conservative approach was chosen.

The predictor multimedia signaling condition was centered with -0.5 for the MIS- and 0.5 for the MIS+ condition. The predictor domain-specific prior knowledge was mean centered in order to facilitate the interpretation of results.

In case of significant interaction terms, simple slopes were tested at +1 standard deviation, and -1 standard deviation of the continuous moderator variable prior knowledge. In addition, for *path b* of the moderated mediation models, simple slopes were tested each at +/-1 standard deviation of the mediator variable.

8.3 Results

Results were analyzed in a five-step procedure. First, it was tested whether there were differences between the signaling conditions with regard to the prior knowledge and the control variables reading comprehension, domain-specific scientific understanding, domain-specific self-concept and interest by means of ANOVAs. Moreover, it was tested whether the signaling groups differed with respect to the time-on-task, in order to ensure that potential differences in learning outcomes are not due to different learning times. Second, regression analyses related to *path c* of the moderated mediation models were conducted. In a third and fourth step, the regression models related to *path a* and *path b* of the moderated mediation models were computed. Finally, in a fifth step the significance of the indirect effects of *path a* and *path b* was determined by means of a joint significance test (Judd et al., 2014; MacKinnon et al., 2002; Muller et al., 2005). For all statistical analyses the α level was set to .05.

8.3.1 Control Variables

The prior knowledge did not differ between the MIS- and MIS+ eBook versions, $F < 1$, nor did the domain-specific scientific understanding, $F < 1$, the domain-specific self-concept, $F < 1$, the domain-specific interest, $F < 1$, or reading comprehension, $F < 1$ (for means and standard deviations see Table 11). Thus, the multimedia signaling conditions were equivalent with respect to control variables. The time-on-task did also not differ significantly between the signaling conditions, $F < 1$.

Table 11

Means and standard deviations (in parentheses) for control variables and the time-on-task as a function of signaling condition

	MIS-	MIS+
	n = 38	*n* = 34
Reading comprehension (correctly identified words)	11.24 (4.65)	12.29 (4.78)
Domain-specific scientific understanding (% correct)	56.84 (22.43)	61.77 (21.10)
Domain-specific academic self-concept	1.47 (0.51)	1.48 (0.54)
Domain-specific interest	1.94 (0.58)	1.95 (0.50)
Domain-specific prior knowledge (% correct)	52.63 (21.01)	50.39 (21.79)
Time-on-task (min)	14.68 (4.75)	15.56 (6.05)

8.3.2 Multimedia Signaling Effect on Learning Outcomes and the Influence of Prior Knowledge (*path c*)

In this chapter results related to *path c* of the moderated mediation models will be reported. Thus, only interaction effects (multimedia signaling x prior knowledge) will be reported. A significant interaction effect indicates that there is a full (or partial) expertise reversal of the multimedia signaling effect.

Regression analyses were computed separately for the learning outcome measures of misconceptions, recall, comprehension, and mapping performance (estimated means and standard errors in Table 12) with prior knowledge, signaling intervention and the related interaction term entered simultaneously as predictors. For recall performance a significant interaction between prior knowledge and signaling was found, β = -0.26, *p* = .022 (see Table 13). A simple slopes analysis revealed that the recall performance between the signaling conditions differed significantly only for LPK learners at one standard deviation below the mean of the prior knowledge measure, *p* = .033, in that they performed better when MIS were present. There were no differences for HPK learners at one standard deviation above the mean, *p* = .257 (see Figure 12).

Table 12

Estimated means and standard errors (in parentheses) of learning outcomes for path c of the moderated mediation models

	MIS-		MIS+	
	LPK	HPK	LPK	HPK
Misconceptions (*z*-standardized)	-0.39 (0.17)	0.48 (0.17)	-0.62 (0.17)	0.54 (0.18)
Recall performance (% correct)	48.47 (3.02)	64.09 (2.87)	57.72 (2.98)	59.23 (3.14)
Comprehension performance (*z*-standardized)	-0.36 (0.14)	0.47 (0.13)	-0.72 (0.13)	0.50 (0.14)
Mapping performance (% correct)	78.65 (3.85)	90.52 (3.65)	89.57 (3.79)	97.95 (4.00)

Note: Means and standard errors were estimated by means of regression analyses with signaling intervention, prior knowledge and the related interaction included as predictors.

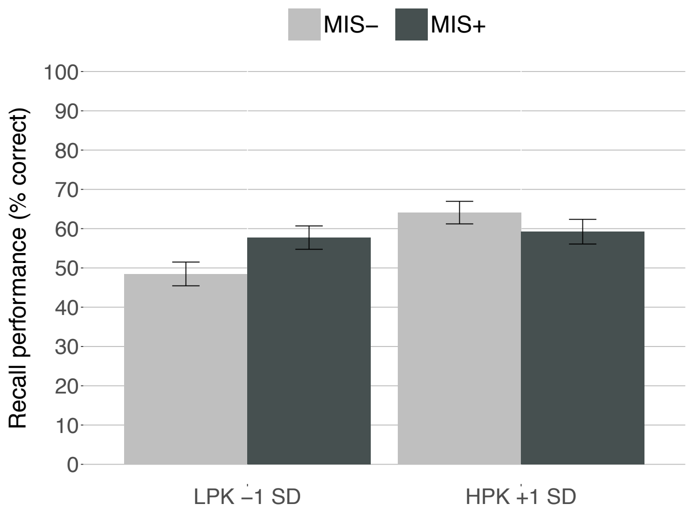


Figure 12. Recall performance as a function of signaling condition at +/-1 *SD* of the prior knowledge measure (error bars: +/- 1 *SE*).

For comprehension performance, β = 0.127, *p* = .152, and the measure of misconceptions, β = 0.085, *p* = .392, the interaction terms were not significant (see Table 13 for results of regression analyses).

For mapping performance, the interaction term was also not significant, β = -.047, *p* = .675, however, there was a significant main effect of prior knowledge, β = .291, *p* = .011, and of signaling, β = .273, *p* = .017 (see Table 13 for results of regression analyses). The estimated means and standard errors in Table 12 reveal that participants in the MIS+ condition scored higher in the mapping task than in the MIS- condition. Moreover, the more prior knowledge participants had the better their mapping performance. However, Table 12 also revealed that the mapping performance measure potentially led to a ceiling effect, because the related scores were very close to the maximum performance possible in this test (100% correct).

Table 13

Results of the regression analyses for predicting dependent variables related to learning outcomes for path c of the moderated mediation models

	Path c										
	Constant		Signaling			Prior knowledge			Signaling x prior knowledge		
Learning outcomes	*B*	SE_B	*B*	SE_B	β	*B*	SE_B	β	*B*	SE_B	β
Misconceptions[a] (*z*-standardized)	0.001	0.09	-0.09	0.17	-.05	0.02	0.004	.58***	0.01	0.01	.09
Recall[b] (*z*-standardized)	-0.01	0.11	0.16	0.22	.08	0.02	0.01	.31**	-0.02	0.10	-.26*
Comprehension[c] (*z*-standardized)	-0.03	0.07	-0.17	0.13	-.11	0.02	0.003	.67***	0.01	0.01	.13
Mapping performance[d] (% correct)	89.11	1.91	9.31	3.81	.27*	0.24	0.09	.29*	-0.08	0.18	-.05

$^{(*)}$ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.
[a] adj. $R^2 = .31$ ($p < .001$).
[b] adj. $R^2 = .14$ ($p = .005$.
[c] adj. $R^2 = .45$ ($p < .001$).
[d] adj. $R^2 = .12$ ($p = .010$).

8.3.3 Multimedia Signaling Effect on Visual Attention and Cognitive Load and the Influence of Prior Knowledge (*path a*)

In this chapter results related to *path a* of the moderated mediation models are presented. Thus, only interaction effects (multimedia signaling x prior knowledge) will be reported. Significant interactions indicate that prior knowledge influences the effect of multimedia signaling on visual attention and/or cognitive load measures.

Results related to the signaling effect on learning outcomes for different prior knowledge levels revealed a significant interaction between the signaling and prior knowledge for the recall measure assessing knowledge about the dissolution process. The pivotal part of the learning material referring to the recall measure was an eBook page explaining the dissolution process on a microscopic level by means of a sequence of four static pictures accompanied by four texts that were spatially integrated each next to one of the pictures (composite texts and pictures). Below the sequence of pictures with integrated text, an explanatory text was displayed to learners. On this page color coding was used to signal relations between the integrated text within the picture and the related pictures as well as between the text below the sequence of pictures and the pictures (see Figure 13).

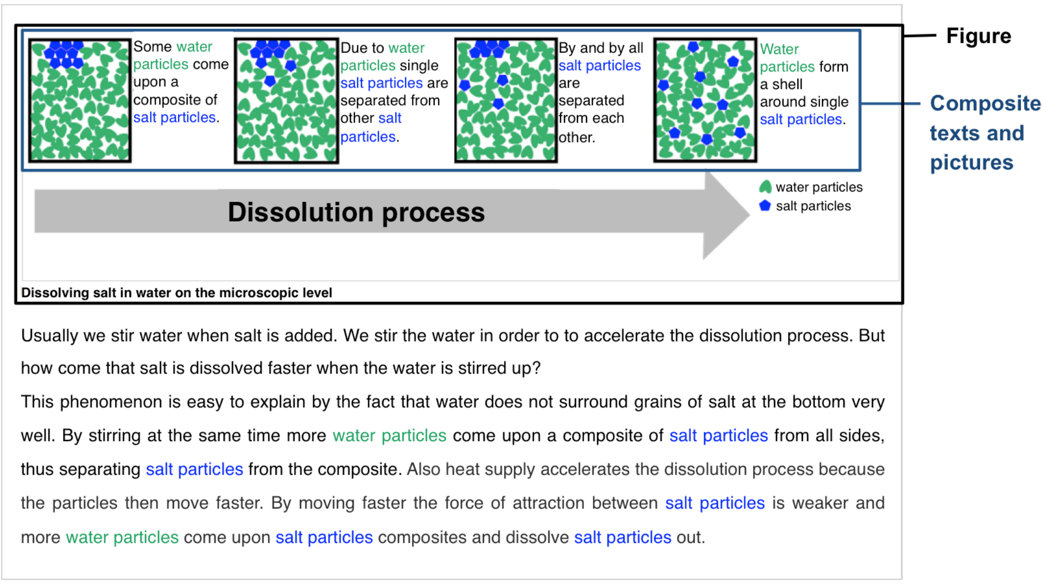


Figure 13. Learning material related to recall performance on the dissolution process on a microscopic level.

In order to shed light on underlying processes of the ERE related to recall performance, the effect of MIS for different prior knowledge levels on visual attention parameters related to the particular eBook page of the learning material (Figure 13) and cognitive load measures was investigated.

Visual attention parameters. Three separate regression models were computed each with (a) fixation time on the figure, (b) the number of transitions within composite texts and pictures within the figure and (c) the time to first fixating compound texts and pictures (estimated means and standard errors in Table 14). Prior knowledge, the signaling intervention as well as the interaction between both variables were included as predictors into each of the models (see Table 15 for results of regression analyses). Results revealed a marginally significant interaction between prior knowledge and the signaling intervention for the fixation time on the whole figure, $\beta = -.205$, $p = .065$ (see Figure 14). Thus, a simple slopes analysis with fixation time on the figure was conducted. Results revealed a significant difference between the signaling conditions only in the LPK group at one standard deviation below the mean, p = .036. LPK learners fixated on the figure longer when signals were present. There were no significant differences in the HPK group (p = .605).

Table 14

Estimated means and standard errors (in parentheses) of visual attention parameters and cognitive load measures for path a of the moderated mediation models

	MIS-		MIS+	
	LPK	HPK	LPK	HPK
Fixation time on figure (s)	20.08 (2.31)	16.95 (2.20)	27.02 (2.28)	15.25 (2.41)
Transitions between composite texts and pictures	12.41 (2.02)	11.68 (1.91)	17.44 (1.99)	9.54 (2.10)
Time to first fixating composite texts and pictures (s)	14.01 (2.01)	6.29 (1.91)	11.51 (2.05)	7.76 (2.10)
Subjective ECL rating	3.03 (0.42)	2.37 (0.40)	2.54 (0.42)	1.74 (0.44)
Subjective GCL rating	6.78 (0.33)	5.50 (0.31)	6.16 (0.32)	6.63 (0.34)
Pupil diameter	3.59 (0.07)	3.49 (0.07)	3.41 (0.07)	3.59 (0.08)

Note: Means and standard errors were estimated by means of regression analyses with signaling intervention, prior knowledge, and their interaction included as predictors.

The interaction between prior knowledge and the signaling intervention for the number of transitions between composite texts and pictures was also marginally significant, β = -.204, p = .079. A simple slopes analysis revealed a marginally significant difference between the signaling conditions for LPK learners, p = .080, in that these learners tended to perform more transitions between composite texts and pictures when signals were present than when they were not. Signaling group differences for the HPK group (p = .453) were not significant (Figure 14).

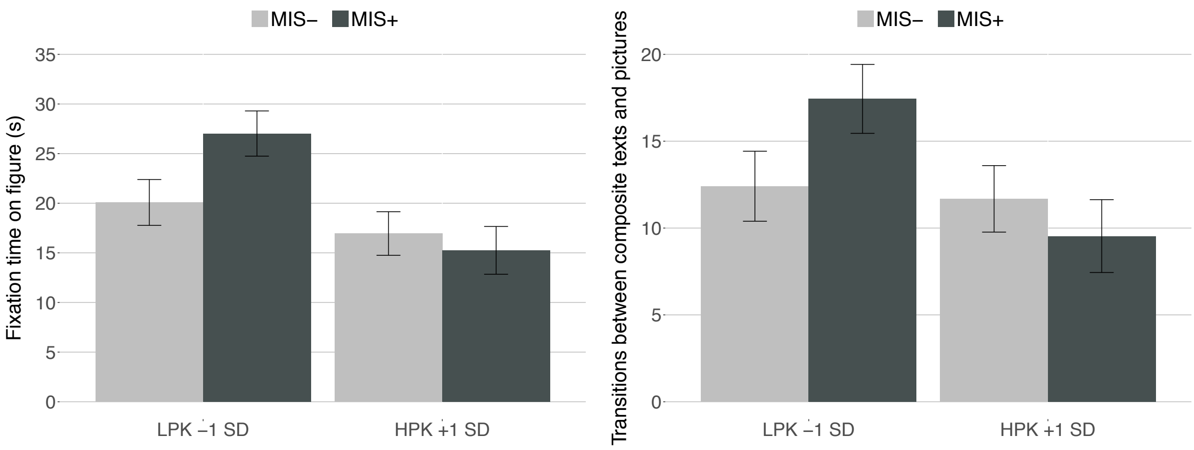


Figure 14. Fixation time on the figure (left panel) and number of transitions between composite texts and pictures (right panel) as a function of signaling at +/-1 *SD* of the prior knowledge measure (error bars: +/- 1 *SE*).

The regression model for the time to first fixating composite texts and pictures revealed no significant interaction effect (signaling x prior knowledge), $\beta = .112$, $p = .331$.

Cognitive load. Three separate regression models were computed each with one of the cognitive load measures as the dependent variable (regression analysis outcomes in Table 15): (a) average pupil diameter of students during learning with the particular eBook page, (b) subjective ratings of GCL and (c) ECL (estimated means and standard errors are displayed in Table 14). Prior knowledge, the signaling intervention as well as the interaction between both variables were included as predictors into the each of the models.

With regard to the average pupil diameter during learning there was a marginally significant interaction between signaling and prior knowledge, $\beta = .228$, $p = .056$. A simple slopes analysis revealed that the pupil diameter differed marginally significant between the signaling conditions for LPK learners, $p = .080$, whereas there were no differences for HPK learners, $p = .332$. The pupil diameter of LPK learners tended to be marginally smaller when MIS were present (see Figure 15).

Table 15

Results of the regression analyses for predicting dependent variables related to visual attention distribution and cognitive load measures for path a of the moderated mediation models

	Path a										
	Constant		Signaling			Prior knowledge			Signaling x prior knowledge		
Mediator variables	*B*	SE_B	*B*	SE_B	β	*B*	SE_B	β	*B*	SE_B	β
Fixation time on figure[a]	19.83	1.15	2.63	2.29	.13	-0.18	0.05	-.35**	-0.20	0.11	-.21(*)
Transitions between composite texts and pictures[b]	12.77	1.00	1.44	2.00	.08	-0.10	0.05	-.25*	-0.17	0.10	-.20(*)
Time to first fixating composite texts and pictures[c]	9.89	1.01	-0.52	2.01	-.03	-0.14	0.05	-.32**	0.09	0.10	.11
Subjective ECL rating[d]	2.43	0.21	-0.55	0.42	-.15	-0.02	0.01	-.21(*)	-0.004	0.02	-.02
Subjective GCL rating[e]	6.27	0.16	0.26	0.32	.09	-0.01	0.01	-.14	0.04	0.02	.30**
Pupil diameter[f]	3.52	0.04	-0.04	0.07	-.07	0.001	0.002	.06	0.01	0.003	.23(*)

(*) $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.
[a] adj. $R^2 = .15$ ($p = .003$).
[b] adj. $R^2 = .07$ ($p = .048$).
[c] adj. $R^2 = .08$ ($p = .034$).
[d] adj. $R^2 = .02$ ($p = .213$).
[e] adj. $R^2 = .09$ ($p = .028$).
[f] adj. $R^2 = .02$ ($p = .239$).

The model for GCL ratings revealed a significant interaction between signaling intervention and prior knowledge, β = .304, p = .009. A simple slopes analysis revealed that the GCL rating differed significantly between the signaling conditions for HPK learners, p = .016, whereas there were no differences for LPK learners, p = .185. For HPK learners, the subjectively rated GCL was higher when MIS were present (see Figure 15).

The ECL measure showed no significant interaction effect, β = -.022, p = .853.

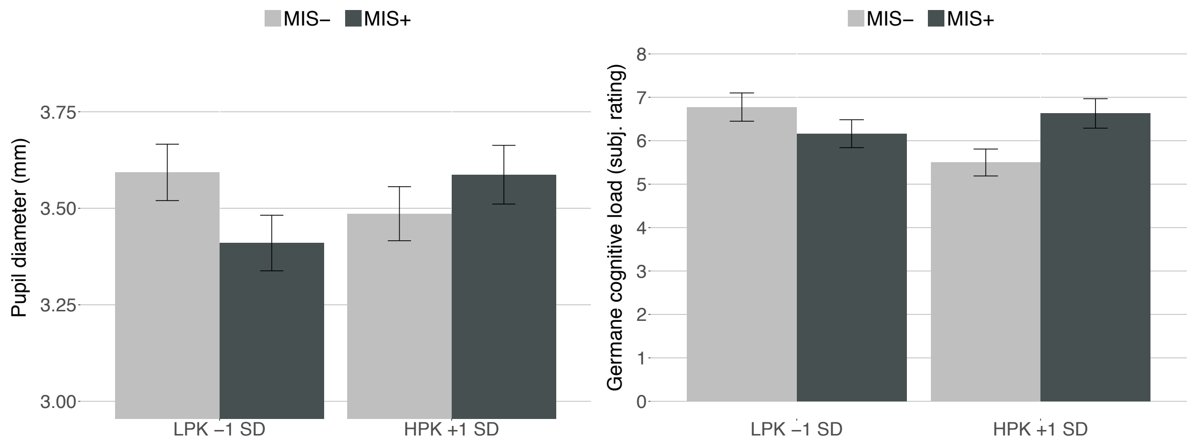


Figure 15. Pupil diameter (left panel) and subjective germane cognitive load rating (right panel) as a function of signaling at +/-1 *SD* of the prior knowledge measure (error bars: +/- 1 *SE*).

8.3.4 Effects of Visual Attention and Cognitive Load on Recall Performance and the Influence of Prior Knowledge (*path b*)

In this chapter results related to *path b* of the moderated mediation models 1 and 2 are presented. The models differed with respect as to whether prior knowledge affects the relation between visual attention and cognitive load measures and recall performance (model 2) or not (model 1). Significant interactions (visual attention/cognitive load x prior knowledge) indicate that prior knowledge influences the effect of visual attention and/or cognitive load measures on recall performance (as hypothesized in model 2). Conversely, significant main effects of visual attention/cognitive load on recall performance indicate that prior knowledge does not affect this relation (as hypothesized in model 1).

In order to investigate the effect of visual attention and cognitive load measures on recall performance, regression analyses for each of the mediator variables were conducted instead of including all mediator variables simultaneously into one regression model. This decision was made because significant correlations were obtained between the fixation time on the figure and transitions between composite texts and pictures, $r = .76$, $p < .001$, as well as between fixation time and the time to first fixation, $r = -.36$, $p = .002$. Moreover, the time to first fixation was correlated with the ECL measure, $r = .24$, $p = .047$. All remaining measures were not correlated, all $r_s < | .21|$. Therefore, the mediator variables were partly dependent and could thus not be included into one regression analysis. Instead, as Allison (1999) suggested, affected variables would have had to be deleted from the regression model, which would have lead to a loss of information regarding effects on *path b*. Against the backdrop of this reasoning, separate regression analyses for each mediator variable were computed. Estimated means and standard errors of the independent variables (mediator variables) are displayed in Table 16. The VIF was below 2.49 and the tolerance value greater than .40, which is in line with the more conservative rule of thumb by Allison (1999). Thus, multicollinearity due to correlations between mediator variables and the moderator prior knowledge was not present in the current data. Predictor variables were mean centered in order to be better able to interpret results.

The results revealed (marginally) significant interactions only between prior knowledge and visual attention distribution measures. The interactions turned out not to be significant for measures of cognitive load (see regression analyses results in Table 17).

Table 16

Estimated means and standard errors (in parentheses) of visual attention parameters and cognitive load measures for path b of the moderated mediation models

	LPK	HPK
Fixation time on figure		
short (-1 SD)	49.47 (3.16)	66.30 (2.88)
long (+1 SD)	55.02 (2.38)	52.46 (4.67)
Transitions between composite texts and pictures		
few (-1 SD)	49.67 (3.14)	65.27 (2.96)
many (+1 SD)	54.60 (2.48)	57.48 (3.66)
Time to first fixating composite texts and pictures		
fast (-1 SD)	55.87 (3.05)	57.27 (3.78)
slow (+1 SD)	53.66 (2.40)	69.711 (5.89)
Pupil diameter		
small (-1 SD)	55.93 (3.20)	62.40 (3.19)
large (+1 SD)	50.44 (3.19)	61.16 (2.85)
Subjective ECL rating		
low (-1 SD)	55.42 (3.24)	61.59 (2.92)
high (+1 SD)	51.59 (2.78)	62.57 (3.81)
Subjective GCL rating		
low (-1 SD)	48.84 (3.69)	63.59 (2.89)
high (+1 SD)	56.57 (3.14)	59.67 (3.35)

Note: Means and standard errors were estimated by means of regression analyses with signaling intervention, prior knowledge and the related interactions included as predictors.

The interaction between prior knowledge and transitions between composite texts and pictures was significant, β = -.28, p = .027 (see Figure 16). However, a simple slopes analysis revealed no significant differences between learners with few and many transitions in the LPK group, p = .199, and in the HPK group, p = .118. Hence, the main effect of prior knowledge was tested for learners performing either few or many transitions by means of recoding only the variable transitions between composite texts and pictures (+/- 1 SD). Results revealed that for learners with few transitions prior knowledge explained a significant amount of variance, β = .58, p = .001, in that these learners performed better in a recall test the more prior knowledge they had. In contrast, the recall performance of learners who performed many transitions did not differ based on their prior knowledge, β = .12, *p* = .451. This may suggest that conducting many transitions may compensate for the otherwise negative effect of having only little prior knowledge.

Table 17

Results of the regression analyses for predicting the dependent variable recall performance on path b of the moderated mediation models

	Path b✚										
	Constant		Mediator			Prior knowledge			Mediator variable x prior knowledge		
Mediator variables	*B*	SE_B	*B*	SE_B	β	*B*	SE_B	β	*B*	SE_B	β
Fixation time on figure[a]	55.81	1.59	-0.20	0.18	-.15	0.17	0.08	.26*	-0.02	0.01	-.38**
Transitions between composite texts and pictures[b]	56.76	1.55	-0.08	0.19	-.05	0.22	0.07	.34**	-0.02	0.01	-.28*
Time to first fixating composite texts and pictures[c]	59.13	1.64	0.29	0.28	.19	0.21	0.08	.33*	0.02	0.01	.32(*)
Subjective ECL rating[d]	57.79	1.59	-0.40	0.92	-.05	0.20	0.08	.32*	0.03	0.04	.09
Subjective GCL rating[e]	57.17	1.56	0.67	1.11	.07	0.21	0.07	.33**	-0.10	0.06	-.19
Pupil diameter[f]	57.48	1.55	-5.42	5.07	-.12	0.20	0.07	.32**	0.16	0.23	.08

✚ The factor signaling was included in all regression analyses related to *path b* and turned out not to be significant. Due to the table structure the factor signaling was not explicitly included in the results table for *path b*.
(*) $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.
[a] adj. $R^2 = .16$ ($p = .003$).
[b] adj. $R^2 = .12$ ($p = .012$).
[c] adj. $R^2 = .09$ ($p = .038$).
[d] adj. $R^2 = .07$ ($p = .073$).
[e] adj. $R^2 = .09$ ($p = .037$).
[f] adj. $R^2 = .07$ ($p = .059$).

For the fixation time on the figure the interaction with prior knowledge was also significant, β = -.38, p = .005 (see Figure 16). Simple slopes analyses revealed only a significant difference for HPK learners having short and long fixation times, p = .028, in that this group performed better when fixating the figure only shortly in contrast to a long fixation duration. Differences between students fixating short or long on the figure were not significant for LPK learners, p = .118. Moreover, the main effect of prior knowledge was tested for learners fixating the figure either for a shorter or longer time (+/- 1 SD). Results revealed that for learners fixating the figure only shortly prior knowledge explained a significant amount of variance, β = .62, p < .001, in that these learners performed better in a recall test the more prior knowledge they had. In contrast, the recall performance of learners who fixated the figure for a longer time did not differ based on their prior knowledge, β = -.09, p = .621.

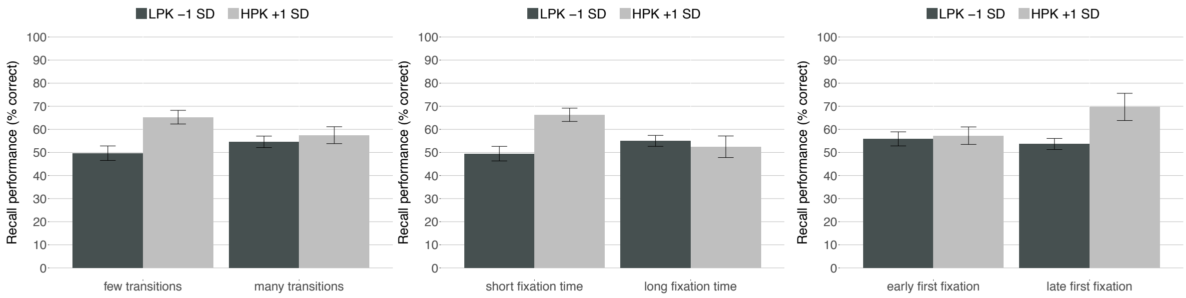


Figure 16. Recall performance as a function of few vs. many transitions between composite texts and pictures (left panel), short vs. long fixation time on figure (middle panel) and time to first fixating composite texts and pictures, right panel) and +/- 1 *SD* of the prior knowledge measure (error bars: +/- 1 *SE*).

A marginally significant interaction was obtained for the time to first fixating composite texts and pictures, β = .32, p = .084 (see Figure 16). Simple slopes analyses however revealed no significant difference between learners fixating composite texts and pictures either early or late during processing the materials for LPK learners, p = .494, and HPK learners, p = .154. Thus, the main effect of prior knowledge was tested for learners fixating composite texts and pictures either early or late.. Results revealed that for learners fixating later prior knowledge explained a significant amount of variance, β = .61, p = .009, in that these learners performed better in a recall test the more prior knowledge they had. In contrast, the recall performance of learners who fixated early during learning did not differ based on their prior knowledge, β = .05, p = .765.

8.3.5 Test of Joint Significance

In this chapter the final step of the moderated mediation analysis is reported. The remaining question is whether the effects obtained for *path a* and *path b* (indirect effects) of the moderated mediation models explain the occurrence of the partial expertise reversal of the multimedia signaling effect on *path c*. In order to determine the significance of the indirect effects the test of joint significance was used (Judd et al., 2014; MacKinnon et al., 2002; Muller et al., 2005) together with bootstrapped 95% confidence intervals using 5,000 bootstrap samples. According to this test, the indirect effect explains the partial expertise reversal of the multimedia signaling effect if each of the effects related to *path a* and *path b* are significant.

Let us recall that model 1 and 2 differed with respect to hypotheses related to *path b* (cf. Figure 11*).* In model 1 a direct effect of visual attention and/or cognitive load measures on learning outcomes was expected. Conversely, in model 2 it was assumed that prior knowledge moderates this effect (indicated by an interaction between visual attention/cognitive load measures and prior knowledge). Since no direct effects of visual attention and cognitive load on recall performance were obtained (see Table 17), model 1 was rejected. Results related to model 2 revealed (marginally) significant interactions between signaling and prior knowledge for *path a* and between mediator variables and prior knowledge for *path b* only for the visual attention distribution measures fixation time on the figure and transitions between texts and pictures. Hence, the test of joint significance was computed for these two measures related to the moderated mediation model 2. Data are displayed in Table 18.

Table 18

Test of joint significance with indirect effect estimates B, SE and bootstrapped lower and upper 95% CIs-values of interaction effects of path a and path b of the moderated mediation model 2

	p-value				95% CI_{Boot}	
	Path a	*Path b*	Indirect effect estimate *B*	SE_B	*LL*	*UL*
Fixation time on figure	.065	.005				
LPK			1.83	1.85	-0.22	7.26
HPK			1.12	2.11	-1.99	6.79
Transitions between composite texts and pictures	.079	.024				
LPK			1.41	1.50	-0.48	6.03
HPK			0.95	1.94	-1.31	7.43

According to the test of joint significance none of the measures revealed significant interaction terms for both *path a* (interaction effects for *path a* are only marginally significant) and *path b* of the moderated mediation model 2 (cf. Figure 11). Moreover, the bootstrapped 95% confidence intervals revealed that all indirect effect estimates were not significant. Thus, the partial expertise reversal of the multimedia signaling effect (*path c*) could not be explained by the hypothesized mediator variables and the influence of prior knowledge.

8.4 Summary & Discussion

The aim of Study 3 was to shed light on the underlying cognitive load and visual attention processes of the expertise reversal of the multimedia signaling effect by means of a moderated mediation analysis. In an eye tracking study, seventh, eighth, and ninth grade students learned with parts of a digital textbook for chemistry education. As in Study 2, an ecologically valid learning material and a strong control group was used. Besides learning outcomes, visual attention parameters, pupil diameter and subjective ECL and GCL ratings were assessed.

The key question related to potential underlying processes was whether the indirect effect of the moderated mediation model 1 or model 2 explains an expertise reversal of the multimedia signaling effect (see Figure 11). In model 1 it was assumed that prior knowledge moderates the effect of multimedia signaling on learning outcomes. It was expected that this ERE can be explained by the effects on visual attention and cognitive load measures. First, it was assumed that the effect of multimedia signaling on visual attention and cognitive load measures is influenced by learners' prior knowledge. This means that LPK and HPK learners would distribute their visual attention differently and experience different cognitive load depending on whether MIS were included in the materials. Second, it was expected that visual attention and cognitive load measures directly affect learning outcome measures. For example, high ECL should lead to a decrease in learning outcomes or many text-picture transitions should increase learning. Model 2 was equal to model 1 except for the effect of visual attention and cognitive load measures on learning outcomes. For this particular effect, model 2 assumed that this effect could also be influenced by prior knowledge. According to this assumption, learners who fixated the figure for longer durations for example would perform better when their prior knowledge was low but perform worse when they had high prior knowledge.

Results of the current study revealed a partial expertise reversal of the multimedia signaling effect in that LPK learners profited from MIS whereas HPK learners were not affected by MIS in their recall performance. But neither model 1 nor model 2 was significant in that their indirect effects explained the partial ERE. Nevertheless, when looking at the single paths in more detail, the pattern of results tends to speak in favor of model 2. Prior knowledge influenced the effect of multimedia signaling on visual attention and cognitive load measures as well as the effect of visual attention measures on recall performance.

LPK learners fixated the figure longer, made more text-picture transitions, and showed smaller pupil diameter when learning with the MIS+ compared to the MIS-version. The latter result points towards the notion that MIS reduce extraneous cognitive load for LPK learners by providing guiding information for cognitive processing of multimedia. Thereby, cognitive resources are made available for deep processing of materials (Chandler & Sweller, 1991; van Gog, 2014), as corroborated by longer fixation times and more frequent transitions.

HPK learners did not show any differences in visual attention and pupil diameter depending on the presence or absence of MIS. Nonetheless, they reported their GCL to be higher when learning with the MIS+ compared to the MIS- version. However, the pattern of results in not in line with the related hypotheses. It was expected that HPK learners refrain from deeper processing in the MIS+ condition, which would have led to a decrease in GCL in contrast to HPK learners in the MIS- condition (cf. McNamara et al., 1996). In the current study HPK learners reported the reverse. What becomes evident when descriptively comparing the pattern of results for pupil diameter and subjective GCL ratings is that these measures seem to be quite similar (cf. Figure 15). Thus, it might be that the GCL item that asked for the concentration during learning reflects some kind of general cognitive load, which cannot be attributed to GCL only. The result related to subjective cognitive load ratings, however, again points towards measurement issues (cf. De Jong, 2010).

The effects of visual attention measures on recall performance were affected by prior knowledge. HPK learners performed better in the recall test when they fixated the picture only shortly compared to when they fixated the picture long. Moreover, for learners performing few transitions, short fixations and that fixate relevant information late during the learning process their recall performance increased with increasing prior knowledge. Conversely, for learners performing many transitions, long fixations and that fixate relevant information early during the learning process prior knowledge had no influence on their learning outcome.

Although the moderated mediation models turned out not to be significant, the individual effects related to cognitive load and visual attention shed light on potential underlying processes of a partial ERE. However, further research is needed to uncover the process level of EREs.

What becomes evident by the approach to use a moderated mediation model in order to shed light on the influence of prior knowledge on processing multimedia material with and without MIS, is that multicollinearity is a general problem when visual attention and cognitive load measures are used as mediators especially for *path b* with mediator variables as independent variables, prior knowledge as moderator variable and learning outcome as dependent variable. In this regression model, two sources of multicollinearity can be present. First, the mediator variables have to be independent because they are added simultaneously as independent variables. The results of the current study as well as results by Schwonke, Berthold, and Renkl (2009), showed that different measures of visual attention can be correlated with each other. As in the current study, it might be that longer fixation times are positively correlated with transitions in that area. Thus, in these situations not all measures can be included simultaneously as independent variables. To combine different measures into one measure as suggested as one possible solution to multicollinearity problems by Allison (1999) and as realized by Schwonke, Berthold, and Renkl (2009), however leads to a loss of informative value for the investigation of the expertise reversal of the multimedia signaling effect. For the related Research Question 5 each of the possible eye tracking measures are relevant and need to be interpreted separately (e.g., fixation time, time to first fixation, and transitions), which is why in the current study separate regression analyses were computed for each mediator variable. A second source of multicollinearity can be introduced by correlations between mediator variables and the moderator prior knowledge. Hence, it is important to compute multicollinearity indicators such as the VIF (Allison, 1999; Mansfield & Helms, 1982) in order to make sure that the collinearity assumption in such regression models is met. Another suggestion might be to conduct studies with HPK learners only to investigate processing differences for learning material with and without MIS because there is already multimedia signaling research using eye tracking methodology for LPK learners (e.g., Jamet, 2014; Ozcelik et al., 2009; Ozcelik et al., 2010; Scheiter & Eitel, 2015). With a HPK sample it would then be sufficient to compute a simple mediation analysis (similar to the study by Scheiter and Eitel [2015] with LPK students) and at least prevent one of the sources of multicollinearity namely between the prior knowledge measure and measures of visual attention.

9. General Discussion

9.1 Summary of Results

The present thesis investigated one the most frequently used instructional design recommendations for multimedia materials, namely, the signaling principle. According to the signaling principle, correspondences between elements in texts and pictures should be highlighted by means of visual or discursive signals (van Gog, 2014). Signals are assumed to support learners in selecting relevant information in texts and pictures and organizing them into mode-specific models; moreover, there exist signals that are specifically designed in order to help learners integrate corresponding verbal and pictorial information into a coherent integrated mental model. This integration process of verbal and pictorial information is supposed to be crucial for meaningful learning from multimedia (Mayer, 2009; Schnotz, 2014), but learners often show inadequate and insufficient integration attempts only. Against this backdrop, the present thesis focused on the effect of multimedia integration signals (MIS), which are aimed at fostering integration of text and pictures.

By means of a review of the literature on the effects of signaling in multimedia learning, five potential boundary conditions were derived from the literature. Whereas four of them were related to the design of the materials (e.g., pictorial format of visualizations and pacing of the materials), the fifth potential boundary condition was related to the learners' prior knowledge. With respect to the latter variable, there is evidence suggesting that the effectiveness of various instructional techniques is dependent on the prior knowledge of learners (cf. Expertise Reversal Effect; Kalyuga et al., 2003) in that only learners with low prior knowledge will benefit from an improved instructional design, whereas learners with high prior knowledge will not show better learning outcomes (partial reversal) or even be hindered in their learning (full reversal).

The goal of the dissertation was to investigate (a) whether MIS aid learning, (b) which material-based boundary conditions would affect the effectiveness of MIS, (c) whether there would be an expertise reversal effect also for MIS, (b) whether this ERE would correspond to a partial or full reversal, and (c) how an ERE with respect to MIS could be explained at the cognitive processing level. To answer these questions, a meta-analysis, an experimental field study, and a laboratory experiment in which students' eye movements were recorded were conducted.

Study 1 was conducted as a meta-analysis aimed at investigating the effects of MIS for learning. The meta-analysis revealed a positive small-to-medium overall effect size in favor of MIS. Students who learned with multimedia material including MIS performed better in comprehension and transfer tests than when learning with multimedia material without MIS. This effect tended to be smaller for studies using strong control groups. While boundary conditions related to the design of the materials did not moderate the effects of MIS, as expected the learners' prior knowledge determined whether MIS improved learning: LPK learners profited from MIS, whereas HPK learners did not, indicating a partial ERE. However, the empirical basis of studies using HPK learners in their sample was extremely weak (only four out of 43 effect sizes). Moreover, most studies included in the meta-analysis were lab studies investigating university students learning from rather short multimedia learning materials. Consequently, it was unclear whether the findings of the meta-analysis would hold true in a more ecological valid setting, for instance, when studying curricular contents in school.

Study 2 was conducted as an experimental field study aimed at more systematically investigating the influence of prior knowledge on the multimedia signaling effect in an ecologically valid context. Eighth graders with different levels of prior knowledge learned with a digital multimedia textbook for introductory chemistry education in one of the two versions: (a) a strong control version with text signals that supported only the selection and organization of information from either text or pictures (MIS-), or (b) an experimental version with additional multimedia integration signals to support the integration of information from text and pictures (e.g., color coding, deictic references) (MIS+). Results of a contrast analysis revealed that LPK learners learned better with the MIS+ compared with the MIS- version, whereas adding MIS was detrimental for learning outcomes of HPK learners. This pattern of results was obtained for measures of comprehension and recall as well as for the number of misconceptions students had regarding the learning domain. Hence, Study 2 revealed a full ERE. In order to gain insight into potential underlying processes of an ERE, participants rated their subjective germane and extraneous cognitive load. The full ERE could, however, only partially be explained by cognitive load measures in that HPK learners reported higher extraneous cognitive load in the condition with MIS. Moreover, HPK students learning with the MIS+ version spend more time on task than HPK students learning with the MIS- version, whereas there were no differences in the LPK group. Therefore, the study pointed towards the redundancy explanation related to full EREs stating that MIS induce unnecessary processing of information that is redundant for HPK learners.

Study 3 was an a laboratory experiment in which eye tracking methodology was used in order to investigate the underlying processes of a potential expertise reversal of the multimedia signaling effect. Students from grades seven to nine learned with part of the digital textbook in a MIS+ or MIS- version from Study 2 whereby their eye movements and pupil diameter were recorded. In addition, students rated their extraneous and germane cognitive load. In order to be able to investigate whether potential changes in visual attention and cognitive load based on MIS and the prior knowledge level would explain learning outcomes, moderated mediation analyses were conducted. Results revealed a partial ERE for recall performance only in that LPK learners profited from MIS, whereas there were no differences in performance for HPK students learning with either the MIS+ and MIS-eBook version. Furthermore, MIS affected the viewing behavior and pupil diameter of LPK learners in that they fixated pictures longer, made more text-picture transitions, and tended to have a smaller pupil diameter in the MIS+ compared to the MIS-condition. MIS did not influence HPK learners in their viewing behavior; however, HPK learners indicated that they had concentrated more during learning when learning with the MIS+ in contrast to the MIS- eBook version. Different from what had been expected, neither students' viewing behavior nor their cognitive load ratings and pupil diameter could explain the partial ERE of the MIS effect on recall performance.

In the following, the results of the present dissertation will be discussed with respect to the key questions that guided this dissertation. Moreover, their practical implications as well as strength and limitations of the present dissertation will be addressed hereinafter.

9.2 Is Multimedia Signaling an Effective Design Measure to Foster Learning with Multimedia?

All three studies reported in the present thesis revealed that LPK learners showed better learning outcomes when MIS were included in the learning materials. Hence, the effectiveness of MIS for LPK learners was a stable finding throughout the dissertation studies across a variety of approaches and methods (meta-analysis, ecologically valid field study, eye tracking lab study).

Against the backdrop of the ERE (Kalyuga et al., 2003; Kalyuga, 2014), MIS are assumed to aid LPK learners in constructing a coherent integrated mental model from text and picture. Compared to HPK learners, LPK learners do not have schemas established in their long-term memory that they could use to map verbal and pictorial information. By highlighting corresponding information in text and picture, MIS should guide LPK learners in integrating verbal and pictorial information into a coherent mental representation. Hence, the effect of MIS for LPK learners should not be limited to recall performance but also be evident in measures of comprehension and transfer (cf. Mayer, 2014a). Indeed, the positive effect of MIS was found for comprehension performance in Study 1 as well as in Study 2. Moreover, in Study 2 positive MIS effects were also found with regard to recall performance and a measure of misconceptions for LPK learners. However, results of Study 3 were not in line with those of Study 1 and 2 because in Study 3 no effect of MIS on comprehension performance was obtained. MIS only improved recall performance of LPK learners. This divergent finding related to learning outcomes in Study 3 might be due to the presentation of MIS differing in their distinctiveness. The part of the learning material related to the recall measure included color coding whereas material related to the other learning outcome measures included discursive MIS such as deictic references only. The meta-analysis revealed that the multimedia signaling effect was small-to-medium for visual signals, whereas the effect was not significant for discursive signals. This result is in line with the assumption of Lemarié et al. (2008) that the realization property of a signal (visual versus discursive) affects whether a learner accesses the information provided by the signal. Hence, the overly reliance on discursive MIS for the materials relevant to improving comprehension and reducing misconceptions may explain why there was no effect of MIS for these measures in Study 3.

In conclusion, the present dissertation revealed that highlighting correspondences between text and pictures by means of MIS is an effective instructional support measure for LPK learners. It helps them in memorizing relevant information as well as constructing a coherent mental model, which is reflected in better recall and comprehension performance. As suggested by Study 3, visual signals may be more effective than discursive signals in this regard.

9.3 Which Material-Based Boundary Conditions Moderate the Multimedia Signaling Effect?

Four potential material-based boundary conditions for the multimedia signaling effect were derived from the literature and tested in the meta-analysis: (a) the pacing of materials, (b) the pictorial format, (c) multimedia mapping requirements, and (d) the distinctiveness of MIS.

For the more conservative approach to test all boundary conditions in one analysis, none of the material-based boundary conditions moderated the multimedia signaling effect. The less conservative approach, where each boundary condition was tested in a separate moderation analysis, revealed that only the pacing of materials turned out to be a significant moderator. Multimedia signaling was much more beneficial for system-paced materials, which do not allow the learner to control the presentation of the material in contrast to self-paced materials. The strict timing of the materials in a system-paced setting forces learners to attend to relevant information at the right time. Hence, they might use the information provided by MIS more strongly then when learning from self-paced instruction, which allows learners to go back and forth in their own pace (cf. Ginns, 2005; Tabbers et al., 2004; Schmidt-Weigand et al., 2010). Importantly, this influence of pacing as a moderator was not particularly strong, since it disappeared when the other boundary conditions were considered simultaneously.

9.4 Is there an Expertise Reversal Effect for Multimedia Integration Signals?

The meta-analysis as well as the two empirical studies in the present thesis confirmed the notion in the literature that prior knowledge will affect the degree to which learners will benefit from MIS (Kalyuga, 2007; Mayer & Sims, 1994; Schwonke, Berthold, & Renkl, 2009).

The more conservative approach of the meta-analysis revealed that only the prior knowledge of learners moderated the multimedia signaling effect. LPK learners profited from MIS with regard to their comprehension performance, whereas MIS did not affect comprehension performance for HPK learners. This finding is in line with a partial ERE that was also revealed in Study 3. The expertise reversal of the multimedia signaling effect was also obtained under ecologically valid conditions in Study 2. However, Study 2 showed a full ERE in that LPK learners profited from MIS while in contrast to Study 1 and 3 HPK learners were even hindered in learning. This contradictory pattern of results is discussed in more detail in the next chapter.

Taken these results together, they point towards the importance to consider prior knowledge when using MIS as an instructional support measure in multimedia learning materials. This finding is in line with numerous studies that revealed EREs for different kinds of instructional techniques (cf. Kalyuga, 2007).

9.5 Does the Influence of Prior Knowledge Lead to a Partial or Full Expertise Reversal of the Multimedia Signaling Effect?

The nature of how prior knowledge moderates the effect of MIS remains an open question based on the present studies. Whereas Study 1 as well as the laboratory experiment (Study 3) yielded a partial reversal, the field experiment (Study 2) showed a full reversal (cf. Kalyuga & Renkl, 2010). There are at least three possible factors that may be responsible for these inconclusive results: (a) different underlying cognitive processes, (b) statistical power of studies on the ERE, and (c) categorization of prior knowledge.

First, the assumption that different underlying cognitive processes cause the occurrence of partial and full EREs related to the multimedia signaling is mainly based on different theoretical explanatory approaches derived from the literature related to the two ERE types. A partial ERE might occur because HPK learners are able to apply their schemas in long-term memory to identify and map text-picture correspondences. Thereby, HPK learners can compensate for missing guiding information provided by MIS (ability-as-compensator hypothesis; Mayer & Sims, 1994). Full EREs might be due to the fact that elaborative processes are suppressed MIS are present, resulting in a decrease in GCL, or that MIS induce unnecessary processing of redundant information, which should result in an increase of ECL (cf. chapter 4.2). But results of Study 2 and 3 revealed no results supporting either one of the explanatory approaches for partial and full EREs unambiguously. This may have partly been due to measurement problems related to subjective cognitive load ratings (De Jong, 2010).

Second, the statistical power of studies could be a reason for revealing partial or full EREs. More specifically, the finding of partial EREs might be due to a lack of power regarding statistical analysis to reveal a disordinal interaction. Indeed, the statistical power of Study 3 was lower than in Study 2 because students had to come to the lab or had to be tested in the afternoon in schools individually. This situation probably led to a rather small sample size in Study 3, which in turn might have been a reason for revealing a partial ERE in Study 3. Furthermore, the meta-analysis also revealed a partial ERE with a very weak empirical basis for studies using HPK learners in the sample (k = 4 effect sizes for HPK learners). However, based on the present dissertation and previous studies (e.g., EREs review by Kalyuga, 2007) it is not possible to decide whether the statistical power of studies influences the occurrence of partial or full EREs related to multimedia signaling or whether other factors play a role.

Third, another factor that potentially drives the occurrence of partial or full EREs is the categorization of prior knowledge. Since it is not possible to have a standardized prior knowledge measure in research on the ERE because prior knowledge is dependent on the learning material, results of studies on the ERE may differ in general. Apart from the decrease of statistical power from Study 2 to Study 3, in addition the variance of prior knowledge increased from Study 2 to Study 3. In Study 2 only eighth graders took part, whereas in Study 3 also seventh graders inexperienced in the subject chemistry and very experienced students from grade nine took part. The partial ERE in Study 3 might thus also have been driven by ninth graders who had two or even more years of school experience with science in general and with chemistry in particular. Hence, these HPK learners had automated schemas established related to basic science concepts such as the microscopic level. Hence, they might have been able to ignore instructional support provided by MIS better than HPK learners in grade eight with potentially less elaborated schemas related to the subject chemistry (cf. ability-as-compensator hypothesis; Mayer & Sims, 1994; information-reduction hypothesis; Haider & Frensch, 1999). Nevertheless, the variance in prior knowledge in Study 2 was representative for students who come from the same grade level and school type. Therefore, the LPK, MPK and HPK classification in Study 2 best reflects upon ecologically valid conditions and thus, provides better insight into the nature of the expertise reversal of the multimedia signaling effect in the field.

Against the backdrop of these potential influencing factors for the occurrence of partial and full EREs with respect to multimedia signaling, it cannot be conclusively decided what factors drive these occurrences. The statistical power of studies and the classification of prior knowledge might play an important role, however, systematic analyses are lacking. Thus, future research is needed that systematically takes the statistical power of studies into account for instance as a moderator variable in a meta-analysis on the ERE related to instructional techniques. In order to circumvent measurement and classification problems of prior knowledge, research might compare the effectiveness of MIS for HPK learners with varying years of experience in science subjects (e.g., eighth, ninth and tenth grade of the same school type). In this way, one could investigate whether the effect of MIS changes with increasing experience with science subjects and related concepts. However, the question how partial and full EREs can be explained is not only an empirical problem. The theoretical explanations underlying EREs are rather vague and partly inconsistent. For example, related to the redundancy explanation for a full ERE Schnotz (2010) pointed out that it is questionable why a support measure that does not cognitively overload LPK learners should overload HPK learners. Based on the

CLT, HPK learners are expected to experience overall lower cognitive load than LPK learners because they can use automated schemas in long-term memory. Hence, the redundancy explanation is inconsistent with the CLT itself (Schnotz, 2010). Therefore, the theoretical explanations for EREs are not necessarily helpful for deciding what factors drive the occurrence of partial and full EREs. Moreover, the vagueness of the description of the CLT impedes forming precise hypotheses. Taken these limitations together it becomes evident that they also apply to the research field itself.

9.6 How Can a Partial or Full Expertise Reversal of the Multimedia Signaling Effect be Explained?

Ausubel, Novak, and Hanesian (1978) stated: “The most important factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly” (p. 163). Given the influence of prior knowledge on the effectiveness of MIS revealed by the present dissertation, this notion seems to be relevant in the context of multimedia learning. However, in order to do so one has to understand the needs of learners with differing prior knowledge with respect to processing of multimedia materials. To gain insight into underlying processes of the expertise reversal of the multimedia signaling effect, subjective cognitive load ratings were assessed in Study 2 in a field context. Moreover, in Study 3 in addition to subjective cognitive load ratings eye movements of learners and their pupil diameter during learning was recorded.

With respect to cognitive load measures, items related to extraneous and germane load were assessed. Regarding ECL, it was expected that a full ERE is related to an increase of ECL for HPK learners because they have to process redundant information. The ECL item was the same in Study 2 and 3 asking learners for their difficulty to understand the contents (adapted from Kalyuga et al., 1998). In line with the redundancy explanation of a full ERE, HPK learners in Study 2 rated their ECL higher when MIS were included in the materials. However, this pattern was not revealed in Study 3 with the same ECL item. What needs to be considered in this context is that Study 3 revealed only a partial ERE in that HPK learners where not hindered in learning. The redundancy explanation, thus, does not apply in Study 3.

On the other hand, on a descriptive level HPK learners learning with the MIS+ version tended to perform worse than when they learned with the MIS- version. Therefore, the ECL measure used in Study 2 and 3 will be discussed in the light of the redundancy explanation.

A relevant question in this context is how did ECL ratings of HPK learners differ between Study 2 and 3. When comparing the means of ECL ratings of HPK learners in the two studies for each eBook version it becomes evident that these differed considerably for the MIS+ eBook version. HPK students learning with the MIS+ eBook version showed a large difference in their ECL ratings between Study 2 and 3 (d = 1.80). Conversely, HPK students learning with the MIS- eBook version showed a small-to-medium difference in their ECL ratings between Study 2 and 3 (d = 0.40). The pronounced deviation related to ECL ratings of HPK learners between Study 2 and 3 for the MIS+ eBook version might of course be related to the presentation of only parts of the learning material and hence the MIS used in Study 3 compared to Study 2. Nevertheless, in Study 2 ECL ratings of HPK learners increased when learning with the MIS+ compared to the MIS- version, whereas in Study 3 the ratings decreased when HPK learners learned with the MIS+ in contrast to the MIS- version.

Regarding GCL it was expected that a full ERE occurs due to a suppression of elaboration processes for HPK learners when MIS are present (cf. McNamara et al., 1996). Again, one has to be careful with the interpretation of the GCL rating in Study 3 because only a partial rather than a full ERE was obtained. Different GCL items were used in Study 2 and 3 due to potential measurement errors in Study 2. In Study 2 learners were asked to rate how much effort they invested to understand the contents (adapted from Paas, 1992). In Study 3 the item asked for how much a learner concentrated during learning (cf. Cierniak et al., 2009). Results of Study 3 revealed that HPK learners rated their GCL higher when learning with the MIS+ compared to the MIS- condition. Thus, the GCL pattern in Study 3 was similar to the ECL pattern in Study 2. It seems that although the used ECL and GCL items in Study 2 and Study 3 have been used before in research, they do not distinguish well between types of cognitive load.

Taken together, these findings corroborate the notion that research issues related to subjective cognitive load ratings exist (cf. de Jong, 2010; Schnotz & Kürschner, 2007). It might for example be that in contrast to the conclusion of Paas et al. (2003) learners were unable to contemplate on different sources of cognitive load and rather rated some kind of overall cognitive load experienced during learning. The pattern of results regarding pupil diameter of learners in Study 3 does partly corroborate this assumption since pupil diameter is supposed to be a rather general measure of cognitive load. LPK learners descriptively tended to rate their GCL lower and had a significantly smaller pupil diameter when learning with the MIS+ compared to the MIS- version. Vice versa, for HPK learners the GCL rating was significantly higher when learning with the MIS+ compared to the MIS- condition while a corresponding difference was also present in their pupil diameter, albeit not significant. However, the outlined descriptive relations between the ECL measure in Study 2 and the GCL measure in Study 3 as well as the measure of pupil diameter are speculative and warrant further study.

With regard to eye tracking parameters, Study 3 revealed that MIS led LPK students to fixate on relevant pictorial information longer and make more transitions between texts and pictures (MIS+) compared to when there were only text signals (MIS-). Although this finding is in line with previous research (cf. guiding-attention hypothesis; Ozcelik et al., 2010; integrative processing; Mason, Tornatora, & Pluchino, 2013; O'Keefe et al., 2014) these changes in visual attention did not explain better recall performance for LPK students learning with the MIS+ compared to the MIS- eBook version. Nevertheless, a mediation analysis by Scheiter and Eitel (2015) revealed that more frequent fixations and earlier fixations on relevant information due to the guiding function of MIS explained better learning performance for LPK learners. What moreover becomes evident is that contrary to what is common practice in eye tracking studies in the context of multimedia learning (e.g., Jamet, 2014; Ozcelik et al., 2010) causal relationships need to be tested by means of mediation analyses in order draw conclusions related to underlying processes of the ERE. It is not sufficient to test effects of MIS on visual attention and learning outcome measures separately to conclude that both effects are related.

The visual attention distribution of HPK learners was not affected by the two different eBook versions MIS- and MIS+. Moreover, they were also not influenced by MIS regarding their learning performance indicating a partial ERE. Consequently, it seemed that as hypothesized in the ability-as-compensator hypothesis HPK learners were actually able to ignore the guiding function of MIS because MIS did not affect their viewing behavior. Instead, it seems that they used their background knowledge to guide their processing of the materials (cf. ability-as-compensator hypothesis; Mayer & Sims, 1994; information-reduction hypothesis; Haider & Frensch, 1999).

However, those HPK learners who fixated relevant pictorial information long showed worse recall performance than those who fixated this information only shortly. Albeit not significant but similar to the latter finding, recall performance of HPK learners tended to be better when these learners made only few text-picture transitions compared to when they performed many transitions and recall performance also tended to be better when fixating relevant information late during the learning process than early. These results suggest that although MIS did not influence the visual attention distribution of HPK learners the multimedia processing behavior that is supposed to be supported by MIS (long fixation on pictorial information, many text-picture transitions, early fixation of relevant information) would not be beneficial for expert learners but rather hinder their learning (integrative processing; Mason, Tornatora, & Pluchino, 2013).

To sum up, the present dissertation cannot provide a conclusive answer to the question what processes explain a partial or full expertise reversal of the multimedia signaling effect. Future research is needed that investigates the effectiveness of MIS in the field context with an ecologically valid sample and learning material that assesses not only subjective cognitive load ratings but also more objective measures such a visual attention measures. Moreover, it might be advisable to also assess verbal protocols (e.g., cued retrospective reports on eye movements) in order to gain further insight into the nature of the full expertise reversal of the multimedia signaling effect (e.g., Jarodzka et al., 2010). A more thorough investigation is necessary in order to be able to decide how students have to be taught according to their prior knowledge, as suggested by Ausubel et al. (1978).

9.7 Practical Implications

Against the backdrop of the present dissertation, practical implications of the results must be derived in a differentiated manner. On the one hand, it can be concluded that MIS are an effective instructional support measure for LPK learners learning with multimedia materials. Related to the initial question of the present dissertation, highlighting corresponding elements in texts and pictures seems to be a valid recommendation for the design of multimedia learning material at least when LPK students are supposed to learn with these materials. On the other hand, the latter scenario is rather unlikely to be found in school classes. As Study 2 showed, students in the same grade and same school type differed with respect to what they already knew about the topic of the eBook. HPK students within this sample were hindered in learning when MIS were present with respect to various learning outcome measures. This result imposes an ethical conflict from a normative standpoint. However, further research is needed in order to gain insight into the underlying cognitive processes that lead to a decline in performance for expert learners when MIS are included in the materials. It might be advisable for example to investigate whether it is more beneficial for HPK learners to learn with material without any MIS at all or with a reduced number of MIS. For instance, MIS could only be displayed in parts of the learning material dealing with a misconception that became evident in a pre-test or they could be gradually faded out during the learning process (fading procedure, cf. Renkl, Atkinson, Maier, & Staley, 2002). The latter would be in line with one of the SEASITE principles according to which instructional support should be provided on demand depending on the learners' progress (Renkl, 2002).

If further research adds to clarify the mechanisms of a full expertise reversal of the multimedia signaling effect a practical implication is to consider prior knowledge of learners in the design of learning material (cf. learner-tailored instruction; Kalyuga, 2007). Especially digital learning material, like the digital textbook used in the current studies, could be designed and programmed adaptively to the prior knowledge of the learner. Thereby, well-designed digital learning material can be a key to support students in the best possible way when learning from multimedia.

9.8 Strengths and Limitations

As with any piece of scholarship there are strengths and limitations related to the present dissertation. To begin with, strengths of the present thesis related to the (a) methodological diversity, (b) ecological validity, (c) consideration of learning-oriented but also process-oriented research, and (d) the use of moderated mediation analyses will be outlined in the following.

First, different methods were used to investigate the effectiveness of multimedia signaling, its boundary conditions and underlying processes of a partial or full ERE. A meta-analysis was conducted in order to systematically review the literature and investigate the effectiveness of multimedia signaling and its boundary conditions. Study 2 was conducted as an experimental field study to address limitations of studies included in the meta-analysis. A laboratory experimental eye tracking study was conducted in order to shed light on the processes underlying an ERE. Thus, the studies of the present thesis confirmed the effectiveness of MIS for LPK learners based on a variety of methods. By using different methods, potential research issues such as the common practice to conduct separate analyses related to process and learning outcome measures rather than using a mediation analysis were elucidated that might further advance research in the field.

Second, Study 2 was an ecologically valid field study conducted in schools during regular chemistry lessons to investigate whether results obtained in the meta-analysis can be generalized to situations in the field. The content of the learning material was aligned to the curriculum and rather extensive resulting in a learning time of about four school lessons. In contrast, most studies included in the meta-analysis used rather short materials including few pages and short learning times. Moreover, a strong control group was implemented who learned with a version of the eBook that was well designed. It was important to test the robustness of the multimedia signaling effect with a strong control group because poor control conditions might potentially maximize effects (cf. Schwonke, Renkl, et al., 2009). Importantly, the setting of Study 2 allowed to test whether results obtained mostly in lab studies in the meta-analysis could actually be generalized in a one-to-one fashion to situations in the field. Due to the efforts related to the digitalization of educational materials it is crucial to test instructional support measures in the field once a robust picture of evidence was obtained in lab studies.

Third, in the present thesis the multimedia signaling effect and its boundary conditions were not only investigated with respect to learning outcomes but also by shedding light on the process level regarding cognitive load and visual attention. Addressing processes underlying an ERE empirically was necessary in order to be able to decide which of the existing theoretical explanations best reflect upon the reasons for the occurrence of EREs. Being able to understand underlying processes of the ERE is a prerequisite for designing individualized instructions. Only when the mechanisms of a full ERE are reliably clarified, support measures that aid HPK learners can be designed.

Fourth, to be able to explain the occurrence of an ERE on the process level moderated mediation analyses were conducted. By using this method the indirect effect related to cognitive load and visual attention measures could be tested. Thus, it was possible to decide whether the effect of multimedia signaling on process measures could explain an ERE. Contrary, in prior eye tracking studies it was common practice to investigate causal relationships between eye tracking measures and learning outcomes based on separate analyses (e.g., Jamet, 2014; Ozcelik et al., 2010) rather than using mediation analyses.

However, the present dissertation is also limited by several factors such as (a) the usage of subjective cognitive load ratings, (b) theoretical explanations for EREs, and (c) methodological issues such as the sample sizes and the categorization of prior knowledge.

First, as already discussed, the results obtained for subjective cognitive load ratings corroborated the notion that research issues related to these ratings exist (cf. de Jong, 2010; Schnotz & Kürschner, 2007). The benefit of using these measures was clearly that they could be easily implemented in the studies in particular in the field. Moreover, although several research issues related to cognitive load ratings are well known from the literature they are frequently used in cognitive load research (De Jong, 2010; Schnotz & Kürschner, 2007). However, results for ECL and GCL ratings obtained in the present thesis were not consistent thereby suggesting that the items that were used did not well distinguish between different types of cognitive load. In addition, as cognitive load was rated in the post-test after learning in both studies, learners might have rated a peak rather than an average of their experienced cognitive load during learning (cf. Schmeck et al., 2014). These presumptions, however, are speculative and cannot be resolved by the present data. In the long-run it might be advisable to either develop reliable cognitive load measures that include more than only one item or to use more objective measures such as the pupil diameter or the heart rate variability of students. Moreover, the reliability of a subjective cognitive load measure could be validated by relating it to actual physiological states.

Second, the explanatory approaches for the ERE are rather vague and partly inconsistent. As Schnotz (2010) pointed out the redundancy explanation related to full EREs is inconsistent with the CLT itself. Regarding the CLT, a support measure that does not overload LPK learners should also not overload HPK learners. The redundancy explanation, however, assumes that HPK learners are overloaded due to processing of redundant information. Moreover, research suggests that with increasing expertise learners become better at ignoring unnecessary information (cf. information-reduction hypothesis; Haider & Frensch, 1999). Scheiter and Eitel (2015) even found that LPK learners were able to ignore guiding information that was not helpful for the task. Therefore, the theoretical basis of the ERE is not consistent and might thus be a weak basis for investigating underlying processes of the ERE related to cognitive load.

Third, as with studies on the ERE in general, the categorization of prior knowledge and the sample size of the studies might strongly influence the interpretation of the effectiveness of an instructional technique. As discussed above (cf. chapter 9.5), it is not possible to have a standardized prior knowledge measure in research on the ERE because prior knowledge is dependent on the learning material. Hence, results of studies on the ERE might differ in general. Depending on how much background knowledge the identified HPK learners have they might not be affected or rather hindered in learning. This could be one of the reasons for the occurrence of partial or full EREs. Moreover, the sample size used in studies on the ERE might be relevant because depending on the sample size statistical power might not be sufficient to reveal a disordinal interaction. Hence, low statistical power can also lead to partial EREs. The data of the present thesis does not allow to disentangle these potential sources for the occurrence of partial and full EREs. Further empirical research is needed that systematically takes these two factors into account.

Two aspects of multimedia signaling that were not the focus of Study 2 and 3 but might add to the understanding of the influence of multimedia signaling on learning outcomes for different levels of expertise is the amount of signals and the type of signals used in the materials. Seufert (2003) suggested that MIS might be more or less suitable for supporting multimedia learning processes of learners with differing prior knowledge based on their salience (cf. directivity; Seufert, 2003). This notion was based on Vygotski's concept of the zone of proximal development (ZPD; Vygotski, 1963). Schnotz (2010) pointed out that instructional support should be tailored to the individual ZPD of each learner, which is an individual range between tasks with differing difficulty. The lower limit of the ZPD are tasks that learners can perform without further support whereas the upper limit are tasks that are more difficult but can still be solved by the learner with adequate instructional support. Beyond this range of the ZPD a learner won't perform any meaningful cognitive processes aiming at successful learning because they are either demanded too little effort (below the lower limit) or are overstrained by the task requirements (above the upper limit) (cf. Vygotski, 1963). Against the backdrop of the concept of the ZPD and similar to results obtained by McNamara et al. (1996), it might be that multimedia material including MIS demands too little effort from HPK learners, which is why they refrain from deeper processing. A way to further investigate this notion would be to use a different amount or/and different types of MIS (salient versus discursive) in multimedia materials. In Study 2 and 3 signaling measures were only chosen based on the content, meaning that a more complex text-picture combination with many verbal and pictorial elements that need to be mapped in order to build an integrated mental model required more signals than a simple one. In the meta-analysis by Richter et al. (2016) the moderating role of multimedia mapping requirements was investigated, which might be a proxy for the amount of signaling. The more verbal and pictorial elements need to be mapped, the more signals could potentially be used. However, the meta-analysis revealed no moderating role of mapping requirements of the materials on the signaling effect. Nevertheless, it is an open question whether the amount of MIS potentially help to gain further insight into the ERE and how to overcome a potential decline in performance for HPK learners. Against the backdrop of the concept of the ZPD (Vygotski, 1963), it might for example be that a reduced number of signals is more beneficial for HPK learners, whereas LPK learners perform best when all text-picture correspondences are highlighted by means of MIS.

Another related question that was not addressed by the studies in the present dissertation is whether particular types of MIS are better suited for LPK or HPK learners to support learning. Lemarié et al. (2008) suggested that the salience of a signal influences how easily readers can access a signal. Accordingly, MIS with a rather salient visual appearance such as color coding might be more easily accessible by learners than discursive signals such as deictic references. As a consequence, salient visual signals might be better suited for LPK learners who profit from instructional guidance because they can access them more easily than discursive signals. Conversely, HPK learners may not (always) need instructional guidance and therefore may want to decide whether to access signaled information. In this situation, less salient discursive signals, which can be more easily ignored, might support HPK learners better in their effort to integrate verbal and pictorial information than the more salient visual signals. The role of types of signals for learners with different prior knowledge levels should thus be subject for further studies.

Summary

The aim of the present dissertation was to investigate one of the most frequently used instructional support measures for multimedia learning, namely, signaling of text-picture relations and its boundary conditions. Learning with multimedia involves processing of information provided by text as well as by corresponding pictures. Numerous studies have shown that learning with text and picture was more beneficial regarding learning performance than when learning from text alone (Mayer, 2014a). When learning with multimedia the crucial step on the processes level is the integration of verbal and pictorial information into a coherent integrated mental representation (cf. Butcher, 2014; Mayer, 2009; Schnotz, 2014). The establishment of such a coherent integrated mental representation in turn is supposed to be necessary for meaningful learning with multimedia. However, since students often struggle in particular with the integration of information from texts and pictures (e.g., Hannus & Hyönä, 1999; Renkl & Scheiter, 2015) several instructional support measures such as highlighting of corresponding elements in text and picture by means of multimedia integration signals (MIS) were recommended (Van Gog, 2014). MIS can be implemented for example by means of color coding (highlighting corresponding elements in text and picture in the same color) or by means of deictic references (text that refers to elements in the picture, e.g. "In the picture you can see element x…"). Besides supporting the selection and organization of information from texts and pictures, MIS are supposed to mainly foster the integration of verbal and pictorial information into a coherent mental representation. Hence, MIS were expected to support multimedia learning.

To investigate the effectiveness of MIS as an instructional support measure for multimedia learning and its boundary conditions a meta-analysis was conducted (Study 1). Results revealed that the domain-specific prior knowledge influenced the effectiveness of MIS. MIS were beneficial for low domain-specific prior knowledge (LPK) learners whereas comprehension performance of high domain-specific prior knowledge (HPK) learners was not affected. Therefore, the result of the meta-analysis was in line with the assumption that the effectiveness of instructional techniques is dependent on learners' prior knowledge as stated by the expertise reversal effect (ERE; Kalyuga et al., 2003). However, research on the ERE did not

only reveal partial EREs in that LPK learners profit whereas the performance of HPK learners is not affected by an instructional support measure (as revealed by the meta-analysis). Rather also full EREs were obtained showing that HPK learners were hindered in learning by an instructional technique whereas LPK learners profited regarding learning outcomes (cf. Kalyuga, 2007).

In order to be able to investigate whether the finding of the meta-analysis could be generalized in to situations in the field, an ecologically valid study was conducted in schools (Study 2). Contrary to most of the studies included in the meta-analysis, in Study 2 (a) comprehensive ecologically valid learning material was used, (b) not only LPK but also HPK learners were included in the sample, and (c) the condition with material including MIS (MIS+) was compared to a strong control condition (MIS-). The related research question was whether there would be a partial or full ERE related to MIS in the field. Moreover, it was hypothesized that extraneous and germane cognitive load add to the explanation of a partial or full ERE. Results of Study 2 revealed a full ERE. LPK learners profited from learning with the multimedia eBook about a chemistry model including text signals and MIS (MIS+) compared to a version with text signals and basic MIS only (MIS-). Conversely, HPK learners were hindered in learning. This pattern was stable among different learning outcome measures. Moreover, HPK learners tended to experience more difficulty during learning and spend more time with the eBook in the MIS+ compared to the MIS- condition.

To gain further insight into underlying processes of the expertise reversal of the multimedia signaling effect, a laboratory experiment including eye tracking methodology was conducted (Study 3). Study 3 used a similar sample, parts of the learning material, and the same signaling manipulation as in Study 2. The related moderated mediation hypothesis was that MIS would alter visual attention distribution and cognitive load differently based on learners' prior knowledge. Moreover, visual attention distribution and cognitive load measures were expected to either directly affect learning outcomes or to affect learning outcomes depending on prior knowledge. In turn, these effects of process measures and prior knowledge were assumed to explain a potential ERE. Results revealed a partial ERE for recall performance only. Furthermore, MIS influenced the viewing behavior and pupil diameter of LPK learners in that they fixated pictures longer, made more text-picture transitions, and tended to have a smaller pupil diameter in the MIS+ compared to the MIS- condition. MIS did not influence HPK learners in their viewing behavior; however, they indicated that they had concentrated more during learning when learning with the MIS+ in contrast to the MIS- eBook version. But different from what was hypothesized in the moderated mediation models neither learners' viewing behavior nor their cognitive load ratings and pupil diameter did explain the partial ERE.

To sum up, the present dissertation confirms the effectiveness of multimedia signaling as an instructional support measure for multimedia learning, however, limited to LPK learners only. Thus, MIS seem to support the integration of verbal and pictorial information into a coherent integrated mental model for these types of learners. In line with evidence on the ERE, HPK learners were not affected with regard to learning (partial ERE) or were even hindered in learning (full ERE) when MIS were present (cf. Kalyuga, 2007; Kalyuga & Renkl, 2010). The students' viewing behavior as well as their cognitive load ratings and pupil diameter did not explain the occurrence of EREs related to MIS.

A reason for concern can be seen in the finding of a full ERE in the field study under ecologically valid conditions for all learning outcome measures (Study 2). If future research would corroborate the finding of a full ERE related to MIS, from a normative standpoint an ethical conflict would arise. Classes are supposed to be heterogeneous with respect to their prior knowledge about specific topics (Slavin, 1987). Recommending multimedia signaling as a general support measure for multimedia learning might entail learning drawbacks for HPK learners in a class. In order to counteract this ethical conflict, individualized learning materials could be used (cf. Kalyuga, 2007). The increasing use of digital learning material in education provides the opportunity to develop adaptive instructions based on the current knowledge level of learners. However, in order to do so further insight into the nature of the full expertise reversal of the multimedia signaling effect is necessary to be able to design instructions best supporting HPK learners. Future research should thus consider to test the effectiveness of MIS for in different settings for example by manipulating the salience of MIS, the amount of MIS or fading out MIS during the course of learning (cf. Renkl et al., 2002).

Zusammenfassung

Das Ziel der vorliegenden Dissertation war es, die Effektivität der instruktionalen Unterstützungsmaßnahme für multimediales Lernen, das Signaling von Text-Bild Korrespondenzen und deren Randbedingungen zu untersuchen. Das Lernen mit Multimedia erfordert die kognitive Verarbeitung von Informationen aus Text und korrespondierendem Bild. In zahlreichen Studien wurde gezeigt, dass das Lernen mit Text und Bild zu besseren Lernergebnissen führt als das Lernen lediglich mit Text (Mayer, 2014a). Der zentrale kognitive Verarbeitungsschritt beim Lernen mit Multimedia ist die Integration von verbaler und piktorialer Information in eine kohärente, integrierte mentale Repräsentation (Butcher, 2014; Mayer, 2009; Schnotz, 2014). Es wird angenommen, dass diese kohärente, integrierte Repräsentation die Voraussetzung für erfolgreiches Lernen mit Multimedia ist. Da jedoch festgestellt wurde, dass Lernende oftmals Schwierigkeiten bei der Integration von verbaler und piktorialer Information haben (z.B. Hannus & Hyönä, 1999; Renkl & Scheiter, 2015), wurden verschiedene instruktionale Unterstützungsmaßnahmen für das Lernen mit Multimedia entwickelt und getestet. Eine dieser empfohlenen Maßnahmen ist das Hervorheben von korrespondierenden Elementen in Text und Bild durch sogenannte Multimedia Integration Signals (MIS) (Van Gog, 2014). MIS können beispielsweise durch Farbkodierungen (Hervorhebung korrespondierender Elemente in Text und Bild in der gleichen Farbe) oder deiktische Hinweise (Text der auf das Bild referenziert, z.B.: “Im Bild kannst du sehen wie Element x…”) umgesetzt werden. Es wird angenommen, dass MIS Selektions- und Organisationsprozesse relevanter Informationen in Text und Bild fördern. Allerdings sollten MIS Lernende vor allem bei der Integration verbaler und piktorialer Informationen zu einer kohärenten, integrierten Repräsentation unterstützen.

Um die Effektivität und die Randbedingungen von MIS als instruktionale Unterstützungsmaßnahme für multimediales Lernen zu untersuchen, wurde eine Meta-Analyse durchgeführt (Studie 1). Die Ergebnisse zeigen, dass das domänenspezifische Vorwissen die Effektivität von MIS beeinflusst. MIS unterstützen lediglich Lernende mit geringem domänenspezifischen Vorwissen beim Verstehen der Inhalte, wohingegen die Verstehensleistung von Lernenden mit hohem Vorwissen nicht beeinflusst wird. Die Ergebnisse der Meta-Analyse stehen damit in Einklang mit der Annahme, dass die Effektivität instruktionaler Techniken von dem Vorwissen der Lernenden abhängig ist (Expertise Reversal Effekt, ERE; Kalyuga et al., 2003). Befunde aus dem Bereich der ERE-Forschung umfassen zum einen partielle EREs, die zeigen, dass die Lernleistung von Lernenden mit hohem Vorwissen von einer instruktionalen Maßnahme nicht beeinflusst wird (wie in dem Ergebnis der Meta-Analyse), wohingegen Lernende mit geringem Vorwissen bezüglich ihrer Lernleistung profitieren. Zum anderen umfassen die Befunde vollständige EREs, die zeigen, dass Lernende mit geringem Vorwissen ebenfalls von einer Unterstützungsmaßnahme profitieren, wohingegen allerdings Lernende mit hohem Vorwissen beim Lernen beeinträchtigt werden (Kalyuga, 2007).

Um zu untersuchen, inwieweit das Ergebnis der Meta-Analyse generalisiert werden kann, wurde eine ökologisch valide Feldstudie in Schulen durchgeführt (Studie 2). Im Gegensatz zu einem Großteil, der in die Meta-Analyse eingeschlossenen Studien, wurde in Studie 2 (a) umfangreiches ökologisch valides Lernmaterial genutzt, (b) es wurden Lernende mit geringem und hohem Vorwissen in die Stichprobe eingeschlossen, und (c) die Materialversion mit MIS (MIS+) wurde mit einer starken Kontrollgruppe (MIS-) verglichen. Die daraus resultierende Forschungsfrage war, welche Art von ERE bezüglich des Einsatzes von MIS in einer ökologisch validen Studie in Schulen resultieren würde (Studie 2). Darüber hinaus wurde angenommen, dass verschiedene Arten von kognitiver Belastung (extrinsisch, lernbezogen) zur Erklärung beitragen können, warum Lernende mit hohem Vorwissen nicht von MIS profitieren oder sogar in ihrem Lernen beeinträchtigt werden. Die Ergebnisse der Studie 2 zeigten einen vollständigen ERE stabil für alle Lernmaße: Lernende mit geringem Vorwissen profitierten beim Lernen mit einer E-Book Version, die Hervorhebungen im Text und MIS verwendete (MIS+), verglichen mit einer Version, die lediglich Hervorhebungen im Text und nur grundlegende MIS enthielt (MIS-). Im Gegensatz dazu zeigten Lernende mit hohem Vorwissen eine schlechtere Lernleistung, wenn sie mit der Version MIS+ lernten im Vergleich zu der Version MIS-. Zudem gab diese Gruppe an, mehr Schwierigkeiten beim Lernen mit der Version MIS+ gehabt zu haben und sie nahmen sich außerdem mehr Zeit für das Lernen, wenn sie mit der Version MIS+ lernten.

Um einen besseren Einblick in zugrundeliegende Prozesse des ERE bezogen auf multimediales Signaling zu erhalten, wurde eine laborexperimentelle Eye-Tracking Studie mit einer ähnlichen Stichprobe und einem Auszug des Lernmaterials mit der selben Manipulation (MIS+/MIS-) wie in Studie 2 durchgeführt (Studie 3). Die damit verbundene moderierte Mediationshypothese lautete, dass der Effekt von MIS auf die Verteilung visueller Aufmerksamkeit und die kognitive Belastung durch das Vorwissen der Lernenden beeinflusst werden würde. Darüber hinaus wurde angenommen, dass die Verteilung visueller Aufmerksamkeit und die kognitive Belastung die Lernergebnisse entweder direkt beeinflussen oder, dass dieser Effekt ebenfalls von dem Vorwissen der Lernenden beeinflusst werden würde. Diese Effekte, bezogen auf Prozessmaße und das Vorwissen, sollten wiederum einen potenziellen ERE bezogen auf die Effektivität von MIS erklären. Die Ergebnisse von Studie 3 ergaben einen partiellen ERE lediglich für die Erinnerungsleistung. Darüber hinaus beeinflussten MIS das Blickverhalten und den Pupillendurchmesser (als Maß für kognitive Belastung) von Lernenden mit geringem Vorwissen. Diese Gruppe fixierte Bilder länger, führte mehr Transitionen zwischen Text und Bild durch und neigte zu einem kleineren Pupillendurchmesser, wenn sie mit der Version MIS+ im Gegensatz zu der Version MIS- lernten. Lernende mit hohem Vorwissen hingegen wurden durch die MIS-Manipulation weder in ihrer visuellen Aufmerksamkeit noch in ihrem Pupillendurchmesser beeinflusst. Sie gaben jedoch an sich stärker konzentriert zu haben, wenn das Material MIS enthielt. Abweichend von den getroffenen Annahmen in den moderierten Mediationsmodellen, erklärten weder die visuelle Aufmerksamkeit noch die kognitive Belastung den partiellen ERE für MIS.

Zusammenfassend lässt sich sagen, dass die vorliegende Dissertation die Effektivität von multimedialem Signaling als instruktionale Unterstützungsmaßnahme beim Lernen mit Multimedia bestätigt, jedoch ausschließlich für Lernende mit geringem Vorwissen. Basierend auf diesen Ergebnissen kann angenommen werden, dass MIS den Integrationsprozess von verbalen und piktorialen Informationen in eine kohärente, integrierte Repräsentation für diese Gruppe von Lernenden unterstützen. In Einklang mit Befunden bezüglich des ERE, ergaben die vorliegenden Studien, dass Lernende mit hohem Vorwissen durch das Vorhandensein von MIS im Lernmaterial bezüglich ihres Lernergebnisses entweder nicht beeinflusst oder sogar in ihrem Lernen beeinträchtigt wurden. Das Blickverhalten und die kognitive Belastung erklärten den Einfluss von Vorwissen auf die Effektivität von MIS nicht.

Ein Grund zur Sorge ist das Auftreten eines vollständigen EREs unter ökologisch validen Bedingungen im Feld für alle erhobenen Lernmaße (Studie 2). Falls zukünftige Forschung dieses Ergebnis weiter stützen sollte, würde dies unter normativen Gesichtspunkten einen ethischen Konflikt bedeuten. Schulklassen sind bezüglich des vorhandenen Vorwissens zu bestimmten Themen heterogen (Slavin, 1987). Die Empfehlung von multimedialem Signaling als grundlegende Unterstützungsmaßnahme für multimediales Lernen kann folglich für Lernende mit hohem Vorwissen in einer Klasse Nachteile bezüglich ihres Lernerfolgs bedeuten. Diesem ethischen Konflikt kann jedoch individualisiertes Lernmaterial entgegengesetzt werden (vgl. Kalyuga, 2007). Der zunehmende Einsatz von digitalem Lernmaterial im Bildungsbereich bietet die Möglichkeit, Lernumgebungen adaptiv, basierend auf dem aktuellen Wissensstand eines Lernenden, zu gestalten. Dafür müssen allerdings zukünftig weitere Erkenntnisse bezüglich der zugrundeliegenden Prozesse von vollständigen EREs des multimedialen Signaling Effekts gewonnen werden, um in der Lage zu sein Instruktionen zu entwickeln, die Lernende mit hohem Vorwissen bestmöglich unterstützen. Diesbezüglich könnte zukünftige Forschung die Effektivität von MIS in verschiedenen Varianten untersuchen. Beispielsweise könnte die Salienz von MIS, die Anzahl der MIS oder die graduelle Ausblendung von MIS während des Lernprozesses (Renkl et al., 2002) Aufschluss über effektive Instruktionen für Lernende mit hohem Vorwissen liefern.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods, 2*, 315–328. doi: 10.1037/1082-989x.2.4.315

Allison, P. D. (1999). *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press.

*Arslan-Ari, I. (2013). *Examining the effects of cueing and prior knowledge on learning, mental effort, and study time in a complex animation*. Texas Tech University, Lubbock, USA.

Attar, N., Schneps, M. H., & Pomplun, M. (2016). Working memory load predicts visual search efficiency: Evidence from a novel pupillary response paradigm. *Memory & Cognition, 44,* 1038-1049. doi:10.3758/s13421-016-0617-8

Ausubel, D., Novak, J., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York: Holt, Rinehart & Winston.

Aydeniz, M., & Kotowski, E. L. (2012). What do middle and high school students know about the particulate nature of matter after instruction? Implications for practice. *School Science and Mathematics, 112*, 59–65. doi: 10.1111/j.1949-8594.2011.00120.x

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 206-226). New York, NY: Cambridge University Press.

*Baetge, I., & Seufert, T. (2010). Effects of support for coherence formation in computer science education. In M. Hopp, & F. Wagner (Eds.), *EARLI SIG 6/7 Meeting 2010. Instructional design for motivated and competent learning in a digital world* (pp. 143–145). Ulm: Ulm University.

Baumert, J., Lehmann, R., Lehrke, M., Clausen, C., Hosenfeld, I., Neubrand, J., … Günther, W. (Eds.). (1998). *Testaufgaben Naturwissenschaften TIMSS 7./8. Klasse* (Population 2). Materialien aus der Bildungsforschung Nr. 60. Berlin: Max-Planck-Institut für Bildungsforschung.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276–292. doi:10.1037/0033-2909.91.2.276

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101. doi:10.2307/2533446

Belsey, D. A., Kuh, E., & Welsch, R. E. (2004). *Regression diagnostics: Identifying influential data and sources of collinearity*. New Jersey: Wiley.

*Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*, 70–87. doi:10.1037/a0013247

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*, 325–341. doi:10.1016/j.learninstruc.2004.06.006

*Boucheix, J. M., & Guignard, H. (2005). What animated illustrations conditions can improve technical document comprehension in young students? Format, signaling and control of the presentation. *European Journal of Psychology of Education, 20*, 369–388. doi:10.1007/bf03173563

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

Boucheix, J. M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction, 19*, 112-127. doi:10.1016/j.learninstruc.2008.03.004

Bracht, G. H. (1970). Experimental factors related to aptitude-treatment interactions. *Review of Educational Research, 40,* 627-645. doi:10.2307/1169460

Bundesministerium für Bildung und Forschung. (2016). *Sprung nach vorn in der digitalen Bildung [A leap forward in digital education]* [Press release]. Retrieved from https://www.bmbf.de/de/sprung-nach-vorn-in-der-digitalen-bildung-3430.html

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 174–205). New York, NY: Cambridge University Press.

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

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 21–30). Cambridge: Cambridge University Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152. doi:10.1207/s15516709cog0502_2

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*, 315-324. doi:10.1016/j.chb.2008.12.020

Clark, W., & Luckin, R. (2013). *What the research says: iPads in the classroom*. London: London Knowledge Lab, Institute of Education.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

De Groot, S. G., & Gebhard, J. W. (1952). Pupil size as determined by adapting luminance. *Journal of the Optical Society of America, 42*, 492-495. doi:10.1364/josa.42.000492

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*, 105-134. doi:10.1007/s11251-009-9110-0

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

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

*De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010b). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education, 55*, 681–691. doi:10.1016/j.compedu.2010.02.027

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous and germane load. *Journal of Educational Psychology, 100*, 223–234. doi:10.1037/0022-0663.100.1.223

Del Re, A. C. (2014). compute.es: Compute effect sizes (Version R package version 0.2-4). Retrieved from http://cran.r-project.org/web/packages/compute.es

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245. doi:10.1037/0033-295x.102.2.211

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods, 10*, 444–467. doi:10.1037/1082-989x.10.4.444

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 665–694. doi:10.1348/000711010x502733

Fisher, R. A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society, 85*, 87–94. doi:10.2307/2340521

*Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction, 20*, 216–224. doi:10.1016/j.learninstruc.2009.02.021

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for life in a digital age: The IEA International Computer and Information Literacy Study International Report.* Cham: Springer. doi:10.1007/978-3-319-14222-7

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2*, 33–67. doi:10.1207/s15328031us0201_03

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*, 523–552. doi:10.1007/s10648-011-9174-7

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331. doi:10.1016/j.learninstruc.2005.07.001

Grüß-Niehaus, T. (2010) *Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht - der Effekt von Methoden progressiver und kollaborativer Reflexion [Comprehension of the dissolution concept in chemistry education].* Gottfried Wilhelm Leibniz Universität, Hannover, Germany.

Hager, W. (2002). The examination of psychological hypotheses by planned contrasts referring to two-factor interactions in fixed-effects ANOVA. *Methods of*

Psychological Research Online, 7, 49–77. Retrieved 14 October 2016, from http://www.dgps.de/fachgruppen/methoden/mpr-online/issue18/art3/Hager.pdf.

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied, 5*, 129–151. doi:10.1037/1076-898x.5.2.129

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. doi:10.1006/ceps.1998.0987

*Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414–434. doi:10.1037/0022-0663.90.3.414

Hartley, J. (2004). Designing instructional and informational text. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (2nd ed., pp. 917–947). Mahwah, NJ: Erlbaum.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504. doi:10.1037//1082-989x.3.4.486

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717–742. doi:10.1006/jmla.1993.1036

Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemical Education Research and Practice*, *12*, 184–192. doi: doi:10.1039/c1rp90023d

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190–1192. doi:10.1126/science.143.3611.1190

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hollstein, A. (2001) *Computergestütztes Lernen auf Basis konstruktivistischer Lerntheorien am Beispiel der Einführung in das Kugelteilchenmodell [Computer-supported learning on the basis of constructivist learning theories through the example of the introduction of the particle model].* University of Duisburg-Essen, Essen, Germany.

Homer, B. D., & Plass, J. L. (2009). Expertise reversal for iconic representations in science visualizations. *Instructional Science, 38*, 259–276. doi:10.1007/s11251-009-9108-7

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275–292. doi:10.1111/1468-2389.00156

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722-738. doi:10.1016/j.learninstruc.2007.09.013

IBM Corp. (2013). IBM SPSS Statistics for Mac (Version 22.0). Armonk, NY: IBM Corp.

Initiative D21. (2014). *Medienbildung an deutschen Schulen: Handlungsempfehlungen für die digitale Gesellschaft [Media education in German schools: Recommendations for action for the digital society]*. Berlin, Germany: atene KOM GmbH.

*Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior, 32*, 47–53. doi:10.1016/j.chb.2013.11.013

*Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and Instruction, 18*, 135–145. doi:10.1016/j.learninstruc.2007.01.011

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. doi:10.1016/j.learninstruc.2009.02.019

*Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329–345. doi:10.1080/0144341970170307

*Johnson, A. M., Butcher, K., Ozogul, G., & Reisslein, M. (2013). Learning from abstract and contextualized representations: The effect of verbal guidance. *Computers in Human Behavior, 29*, 2239–2247. doi:10.1016/j.chb.2013.05.002

*Johnson, A. M., Butcher, K. R., Ozogul, G., & Reisslein, M. (2014). Introductory circuit analysis learning from abstract and contextualized circuit representations: Effects of diagram labels. *IEEE Transactions on Education, 57,* 160–168. doi:10.1109/te.2013.2284258

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*, 178-191. doi: 10.1037/a0026923

*Johnson, A. M., Ozogul, G., Moreno, R., & Reisslein, M. (2013). Pedagogical agent signaling of multiple visual engineering representations: The case of the young female agent. *Journal of Engineering Education, 102*, 319–337. doi:10.1002/jee.20009

*Johnson, A. M., Ozogul, G., & Reisslein, M. (2014). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*, 97–115. doi:10.1111/jcal.12078

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Harvard University Press.

Judd, C. M. (2000). Everyday data analysis in social psychology: Comparisons of linear models. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 370-392). Cambridge, UK: Cambridge University Press.

Judd, C. M., Yzerbyt, V. Y., & Muller, D. (2014). Mediation and Moderation. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of Research Methods in Personality and Social Psychology* (2nd ed., pp. 653-676), New York, NY: Cambridge University Press. doi:10.1017/cbo9780511996481.030

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

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583–1585. doi:10.1126/science.154.3756.1583

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509-539. doi:10.1007/s10648-007-9054-3

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 576–597). New York, NY: Cambridge University Press.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychlogist, 38*, 23–31. doi:10.1207/s15326985ep3801_4

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. doi:10.1518/001872098779480587

*Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371. doi:10.1002/(sici)1099-0720(199908)13:4<351::aid-acp589>3.0.co;2-6

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209–215. doi:10.1007/s11251-009-9102-0

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10. doi:10.1016/s0959-4752(01)00014-7

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

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242*, 396–402. doi:10.1148/radiol.2422051997

*Kühl, T., Scheiter, K., & Gerjets, P. (2012). Enhancing learning from dynamic and static visualizations by means of cueing. *Journal of Educational Multimedia and Hypermedia, 21*, 77–88.

Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior, 27*, 29–35. doi:10.1016/j.chb.2010.05.008

Lemarié, J., Lorch, R. F. J., Eyrolle, H., & Virbel, J. (2008). SARA: A text-based and reader-based theory of signaling. *Educational Psychologist, 43*, 27–48. doi:10.1080/00461520701756321

*Lin, L. (2011). *Learning with multimedia: Are visual cues and self-explanation prompts effective?* Arizona State University, Phoenix, USA.

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology, 75*, 402–412. doi:10.1037/0022-0663.75.3.402

Lorch, R. F. (1989). Text signaling devices and their effects on reading and memory processes. *Educational Psychology Review, 1*, 209–234. doi: 10.1007/bf01320135

Lorch, R. F., & Lorch, E. P. (1995). Effects of organizational signals on text-processing strategies. *Journal of Educational Psychology, 87*, 537–544. doi:10.1037/0022-0663.87.4.537

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 227–246). New York, NY: Cambridge University Press.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104. doi:10.1037/1082-989x.7.1.83

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician, 36*, 158–160. doi:10.1080/00031305.1982.10482818

Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. *Learning and Instruction*, *11*, 305-329. doi:10.1016/s0959-4752(00)00035-9

*Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly, 48*, 199–214. doi:10.1002/rrq.41

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. doi:10.1016/j.compedu.2012.07.011

*Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*, 377–389. doi:10.1037/0022-0663.93.2.377

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19. doi:10.1207/s15326985ep3201_1

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*, 760–769. doi:10.1037/0003-066x.63.8.760

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Mayer, R. E. (2014a). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 1–24). New York, NY: Cambridge University Press.

Mayer, R. E. (2014b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397. doi:10.1037/0022-0663.93.2.390

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52. doi:10.1207/s15326985ep3801_6

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401. doi:10.1037/0022-0663.86.3.389

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 11*, 368–401. doi:10.1037/1082-989x.11.4.386

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1–43. doi:10.1207/s1532690xci1401_1

*McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology, 34*, 143–153. doi:10.1016/j.cedpsych.2008.12.003

Miles, J. (2005). Tolerance and variance inflation factor. In B. S. Everitt, & D. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 1, pp. 2055–2056). Hoboken, NJ: John Wiley and Sons. doi:10.1002/0470013192.bsa683

*Moreno, R., Reisslein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention during learning. *Journal of Media Psychology: Theories, Methods, and Applications, 22*, 52–60. doi:10.1027/1864-1105/a000008

Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education, 79*, 739-744. doi:10.1021/ed079p739

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863. doi:10.1037/0022-3514.89.6.852

Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education, 69,* 191-196. doi: 10.1021/ed069p191

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics, 2*, 149–155. doi:10.3758/bf03210310

Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*, 237–258. doi:10.1007/s11251-009-9106-9

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: An examination of fixation patterns in a science simulation. *Computers in Human Behavior, 35*, 234–242. doi:10.1016/j.chb.2014.02.040

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science, 38*, 217–236. doi:10.1007/s11251-009-9109-6

*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. doi:10.1016/j.chb.2009.09.001

*Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*, 445–453. doi:10.1016/j.compedu.2009.03.002

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

Paas, F., Tuovinen, J. E., Tabbers, H. K., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. doi:10.1207/s15326985ep3801_8

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115-121. doi: 10.1007/s10648-010-9133-8

*Paik, E. S., & Schraw, G. (2013). Learning with animation and illusions of understanding. *Journal of Educational Psychology, 105*, 278–289. doi:10.1037/a0030281

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255-287. doi:10.1037/h0084295

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, 50*, 157–175. doi:10.1080/14786440009463897

Petermann, K., Friedrich, J., & Oetken, M. (2009). Test zur Diagnose von Schülervorstellungen zum Teilchenkonzept [Test for the diagnosis of students' conceptions related to the particle concept]. *Praxis der Naturwissenschaften - Chemie in der Schule*, *58*, 41-43.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology, 60*, 211–229. doi:10.1080/17470210600673818

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185–227. doi:10.1080/00273170701341316

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*, 1457–1506. doi:10.1080/17470210902816461

R Development Core Team. (2008). R: A language and environment for statistical computing [http://www.R-project.org]. Vienna, Austria: R Foundation for Statistical Computing.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*, 529–556. doi:10.1016/s0959-4752(01)00030-5

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education, 70*, 293–315. doi:10.1080/00220970209599510

Renkl, A., & Scheiter, K. (2015). Studying visual displays: How to instructionally support learning. *Educational Psychology Review,* 1-23. doi:10.1007/s10648-015-9340-4

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19-36. doi: 10.1016/j.edurev.2015.12.003

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641. doi:10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park: Sage.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast Analysis.* Cambridge, UK: Cambridge University Press.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*, 328–364. doi:10.3102/0034654313500826

Schanze, S. (2002). *Wissenserwerb mithilfe der internetbasierten Lernumgebung ChemNet: Eine empirische Untersuchung zum Lernen mit linearen und vernetzten Hypertexten [Knowledge acquisition by means of the online* *learning environment ChemNet]*. Christian-Albrechts-Universität zu Kiel, Kiel, Germany.

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*, 1-24. doi:10.1371/journal.pone.0130608

*Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction, 36*, 11–26. doi:10.1016/j.learninstruc.2014.11.002

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, 23*, 1209–1214. doi:10.1002/acp.1524

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2014). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science, 43,* 93-114. doi:10.1007/s11251-014-9328-3

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction, 20*, 100-110. doi:10.1016/j.learninstruc.2009.02.011

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007) *Lesegeschwindigkeits- und verständnistest für die Klassenstufen 6-12 (LGVT 6-12) [Reading speed and comprehension test for grades 6-12].* Göttingen: Hogrefe.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66. doi:10.1037/0033-295x.84.1.1

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science, 38*, 315–323. doi:10.1007/s11251-009-9104-y

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 72-103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141–156. doi:10.1016/s0959-4752(02)00017-8

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508. doi:10.1007/s10648-007-9053-4

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology, 106*, 974–989. doi:10.1037/a0037054

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. doi:10.1002/acp.1526

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258-266. doi: 10.1016/j.chb.2008.12.011

*Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*, 227–237. doi:10.1016/s0959-4752(02)00022-1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127–190. doi:10.1037/0033-295x.84.2.127

Shute, V. J., & Gluck, K. A. (1996). Individual differences in patterns of spontaneous online tool use. *Journal of the Learning Sciences, 5*, 329–355. doi:10.1207/s15327809jls0504_2

Slavin, R. E. (1987) A theory of school and classroom organization. *Educational Psychologist*, *22*, 89-108. doi: 10.1207/s15326985ep2202_1

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi:10.1023/a:1022193728205

Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science, 38*, 441-453. doi:10.1007/s11251-009-9119-4

*Tabbers, H. K., Martens, R. L., & Merriënboer, J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81. doi:10.1348/000709904322848824

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167–174. doi:10.1111/j.1469-8986.2003.00148.x

Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 263–278). New York, NY: Cambridge University Press.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26. doi:10.1080/00461520701756248

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99. doi:10.1016/j.learninstruc.2009.02.009

Van Gog, T., van Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*, 325–331. doi:10.1016/j.chb.2008.12.021

Van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction, 24*, 15–25. doi:10.1016/j.learninstruc.2012.08.005

*Van Oostendorp, H., Beijersbergen, M. J., & Solaimani, S. (2008). Conditions for learning from animations. *Proceedings of the 8th International Conference for the Learning Sciences: Vol. 2.* (pp. 438–445). Utrecht, The Netherlands: International Society of the Learning Sciences.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. doi:10.1037/1082-989x.10.4.428

Vygotski, L. S. (1963). Learning and mental development at school age. In B. Simon & J. Simon (Eds.), *Educational psychology in the U.S.S.R.* (pp. 21–34). London, UK: Routledge & Kegan Paul.

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009) Überprüfung einer Kurzskala intrinsischer Motivation (KIM) [Testing a short scale of intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 31-45.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. doi:10.1037/0003-066x.54.8.594

Yezierski, E. J., & Birk, J. P. (2006). Misconceptions about the particulate nature of matter. Using animations to close the gender gap*. Journal of Chemical Education, 83*, 954-960. doi: 10.1021/ed083p954

ZPG-Chemie (2011, Febuary). Das Teilchenmodell in Klasse 8 [The Particle Model of Matter in grade 8]. Retrieved Feburary 28, 2016, from Landesakademie für Fortbildung und Personalentwicklung an Schulen website, http://lehrerfortbildung-bw.de/faecher/chemie/gym/fb2/modul2/

List of Figures

List of Tables

Appendices

Appendix A

Verification items assessing domain-specific prior knowledge in Study 2.

Kreuze die richtigen Aussagen an. Es können mehrere Aussagen richtig sein.

	Aussagen	**Richtig**
a)	Die Bewegung der kleinsten Teilchen kommt nie zum Stillstand.	☐
b)	Zwischen den Teilchen, die einen Stoff bilden, ist Luft.	☐
c)	Die Bewegung der kleinsten Teilchen eines Gases wird mit der Zeit immer langsamer.*	☐
d)	Die einzelnen Schwefelteilchen sind gelb.	☐
e)	Kleinste Teilchen können nicht schmelzen.	☐
f)	Wenn kleine Teilchen eines Gases gegen ein Hindernis treffen, zerbrechen sie.	☐
g)	Wenn die kleinsten Teilchen eines Gases erwärmt werden, vergrößern sie sich.	☐
h)	Beim Lösen von Salz in Wasser verschwinden die Salzteilchen.	☐
i)	Wenn eine Flüssigkeit verdunstet, dann löst sie sich in nichts auf.	☐
j)	Die kleinsten Teilchen eines Gases können sich auflösen.	☐
k)	Zwischen den Teilchen, die einen Stoff bilden, ist nichts.	☐
l)	Je schneller sich die Teilchen bewegen, desto größer ist der Druck.*	☐
m)	Zwischen den einzelnen Wasserteilchen befindet sich Wasser in flüssiger Form.	☐
n)	Verkleinert man das Volumen bei gleichbleibender Teilchenanzahl und Energie, erhöht sich der Druck.	☐
o)	Es kann beliebig viel Salz in 100 ml Wasser aufgelöst werden.	☐

* Items were removed due to negative corrected item-total correlations.

Appendix B

Multiple-choice items assessing domain-specific prior knowledge in Study 3.

Bei den folgenden Fragen gibt es jeweils immer nur eine richtige Antwort, kreuze die richtige Antwort an!

1) Die kleinsten Teilchen von Stoffen…

… sind nur in Feststoffen nachweisbar.	□
… sind im gasförmigen Zustand verschwunden.	□
… verschwinden, wenn sie zu stark erhitzt werden.	□
… verschwinden, wenn die Temperatur unter 0°C sinkt.	□
… verschwinden nicht in gasförmigen Stoffen.	□

2) Welche Aussage ist korrekt? Zwischen den Teilchen, die einen Stoff bilden, …

… ist nichts.	□
… ist Wasser, wenn es sich um einen flüssigen Stoff handelt.	□
… ist Luft.	□
… ist Wasserdampf, wenn es sich um einen gasförmigen Stoff handelt.	□
… sind Staub und Schadstoffe.	□

3) Max fragt sich, ob er einen Lufterfrischer schneller in einem warmen oder in einem kalten Raum riechen kann. Er entscheidet sich ein Experiment durchzuführen: Er kühlt den Raum auf 10°C ab, schließt den Lufterfrischer an und misst die Zeit, bis der Duft des Lufterfrischers die Tür erreicht. Am nächsten Tag erwärmt er den gleichen Raum auf 30°C, schließt einen neuen Lufterfrischer an und misst erneut die Zeit, bis der Duft die Tür erreicht. Welche Vermutung hast Du über das Ergebnis des Experimentes? Kreuze eine der folgenden Möglichkeiten an: Der Duft erreicht die Tür…

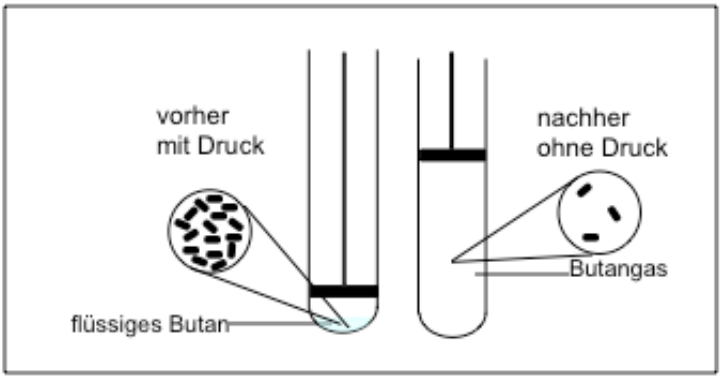
… bei beiden Temperaturen in der gleichen Zeit.	□
… bei 30°C langsamer, weil die Teilchen größer werden.	□
… bei 10°C schneller, weil sich die Teilchen schneller bewegen.	□
… bei 30°C schneller, weil sich die Teilchen schneller bewegen.	□
… bei 10°C langsamer, weil die Teilchen kleiner werden.	□

4) Stelle Dir vor, man könnte die Bewegungen der kleinsten Teilchen eines Gases sehen. Was denkst Du, welche Aussage korrekt ist? Die Bewegung der kleinsten Teilchen eines Gases…

… kommt bei gleichbleibender Temperatur nicht zum Stillstand.	☐
… wird bei gleichbleibender Temperatur aufgrund der Reibungsverluste langsamer.	☐
… verliert (unabhängig von der Temperatur) an Geschwindigkeit, da die Teilchen gegen Luftteilchen stoßen.	☐
… nimmt (unabhängig von der Temperatur) an Geschwindigkeit zu, da die Teilchen gegen Luftteilchen stoßen.	☐
… wird bei gleichbleibender Temperatur aufgrund der Erdanziehungskraft schneller.	☐

5) In einem Versuch wird die Hülse einer Handpumpe vollständig mit Butangas gefüllt. Das Gas wird mit der Handpumpe unter Druck in flüssiges Butan verwandelt. Wird der Kolben der Pumpe gelöst, so verdampft das flüssige Butan wieder.

In der Abbildung ist das Verdampfen von Butan im Teilchenmodell dargestellt.



Wie stellst Du Dir den Raum zwischen den Teilchen im Butangas vor? Kreuze bitte an. Ich stelle mir vor, dass…

… zwischen den Teilchen auch Butangas vorhanden ist.	☐
… Wasserteilchen zwischen den Butanteilchen sind.	☐
… sich Luft zwischen den Teilchen befindet.	☐
… ein unsichtbarer Stoff zwischen den Teilchen ist.	☐
… der Raum zwischen den Teilchen leer ist.	☐

6) Wie kann man mithilfe des Teilchenmodells erklären, dass in einem geschlossenen Gefäß Druck entsteht, wenn man das darin befindliche Gas erhitzt?

Die Teilchen eines Gases sind normalerweise klein und mit steigender Temperatur beginnen sie, sich auszudehnen, was sich durch Druck	☐

bemerkbar macht.	
Die Teilchen eines Gases sind normalerweise fest miteinander verbunden. Mit steigender Temperatur beginnen sie, sich zu bewegen, was sich durch Druck bemerkbar macht.	☐
Die Teilchen eines Gases sind ständig in Bewegung. Mit steigender Temperatur erhöht sich deren Geschwindigkeit, was sich durch Druck bemerkbar macht.	☐
Die Teilchen eines Gases sind fest miteinander verbunden. Mit steigender Temperatur beginnen sie, sich zu lösen, was sich durch Druck bemerkbar macht.	☐
Die Teilchen eines Gases sind ständig in Bewegung. Mit steigender Temperatur beginnen sie, sich miteinander zu verbinden, was sich durch Druck bemerkbar macht.	☐

7) Bei 0°C gilt für die Teilchen von festen Stoffen:

Die Teilchen bewegen sich ungeordnet.	☐
Die Anziehungskräfte sind überwunden.	☐
Es liegen überhaupt keine Anziehungskräfte vor.	☐
Die Teilchen schwingen um ihre Plätze, weil Anziehungskräfte vorliegen.	☐
Die Teilchen stehen still, weil Anziehungskräfte vorliegen.	☐

8) Welche Aussage ist korrekt? Es gibt drei Aggregatzustände. Diese besagen, dass…

… derselbe Stoff in unterschiedlichen Zuständen (fest, flüssig und gasförmig) vorkommen kann.	☐
… es drei verschiedene Arten an Stoffen gibt, nämlich: fest, flüssig und gasförmig.	☐
… derselbe Stoff in unterschiedlichen Zuständen (fest, flüssig und gasförmig) vorkommen kann, wobei der flüssige Zustand immer den Übergang zwischen „fest“ und „gasförmig“ darstellt.	☐
… die in den Stoffen enthaltenen Teilchen entweder fest, flüssig oder gasförmig sind.	☐
… derselbe Stoff nicht in unterschiedlichen Zuständen (fest, flüssig und gasförmig) vorkommen kann.	☐

9) In gasförmigen Stoffen…

… bewegen sich die Teilchen geordnet.	☐
… sind die Anziehungskräfte nahezu überwunden und die Teilchen berühren sich selten.	☐
… liegen überhaupt keine Anziehungskräfte vor.	☐
…. sind die Abstände zwischen den Teilchen gleich groß.	☐
… sind die Abstände zwischen Teilchen klein, weil Anziehungskräfte vorliegen.	☐

10) In flüssigen Stoffen…

… sind die Abstände zwischen den Teilchen gleich groß.	☐
… sind die Anziehungskräfte überwunden und die Teilchen berühren sich nicht.	☐
… liegen überhaupt keine Anziehungskräfte vor.	☐
… sind die Teilchen geordnet, weil Anziehungskräfte vorliegen.	☐
… bewegen sich die Teilchen ungeordnet.	☐

11) Die kleinsten Teilchen von Stoffen…

… verändern ihre Temperatur, je nachdem, welche Außentemperatur herrscht.	☐
… sind wärmer, je näher sie zusammen sind.	☐
… ziehen sich bei Abkühlung zusammen, um Wärme zu speichern.	☐
… können nicht über Temperatur beschrieben werden, da Temperatur eine Stoffeigenschaft ist.	☐
… sind genauso warm, wie der Stoff, der aus ihnen besteht.	☐

12) Ein Topf mit Wasser wird auf eine heiße Herdplatte gestellt und fängt schnell an zu kochen. Ein Glasdeckel wird auf den Topf gelegt und es beginnen sich Wassertropfen im Inneren des Topfes am Glasdeckel zu bilden. Was ist passiert?

Wasserdampf ist sublimiert, die Teilchen berühren sich und Anziehungskräfte wirken.	☐
Der Dampf reagiert mit der Luft, dadurch entstehen Wassertropfen.	☐
Wasserdampf ist kondensiert, die Teilchen berühren sich und Anziehungskräfte wirken.	☐
Wasserdampf ist kondensiert und Anziehungskräfte sind überwunden.	☐
Wasserstoff und Sauerstoff haben zu Wasser reagiert.	☐

13) Wenn 24°C warmes Wasser auf 0°C abgekühlt wird, dann bedeutet das für die Teilchen, dass sie…

… weniger organisiert vorliegen.	☐
… auseinanderbrechen.	☐
… um ihre Plätze schwingen.	☐
… sich frei im Raum bewegen.	☐
… nicht mehr um ihre Plätze schwingen.	☐

14) Welcher Vergleich zwischen einem Feststoff und einem Gas ist korrekt? Im Vergleich zu einem Feststoff sind die…

… Teilchen in einem Gas dichter angeordnet und schwingen um ihre Plätze.	□
… Anziehungskräfte zwischen den Teilchen in einem Gas geringer.	□
… Teilchen in einem Gas leichter.	□
… Teilchen in einem Gas größer und schwingen um ihre Plätze.	□
… Teilchen in einem Gas dichter angeordnet und bewegen sich nicht.	□

15) Wie kann man den Druck in einem geschlossenen Behälter mit Gas erhöhen?

Volumenvergrößerung	□
Anzahl der Teilchen verringern	□
Energie verringern	□
Volumenverringerung	□
Ein anderes Gas unter gleichen Bedingungen verwenden.	□