Facilitating and Inhibiting Learning by the Spatial Contiguity of Text and Graphic: How Does Cognitive Load Mediate the Split-Attention and Expertise Reversal Effect?

Der Einfluss räumlicher Kontiguität zwischen Text und Bild auf den Wissenserwerb: Wie vermittelt kognitive Beanspruchung den Split-Attention und Expertise-Reversal Effekt?

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

vorgelegt von
Dipl.-Psych. Gabriele Cierniak
aus Pforzheim

Tübingen
2011

Tag der mündlichen Qualifikation	26.07.2011
Dekan	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter	Prof. Dr. Peter Gerjets
2. Berichterstatter	Prof. Dr. Stephan Schwan

THOUGHTS ON SCIENCE

As it is impossible to classify Psychology as a truly hard science I want to frame my dissertation with thoughts on science of two famous hard science researchers:

Science also teaches you to avoid numbers when they would make no sense. One can certainly assess scientific performance, students' satisfaction, success in teaching, and sometimes even originality, but no true scientist would do so by numbers…Giving a number to something that cannot be accurately quantified is bad science.

(Gottfried Schatz, biochemist and international laureate, 2005)

Science is the art of acquiring knowledge in such a manner that coherent structures of understanding can be erected on the basis of a critical evaluation of evidence.

(Ragnar A. Granit, neurophysiologist and Nobel laureate, 1977)

According to Schatz my dissertation presents rather "*bad science"* because it assigns numbers to fuzzy concepts like subjective cognitive load and learning success. However, according to Granit it is a *"scientific piece of art"* that contributes to our understanding of the relation between instructional design and knowledge acquisition.

For Diana

When I seek you,
I find you in brave and in fierce,
in demanding and in entertaining qualities.

When I've found you,
I embrace you gently.

Then, the feeling of completeness sounds in me.

Für Diana

Wenn ich nach dir suche,
finde ich dich in Mut und Kampfesgeist,
in Forderung und Unterhaltsamkeit.

Wenn ich dich gefunden habe,
umarme ich dich in Liebe und Sanftmut.

Dann erklingt das Gefühl der Vollständigkeit in mir.

ACKNOWLEDGMENT

This dissertation benefited a lot from the research field in which it was generated.

My THANKS go to my supervisors Prof. Dr. Peter Gerjets and Prof. Dr. Katharina Scheiter for accompanying and supporting the whole process of my dissertation. I also want to thank Prof. Dr. Stephan Schwan for his interest and engagement.

Moreover, I thank Prof. Dr. Dr. Friedrich Hesse and all other professors of the virtual PhD-Program (vgk) "Knowledge Acquisition and Knowledge Exchange with New Media" sponsored by the German Research Association (DFG). The program as well as the KMRC (Knowledge Media Research Center, Tuebingen) offered me a wonderful research platform with many opportunities.

I say THANK YOU to all my former “vgk-fellows” and PhD students at the KMRC – especially Kristin K., Jessica D., Eva M., Antonia B., Daniel W., Eva Sch., Christina Sch., Birgit I., and Maria O.!

I am happy that I had the opportunity to meet unique researchers of two different theoretical approaches relevant in my dissertation who shared their ideas with me.

Special thanks go to Prof. Paul Ayres not only for interesting discussions but also for all the motivational support: THANKS a lot ☺!

Furthermore, I cordially thank Prof. Danielle McNamara for informative discussions and challenging questions.

During writing this dissertation lots of things happened outside the research world in my closest surrounding which influenced me and thereby the final face of this work.

I say THANK YOU to my Mum and my brother Martin for being a stable basis.

I am very grateful to my sisters Diana and Sandra who gave life to two wonderful nieces and an amazing nephew.

I am especially thankful that my brother Richard returned home safely from the war in Afghanistan.

Moreover, I thank Gülestan F., Sandra V., and Marko T. for being my close friends and for discussing anything important in life - related but also not related to my PhD project. I say also thank you to my Gestalt friends who have accompanied me to the way following the PhD phase.

I dedicate this dissertation to my beloved sister Diana who vitally reinforced me to start this work but who died after suffering before I finalized it.

CONTENTS DIRECTORY

FIGURES DIRECTORY

TABLES DIRECTORY

Zusammenfassung

Allgemeiner Hintergrund. Um Effekte des Instruktionsdesigns auf den Lernerfolg zu erklären, beziehen sich derzeit viele Forscher auf die kognitive Beanspruchung als vermittelnder Faktor. Hierbei werden drei Arten der kognitiven Beanspruchung, die auf der Cognitive Load Theory (CLT; Sweller, Van Merriënboer, & Paas, 1998) basieren, unterschieden: (1) notwendige Beanspruchungen bedingt durch die Inhaltskomplexität (intrinsic cognitive load), (2) lern-hemmende Beanspruchungen bedingt durch unvorteilhaftes Instruktionsdesign (extraneous cognitive load) sowie (3) lern-förderliche Beanspruchungen (germane cognitive load) bedingt durch gelungenes Instruktionsdesign. Mit ihrem starken Fokus auf die Reduktion der lern-hemmenden Beanspruchung verführt die Cognitive Load Theory oft zu der Annahme, dass Lernen erst dann erfolgreich stattfindet, wenn es mühelos von Statten geht.

Hintergrund 1. Das Phänomen beim Multimedialernen, dass Lernende mit geringem Vorwissen anhand integrierter Instruktionsformate, bei denen Text und dazugehörige Illustrationen räumlich benachbart dargestellt sind, mehr lernen als anhand geteilter Instruktionsformate, bei denen Text und dazugehörige Illustrationen weiter voneinander entfernt dargestellt sind, bezeichnet man als Split-Attention Effekt. Die meisten Instruktionsforscher nehmen an, dass Lernende bei geteiltem Format durch lern-hemmende Beanspruchungen kognitiv überfordert sind. Einige Forscher gehen mittlerweile jedoch davon aus, dass Lernende außerdem durch ein integriertes Format zu lern-förderlichen Verarbeitungsprozessen angeregt werden. Eine Analyse der bisherigen Literatur ergab, dass kaum empirische Befunde vorliegen, die geeignet sind, eine der beiden Erklärungen des Split-Attention Effekts zu untermauern.

Ziel 1. Ein Anliegen dieser Arbeit war zu untersuchen, welcher der angenommenen kognitiven Mechanismen dem Split-Attention Effekt zugrunde liegt, um z.B. Lehrern ein angemessenes Verständnis von Lernmechanismen vermitteln zu können. Zu diesem Zweck wurden in einem ersten Schritt multiple Skalen entwickelt, welche die drei Beanspruchungsarten individuell erfassen sollten.

Experiment 1. Teilnehmer der ersten Studie waren 103 Studierende (Nicht-Medizinstudierende) der Universität Tübingen, welche die physiologischen Vorgänge in der Niere entweder anhand eines integrierten oder geteilten Instruktionsformats zu lernen hatten. Die kognitive Gesamtbeanspruchung wurde bei Zweidritteln der TeilnehmerInnen während des Lernens anhand der Leistung in einer von zwei Zweitaufgaben gemessen, die sich in den verwendeten Reizen unterschieden. Dies resultierte in einem 2x3 Versuchsdesign. Die drei Beanspruchungsarten wurden anhand multipler subjektiver Ratingskalen direkt im Anschluss an die Lernphase erfasst, wobei zur Erfassung der notwendigen Beanspruchung nach der Inhaltsschwierigkeit, zur Erfassung der lern-hemmenden Beanspruchung nach der Materialschwierigkeit und zur Erfassung der lern-förderlichen Beanspruchung nach der Konzentrationsstärke gefragt wurde. Abschließend hatten die TeilnehmerInnen vier Wissenstests zu beantworten (Fachbegriffe, Bildbeschriftungen, komplexes Faktenwissen und Transferwissen).

Ergebnisse 1. Ein Split-Attention Effekt zeigte sich bei drei der vier Wissenstests (Fachbegriffe, Bildbeschriftungen und komplexes Faktenwissen). Die Leistungen in der Zweitaufgabe wurden nicht vom Instruktionsformat beeinflusst. Multiple Mediationsanalysen ergaben jedoch, dass die subjektive Einschätzung der Materialschwierigkeit ca. 8% und die der Konzentrationsstärke ca. 9% des Split-Attention Effekts auf das Bildbeschriftungswissen vermittelt haben. Weitere Mediationsanalysen ergaben, dass die Einschätzung der Materialschwierigkeit ca. 16% und die der Konzentrationsstärke ca. 18% des Effekts auf komplexes Faktenwissen vermittelt haben.

Fazit 1. Die Ergebnisse unterstützen die Annahme, dass geteilte Formate nicht nur lern-hemmende Wirkung haben, sondern dass integrierte Formate zusätzlich lern-förderliche Prozesse anregen.

Hintergrund 2. Entsprechend des Split-Attention Effekts wird das integrierte Instruktionsformat beim Gestalten multimedialer Lernmaterialien empfohlen. Diese Empfehlung sollte jedoch nur bei Lernenden mit geringem Vorwissen vorbehaltlos umgesetzt werden, da sich gezeigt hat, dass Lernende mit hohem Vorwissen nicht von integrierten Formaten profitieren oder womöglich sogar schlechter damit abschneiden. Diese Umkehrung des Split-Attention Effekts bezeichnet man als Expertise-Reversal Effekt. Eine etablierte Erklärung des Effekts lautet, dass Lernende mit hohem Vorwissen bei geteiltem Format ihr Vorwissen aktiv einsetzen können und dadurch eine erhöhte lern-förderliche Beanspruchung erfahren, wohingegen sie durch ein integriertes Format an dieser lern-förderlichen Beanspruchung gehindert werden. Forscher der Cognitive Load Theory zweifeln diese Erklärung allerdings an, weil sie der Ansicht sind, dass Lernende mit hohem Vorwissen bei integriertem Format vor allem unter lern-hemmender Beanspruchung leiden. Eine Analyse der bisherigen Literatur ergab, dass kaum empirische Befunde vorliegen, die es erlauben, eine der Erklärungen des Expertise-Reversal Effekts zu untermauern oder zu widerlegen.

Ziel 2. Ein weiteres Anliegen dieser Arbeit richtet sich darauf, welcher kognitive Mechanismus dem Expertise-Reversal Effekt zugrunde liegt. Um dieses Ziel zu erreichen wurden neben den subjektiven Ratingskalen auch die Analyse des Blickverhaltens der Lernenden eingesetzt, um weitergehende Aufschlüsse über kognitive Verarbeitungsprozesse beim Lernen mit Multimedia zu erhalten.

Experiment 2. An der zweiten Studie nahmen 60 Studierende der Universität Tübingen teil. Während die eine Hälfte aus Medizinstudierenden bestand, die ein hohes Vorwissen aufwiesen, bestand die andere Hälfte aus Nicht-Medizinstudierenden mit geringem Vorwissen. Die TeilnehmerInnen hatten die physiologischen Vorgänge in der Niere entweder anhand eines integrierten oder anhand eines geteilten Instruktionsformats zu lernen. Dies resultierte in einem 2x2 Versuchsdesign. Die drei kognitiven Beanspruchungsarten wurden wieder direkt nach dem Lernen anhand der o.g. multiplen subjektiven Ratingskalen erfasst. Während dem Lernen wurden die Blickbewegungen der TeilnehmerInnen erfasst. Abschließend hatten die TeilnehmerInnen die vier Wissenstests zu beantworten (Fachbegriffe, Bildbeschriftungen, komplexes Faktenwissen und Transferwissen).

Ergebnisse 2. Der Expertise-Reversal Effekt zeigte sich bei zwei der vier Wissenstests (Bildbeschriftungen und komplexes Faktenwissen) – während TeilnehmerInnen mit geringem Vorwissen und integriertem Format einen höheren Lernerfolg hatten als solche mit geteiltem Format, zeigten sich für die TeilnehmerInnen mit hohem Vorwissen keine Unterschiede zwischen den Instruktionsbedingungen. Mediierte Moderationsanalysen ergaben, dass die lern-förderliche Beanspruchung diesen Interaktionseffekt auf das Bildbeschriftungswissen zu ca. 19% und auf komplexes Faktenwissen zu ca. 38% mediierte. Die Blickbewegungsdaten zeigten, dass TeilnehmerInnen mit geteiltem Format Text und Bild eher unabhängig voneinander verarbeiten, wohingegen TeilnehmerInnen mit integriertem Format häufiger zwischen korrespondierenden Textstellen und Bildausschnitten hin und her springen.

Fazit 2. Die Ergebnisse sprechen dafür, dass der Expertise-Reversal Effekt durch eine Erhöhung der lern-förderlichen Beanspruchung bei Lernenden mit hohem Vorwissen und geteiltem Format verursacht wird.

Gesamtfazit. Die Messung der drei Belastungsarten mittels multipler subjektiver Ratingskalen wird als relativ erfolgreich eingeschätzt. Allerdings bleibt die Messung der drei Belastungsarten weiterhin eine methodische Herausforderung, die weiterer Validierungen durch Prozessmaße wie Blickbewegungen oder durch neuronale Korrelate (z.B. durch EEG) bedarf. Zukünftige Forschungen, die Beanspruchungsmessungen mit Blickbewegungsdaten und neuronalen Korrelaten kombinieren, könnten einerseits Hinweise für geeignete Verarbeitungsstrategien liefern, um darauf aufbauend entsprechende Lerntrainings zu entwickeln, andererseits könnten sie genutzt werden, um die Entwicklung computerbasierter Lernumgebungen, die sich automatisch an die kognitive Beanspruchung von Lernenden adaptieren, zu ermöglichen.

Abstract

Overall Background. Many instructional design researchers refer to the triarchic model of cognitive load that is based on Cognitive Load Theory (CLT; Sweller, Van Merriënboer, & Paas, 1998) to describe how instructions influence necessary (*intrinsic* cognitive load caused by content difficulty), inhibiting (*extraneous* cognitive load caused by poor design), and supportive (*germane* cognitive load caused by good design) cognitive processes, which in turn, influence learning outcomes. Instructional design effects within multimedia research are predominantly explained by these three types of cognitive load. With its strong emphasis on reducing extraneous cognitive load CLT tempts to make instructors believe that learning takes place easily if cognitive load is reduced.

Background 1. The instructional design phenomenon that low-knowledge learners benefit from integrated formats where text is spatially integrated into the corresponding part of a graphic but suffer from separated formats where text and graphic are not presented close to each other is called *split-attention effect*. In explaining the split-attention effect many CLT researchers argue that low-knowledge learners with separated format suffer from high extraneous cognitive load only, whereas some researchers argue that low-knowledge learners with integrated format benefit additionally from germane cognitive load. A thorough literature review revealed that there is only very limited evidence in favor of or against the two cognitive load explanations of the split-attention effect.

Aims 1. To gain basic knowledge about cognitive learning mechanisms involved in the split-attention effect and to be able to inform instructors about these mechanisms of multimedia learning appropriately, the aim of this dissertation were twofold. The first aim was to generate scales to measure intrinsic, extraneous, and germane cognitive load separately. A second aim was to test the different cognitive load explanations of the split attention effect as comprehensively as possible.

Experiment 1. To investigate the cognitive load mechanisms underlying the split-attention effect, 103 (non-medical) students from the University of Tuebingen served as participants. They learned with a computer based learning environment either designed in integrated or separated format about the physiological functioning of the kidney. Overall cognitive load was measured by secondary task performance, whereas intrinsic, extraneous, and germane cognitive load were measured by multiple subjective rating scales. During learning about two thirds of the participants had to perform one out of two secondary tasks (reacting to one of two perceptual stimuli),

whereas one third had not to respond to a secondary task resulting in a 2x3 design. After learning, participants first had to rate (1) the difficulty of the content to be learned (intrinsic cognitive load), (2) the difficulty to learn with the materials (extraneous cognitive load), and (3) their level of concentration (germane cognitive load). Finally, they completed four knowledge tests (terms, labeling, complex facts, and transfer).

Results 1. The split-attention effect was demonstrated on three out of four knowledge tests (terms, labels, and complex facts). Whereas instructional format did not influence secondary task performance, mediation analyses yielded that extraneous cognitive load mediated about 8% of the split-attention effect on knowledge about labeling and about 16% of the effect on knowledge about complex facts. Furthermore, mediation analyses yielded that germane cognitive load mediated about 9% of the effect on knowledge about labeling and about 18% of the effect on knowledge about complex facts.

Conclusion 1. The results corroborate the assumption that the split-attention effect is not only caused by a reduction in extraneous but also by an increase in germane cognitive load.

Background 2. Based on the split-attention effect, multimedia researchers usually recommend integrated formats. However, this recommendation does not hold for high-knowledge learners, because it can be demonstrated that high-knowledge learners do not benefit or even suffer from integrated formats. This phenomenon is called *expertise reversal effect*. In explaining the expertise reversal effect several researchers argue that high-knowledge learners with integrated format are inhibited to invest enough germane cognitive load, whereas other researchers argue that high-knowledge learners with integrated format suffer from too high extraneous cognitive load. A thorough literature review revealed that there is only very limited evidence in favor of or against the different cognitive load explanations of the expertise reversal effect.

Aims 2. To gain knowledge about the cognitive load mechanism underlying the expertise reversal effect, the different explanations were tested. Moreover, participants' viewing behavior was registered to obtain deeper insights into their behavioral (and cognitive) processing.

Experiment 2. To investigate the cognitive load mechanisms underlying the expertise reversal effect, 60 students from the University of Tuebingen served either as high-knowledge learners (30 medical students) or low-knowledge students (30 non-medical students). Participants learned either with integrated or separated format about the physiological functioning of the kidney resulting in a 2x2 design. In addition of measuring the three cognitive load types by means of multiple subjective ratings

scales after learning, participants' eye movements were recorded during learning to learn more about their reading and learning behavior. Subsequent to the cognitive load ratings participants had to answer four knowledge tests as in Experiment 1.

Results 2. The expertise reversal effect was demonstrated for two out of four knowledge tests. Mediated moderation analyses yielded that germane cognitive load mediated about 19% of the expertise reversal effect on knowledge about labeling and about 38% on knowledge about complex facts. Moreover, the eye-tracking measures suggested that participants with separated format tended to process text and graphic in a more isolated way than participants with integrated format who switched more often between text and corresponding parts of the graphic.

Conclusions 2. The results showed that the expertise reversal effect is predominantly caused by germane cognitive load. As a result, instructors should aim at challenging advanced learners thereby increasing germane cognitive load and not preventing them from being overloaded by extraneous cognitive load.

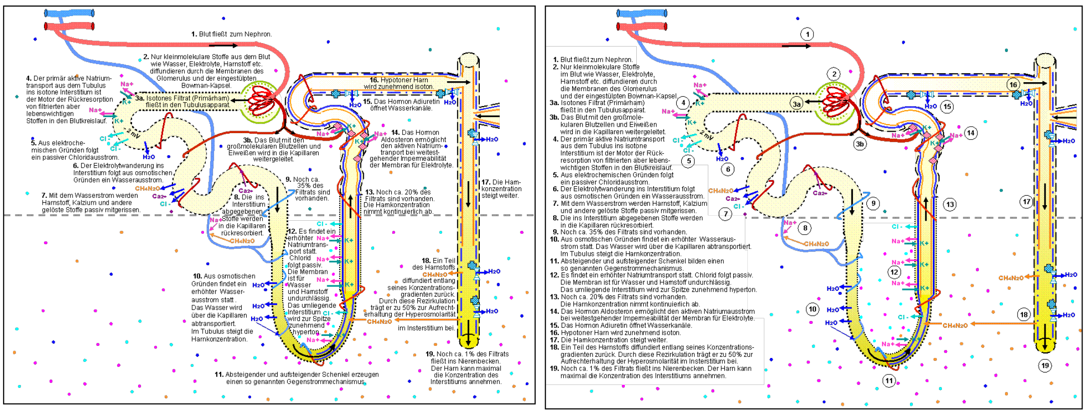
Overall conclusion. First, the results of both experiments stress that learning does not take place without learners' investment of cognitive resources. A reduction in extraneous cognitive load does not seem to be enough to make learning successful. Second, the measurement of different cognitive load types by means of multiple subjective ratings was rather successful but is still a critical issue that needs further validation from process measures like reading behavior or from neuronal correlates (e.g., EEG). Future research on the combination of cognitive load, viewing behavior, and neuronal correlates might support the development of trainings for successful processing strategies as well as the development of computer based learning environments that automatically adapt to learners current configuration with regard to cognitive load.

8

1 Introduction

Since the first PISA (Program for International Student Assessment) results of the year 2000 or at least since Bindé's UNESCO world report "Towards knowledge societies" in 2005, people's interest in and awareness of the importance of knowledge (in contrast to information only) as a primary production resource have increased. Because knowledge is such a central resource, learning processes have gained more and more attention. Central questions are how to help people better learn and how to improve instructions in such a way that knowledge acquisition becomes more effective. Research in instructional design is one important way to find profound answers to these questions. Multimedia research as one branch of instructional design research investigates which type of media should be used in which way to enhance learners' knowledge acquisition. For instance, teachers have to consider three aspects of media: (1) delivery media like books and computers, (2) sensory modalities concerning visual and auditory information, and (3) presentation modes, also termed codality (Brünken, Steinbacher, & Leutner, 2000), like text and illustrations (Mayer, 1997). Whereas the delivery media are meanwhile regarded as less important for learning (Clark, 1994; Mayer, 2003), the sensory modalities and the presentation modes are assumed to be critical for learning, because they build the gateway to the learner's cognitive system (Larkin & Simon, 1987; Schnotz, 1997). Concerning the sensory modalities, it has been argued that narrations (auditory text) may be more supportive than written text, when combined with animated illustrations (Mayer & Anderson, 1991). Concerning the presentation modes or codalities, it was shown that carefully constructed static illustrations enhance learning from text, whereby different types of illustrations serve different functions (Carney & Levin, 2002; Levin, Anglin, & Carney, 1987). For instance, in explaining scientific topics, so called interpretational graphics that explain the text should be added instead of merely decorative pictures. The supporting effect of these illustrations in learning with texts resulted in the so called multimedia principle that recommends adding interpretational graphics to texts (Mayer, 2001; Fletcher & Tobias, 2005). In an analysis of American science text books, Mayer (1993) showed that about half of the space in science textbooks is used for textual information and the other half is used for illustrations. Thus, when multimedia instructions are generated an issue that has to be considered is how to arrange written text and illustrations on the limited space of a textbook page or a computer screen. Because written text and pictures are extensively used in presenting information, it is important to know how learners deal with different arrangements of written text and static illustrations in multimedia learning.

About 15 years ago, multimedia researchers stated that “many studies have shown that graphics can make communication and learning more effective, but we only have recently begun to understand better why and under what conditions they are really effective” (Schnotz & Kulhavy, 1994, p. vi). Today, one can say without doubt that with respect to the conditions multimedia research has accumulated a substantial body of knowledge about how to present text and graphic in order to enhance learning. In his seminal book about multimedia learning, Mayer (2005a) presents seven basic and nine advanced design principles of multimedia learning that describe how to effectively combine different instructional design characteristics on the one hand and learner characteristics on the other hand. One important instructional design characteristic resulting from the limited space of presentation devices like books or computer screens is the spatial contiguity of text and corresponding picture. For instance, Figure 1a shows the blood filtering and urine production processes of the human kidney arranged in an integrated presentation format with high spatial contiguity, whereas Figure 1b shows the same information sources arranged in a separated presentation format with lower spatial contiguity.



1a. Physically integrated presentation format

1b. Physically separated presentation format (split-attention)

Figure 1. Different presentation arrangements of text and corresponding graphic

Numerous studies demonstrated that learners studying multimedia materials (e.g., text and graphic, text and mathematical equation, or equation and diagram) that were physically integrated outperformed learners studying the same materials arranged in a separated format (e.g., Ginns, 2006; Mayer, 2001). This phenomenon is called *spatial contiguity effect* according to the *Cognitive Theory of Multimedia Learning* (*CTML*; Mayer, 2001) and *split-attention effect* according to the *Cognitive Load Theory* (*CLT*; Sweller, Van Merriënboer, & Paas, 1998). The empirical findings resulted in the so called split-attention principle that recommends placing corresponding verbal and pictorial information close to each other (e.g., Ayres & Sweller, 2005). The term split-attention effect results from the impossibility to (visually) process disparate sources at the same time, and therefore being forced to process both information representations successively.

One important learner characteristic that has proven to influence learning is prior knowledge (Dochy, Segers, & Buehl, 1999; Shapiro, 2004). Prior knowledge can have a moderating influence on learning outcomes as already shown by the early *Aptitude-Treatment Interaction* approach (ATI; Cronbach & Snow, 1977). ATI effects describe the phenomenon that instructional methods or presentation formats that are highly effective for less knowledgeable learners can lose their effectiveness or even have detrimental effects when used with more knowledgeable learners. Concerning multimedia learning, there is first evidence that prior knowledge moderates the above mentioned spatial contiguity or split-attention effect. For example, in a series of three studies, Mayer and Gallini (1990) showed that low knowledgeable students with a fully integrated format had higher learning outcomes than low knowledge students with less integrated formats or text only, whereas there were no group differences among students with high prior knowledge. Furthermore, Kalyuga, Chandler, and Sweller (1998) conducted a longitudinal study and showed that novices in electrical circuits benefited from integrated formats compared to separated ones, whereas these students suffered from integrated formats when they were more advanced. This ATI phenomenon led to the *individual differences principle* (Mayer, 2001, 2009) as well as to the *prior knowledge principle* (Kalyuga, 2005; Mayer, 2005a) within the CTML literature, and was subsumed under the term *expertise reversal effect* within the CLT literature (Kalyuga, Ayres, Chandler, & Sweller, 2003). According to these findings, the instructional design should be adapted to learners' prior knowledge to optimize learning.

Although research on the split-attention and expertise reversal effect has specified under which learner characteristics specific instructional multimedia formats are most effective, relative little research has directly investigated the question of *why* integrated formats are more effective for less knowledgeable learners than for more knowledgeable learners and *why* separated formats are more effective for more knowledgeable learners than for less knowledgeable learners. As the CTML and CLT are the theoretical frameworks in instructional design research that offer explanations, they might be used as theoretical starting points. The two frameworks explain the split-attention and expertise reversal effect by referring to two basic mechanisms: (1) the more "negatively" oriented one that is based on the CLT and that focuses on processes inhibiting learning. According to this explanation, inhibiting processes should be removed. (2) The more “positively” oriented one that is based on the CTML and that focuses on processes promoting learning. According to this explanation supporting processes should be enhanced. Both explanations refer to *cognitive load* as the mediating variable which transmits the effect of spatial contiguity as well as the moderating effect of spatial contiguity and prior knowledge on learning outcomes. According to the inhibiting mechanism it is assumed that poor instructions cause unnecessary cognitive load that hinders knowledge acquisition, whereas according to the promoting mechanism it is assumed that good instructions cause cognitive load that is relevant for learning and thus supports knowledge acquisition. In explaining the split-attention effect, the CTML and CLT suggest the two cognitive mechanisms to function in a complementary way, whereas in explaining the expertise reversal effect, the mechanisms are used in a contradictory way. With regard to the potential theoretical explanations, however, research lacks detailed empirical results about the cognitive mechanisms underlying the split-attention and expertise reversal effect. This lack of knowledge about potential mechanisms which are based on cognitive load needs to be overcome by empirical data, because it has implications for teachers and other instructors with regard to whether they should make learning as easy as possible to not overload learners (according to the explanations focusing on cognitive load inhibiting learning) or whether they should try to challenge and activate learners (according to the explanations focusing on cognitive load supporting learning). In order to provide instructors with the most adequate attitude towards learners’ cognitive processes, it is the aim of this dissertation to test both mechanisms postulated, thereby gaining more insights into mediating cognitive processes in multimedia learning.

This dissertation is divided into three main sections. Section I provides the theoretical background with respect to cognitive load. Chapter 2 outlines the CLT and the CTML as the core frameworks of multimedia learning that use cognitive load to explain the split-attention and expertise reversal effect. Chapter 3 presents the state of the art with respect to cognitive load measurement. An overview over important measurement techniques and their general assumptions as well as their contributions and limitations in instructional design research is offered. Section II reports the empirical work provided by this dissertation and is divided into two chapters about the split-attention effect and into two chapters about the expertise reversal effect. Whereas Chapter 4 reviews the literature on the split-attention effect and shows the shortcomings of the existing research concerning potential cognitive load mechanisms, Chapter 5 describes the empirical method, results, and discussion of Experiment 1 that aimed at testing mechanisms underlying the split-attention effect by using objective as well as subjective cognitive load measures. Chapter 6 reviews the literature on the expertise reversal effect and demonstrates the caveats of existing studies. Chapter 7 describes the empirical method, results, and discussion of Experiment 2. The second experiment aimed at testing the explanations of the expertise reversal effect by means of subjective ratings and objective viewing behavior. Section III including Chapter 8 provides a general discussion, a critical reflection on the contributions and limitations of this dissertation, and suggestions for further research.

I. THEORETICAL BACKGROUND

Nothing is more practical than a good theory.

(Old proverb assigned among others to Albert Einstein and Kurt Lewin)

2 Theoretical Frameworks and Cognitive Load Explanations in Multimedia Research

Multimedia learning is defined as learning from visually or auditorily presented text and static or dynamic pictures. The main aim of investigating multimedia learning is to understand how learners process text-picture combinations in order to design instructions that promote learning (Mayer, 2005b). There are several theoretical frameworks in multimedia research that describe and/or explain multimedia learning. For instance, there are the framework of designs, functions and tasks (DeFT; Ainsworth 1999, 2006), the integrated model of multimedia learning and motivation (Astleitner & Wiesner, 2004) or the integrated model of text and picture comprehension (ITPC model; Schnotz, 2005; Schnotz & Bannert, 2003). These three frameworks, however, have not been as widely acknowledged as the *cognitive theory of multimedia learning* (CTML; Mayer, 2001, 2005) and the *cognitive load theory* (CLT; Sweller, Van Merriënboer, & Paas, 1998; Sweller, 2005a). The main reason for the success of the latter two is that well-known instructional design principles on how to promote multimedia learning were derived from and widely investigated within these two frameworks. Among these principles are the multimedia principle, the split-attention principle subsuming the spatial contiguity and the temporal contiguity principle, the modality principle, the redundancy principle, and the prior knowledge principle that is based on the expertise reversal effect (for a detailed overview see Mayer, 2001, 2005a, 2009). Each of these principles is based on a corresponding empirical effect. To explain these effects, both frameworks refer to cognitive load types. With regard to the split-attention effect and expertise reversal effect, however, the details of their explanations show some inconsistencies. These inconsistencies lead to the question whether multimedia research has already understood and described the ongoing learning processes in sufficient detail. To understand these inconsistencies it is important to consider how both frameworks explain multimedia learning. Although both frameworks meanwhile share a triarchic model of cognitive load, they focus on different aspects or types of cognitive load when they describe multimedia learning. Whereas the CLT focuses mainly on cognitive processes that prevent learning (inhibiting mechanism) and thus should be minimized, the CTML focuses mainly on cognitive processes that stimulate learning (promoting mechanism) and thus should be increased. To understand the different foci and with them the inconsistencies in explaining the split-attention (spatial contiguity) and expertise reversal effects, both frameworks will be presented in the following sections. After the description of each framework the explanation of the split-attention and the expertise reversal effect derived from the framework will be outlined. The chapter will end with a summary of the different cognitive load explanations.

2.1 Cognitive Load Theory

The framework of CLT has been developed since the late 1980s by John Sweller and his colleagues (e.g., Sweller, 1988; Chandler & Sweller, 1991). It is a general instructional design framework and not restricted to multimedia learning. Instructional design effects are explained by referring to three different types of cognitive load: (1) intrinsic cognitive load, (2) extraneous cognitive load, and (3) germane cognitive load. This level of generalization has the great advantage that the framework can be applied to very different learning situations that differ with regard to specific processes that can be categorized into a small number of cognitive load types. Therefore, CLT has been described as a unifying “umbrella theory” (Gerjets, Cierniak, Scheiter, 2008). However, this unifying character is only advantageous, in case that one knows which cognitive processes are occurring during learning and if these processes can reliably be assigned to one of the three cognitive load types. The following description of the framework concentrates on its structure presented by Sweller et al. in 1998, thereby not including the analogy between evolution by natural selection and learning introduced by Sweller (2004, 2009b, Sweller & Sweller, 2006) rather recently. In describing the framework the first focus is set on cognitive structures. The second focus is set on cognitive load as the core concept of the framework. Afterwards, the cognitive mechanisms suggested by CLT to explain the split-attention and expertise reversal effects are presented.

2.1.1 Cognitive Architecture and Representations

CLT is based on a memory system with different memory stores as has already been suggested by Atkinson and Shiffrin (1968). CLT concentrates on two memory stores, a working memory with a limited cognitive capacity and a potentially capacity unlimited long-term memory (Sweller, 2005a; Sweller et al., 1998). Knowledge is assumed to be stored in long-term memory. Figure 2 depicts a flowchart of the CLT framework.

Working memory. Before information is stored in long-term memory, it must be processed in working memory. CLT assumes that working memory comprises substructures like a visuo-spatial sketchpad for processing visual information and a phonological loop for verbal information (cf. Baddeley, 1996). In contrast to Baddeley's (1996, 2000) working memory model that assumes a separate substructure that functions as central executive to manage for example knowledge acquisition by means of an integration process, CLT does not assume such a specific cognitive substructure but relies on the means-ends heuristic, well designed instructions, or prior knowledge to manage knowledge acquisition. The critical characteristic of working memory is its limited capacity. Persons can only hold four (Cowan, 2001) to seven (Miller, 1956) single elements simultaneously active in working memory. Without active rehearsal (Baddeley, 1996), new elements can be held only for about 20 seconds in working memory (Peterson & Peterson, 1959). Hence, new complex information that consists of many interacting elements is very hard to handle within the limited working memory and can easily overload it. Cognitive overload, however, prevents the successful acquisition of new knowledge. Although the number of single elements is limited, their size and complexity is not. This offers the opportunity to deal with complex information in working memory, if the elements consist of well organized information packets or so called schemas.

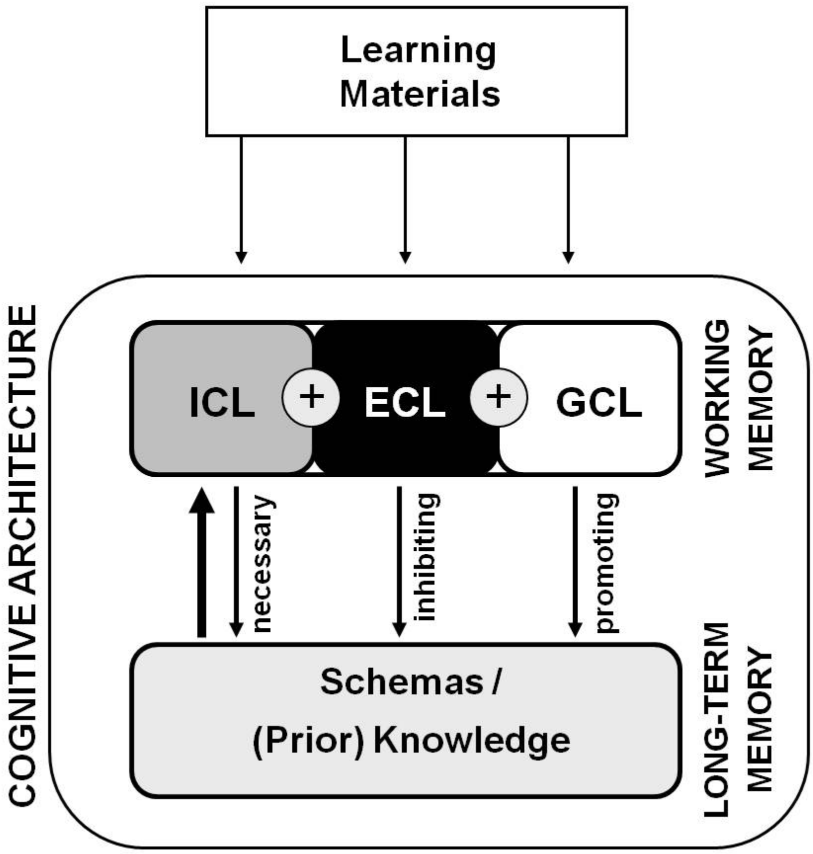


Figure 2. Flowchart of the CLT

Long-term memory and schemas. According to CLT a person's knowledge is stored by means of schemas in long-term memory (Bartlett, 1932; Chi, Glaser, & Rees, 1982; Rumelhart & Norman, 1978). Schemas are mental representations or "organized packets of information about the world" stored in long-term memory (Eysenck, 2006; p. 275). Schemas contain information or knowledge of persons, objects and procedures (so called scripts) in abstract and general form serving as abstract problem categories. Skilled performance and knowledge develops by building more and more complex schemas by means of (1) matching new information with existing information in available schemas (accretion), (2) restructuring existing schemas, if new information does not fit currently available schemas (restructuring), and (3) adjusting the terms to improve accuracy, generalizability, specificity as well as to determine default values (schema tuning; Rumelhart & Norman, 1978). Consequently, prior knowledge organized in schemas guide automatically how further information on the domain is comprehended and integrated into schemas or long-term memory. Thus, schemas influence or help guiding how to process new information (cf. Bransford & Johnson, 1972; Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995). Therefore, persons' prior knowledge or available schemas are an important variable that needs to be considered, when designing instructional materials. If learners do not have prior knowledge that guides or structures their learning process, they depend on the guidance provided by the instructional format. If an instruction does not provide enough guidance that shows a learner the structures and processes that support them to construct meaningful schemas, a learner has to rely on means-ends heuristics demanding the learner to test by coincidence which structures and processes are related with each other and how to understand or solve the issue to be learned. However, such learning by means of means-ends heuristics imposes high cognitive load onto learners' working memory and prevents a successful construction of schemas. Therefore, the aim of CLT is to generate instructional formats that support optimizing learners' cognitive load in working memory by offering a meaningful guidance.

2.1.2 Cognitive Load

Before the above mentioned three types of cognitive load are described in detail, an overview of their historical development, a general overview of their defining characteristics and three basic assumptions about their relations are presented.

Historical development. The construct of cognitive load underwent a profound development from the first description in the late 1980s (e.g., Sweller, 1988; Chandler

& Sweller, 1991) until the description of the framework by Sweller et al. in 1998 about ten years later. At its beginnings, CLT did not distinguish between different cognitive load types. Rather, it was claimed that learners with poorly designed instructions suffer from cognitive overload. Later on, the construct of cognitive load was differentiated into intrinsic and extraneous cognitive load as two distinguishable load types (Sweller, 1993, 1994; Chandler & Sweller, 1994). If both load types exceed working memory capacity, learning is prevented. The concept of cognitive overload is still very important and builds CLT's focus on so called inhibiting processes because they prevent successful learning. The focus of explaining learning success by removing cognitive load (processes irrelevant to learning) from learners' working memory did first change when the construct of germane cognitive load was included into CLT in 1998 (Sweller et al., 1998). Only since that time the general argumentation changed from facilitating learning by reducing cognitive load into enhancing learning by balancing cognitive load, that is, reducing extraneous and increasing germane cognitive load (Van Gog & Paas, 2008).

Defining characteristics. Before information is learned, it must be processed in working memory. Such information processing imposes cognitive load onto the capacity limited working memory. In order to not overload working memory, instructions should be designed by considering the following cognitive load types: (1) intrinsic cognitive load, (2) extraneous cognitive load, and (3) germane cognitive load. The three types differ with regard to their source, their underlying cognitive processes and their effects on schema construction (Paas, Renkl, & Sweller, 2003; Sweller et al. 1998). Table 1 summarizes the defining characteristics of the three load types.

Table 1

The Defining Characteristics of the Three Cognitive Load Types (Gerjets, Scheiter, & Cierniak, 2009)

Load type	Source	Cognitive processes	Effect on schema construction
ICL	Domain complexity x prior knowledge	Holding interacting elements active in WM simultaneously	Is necessary but harmful if exceeding WM capacity
ECL	Poor instructional design characteristics	Processes irrelevant to schema construction	Inhibits schema construction
GCL	Supportive instructional design characteristics	Processes relevant to schema construction	Promotes schema construction

Note. ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load, WM = working memory.

Basic assumptions on relations. There are three basic assumptions concerning the relation of the three load types. First, the three types of cognitive load are assumed to be additive and second, the relations between the three load types are assumed to be asymmetric (Paas, Renkl, & Sweller, 2003). That is, intrinsic cognitive load provides a base load that is irreducible (except via increasing expertise or prior knowledge). If learners deal with a content (intrinsic cognitive load), only the leftovers of working memory capacity are available for cognitive load caused by the instructional design (extraneous and germane cognitive load). If intrinsic cognitive load is low, there is relative much capacity left for load imposed by the instructional design. The higher the intrinsic cognitive load is during learning, the more critical gets a high degree of extraneous and/or germane cognitive load, because the total load should not exceed available working memory resources, if learning is to occur. If the sum of the three cognitive load types requires more working memory resources than a learner has at his or her disposal at a specific point in time during learning, it results in cognitive overload. Third, it is assumed that the composition of intrinsic, extraneous, and germane cognitive load mediates learning outcomes. CLT suggests that the relations between instructional design and cognitive load as well as between cognitive load and learning outcomes are direct. This strict model (see Figure 1) was augmented by Gerjets and Hesse (2004) who included learner activities (e.g., learning goals and strategies) that influence cognitive load. These learner activities are moderated by individual learner characteristics (e.g., epistemological beliefs and attitudes), thereby allowing for more variability in the cognitive load type patterns and learning outcomes. According to this augmented view, differences in learner activities due to self-regulation should also be considered to better predict the relation between instructional designs and load type patterns.

Intrinsic cognitive load. Intrinsic cognitive load is defined by the amount of cognitive load caused by the number of interacting elements that need to be actively held in working memory at the same time to represent the content to be learned. It can also be described as the complexity of the learning content or task. Thus, intrinsic cognitive load is necessary load. Learning contents with low element interactivity or low intrinsic cognitive load require learners to process only few elements at the same time, whereas learning contents with high element interactivity require learners to processes many elements simultaneously resulting in high intrinsic cognitive load. For example, understanding complex sentences or understanding mechanical or biological systems that consist of many interacting processes cause high intrinsic cognitive load. If intrinsic cognitive load is too high, learning outcomes are reduced if the additional extraneous and/or germane cognitive load exceeds working memory capacity. In general, intrinsic cognitive load is “imposed by the basic characteristics of the information rather than by instructional design” (Sweller 1993, p. 6). Hence, intrinsic cognitive load is determined by the learning content and cannot be altered by the instructional design. However, intrinsic cognitive load is influenced by learners’ prior knowledge. If learners already have some prior knowledge, many elements of the content to be learned are already organized in a schema that can be held as one element in working memory, thereby reducing intrinsic cognitive load. The construct of intrinsic cognitive load was introduced into the framework by Sweller (1993, 1994; Chandler & Sweller, 1994) to differentiate more precisely between cognitive load caused by the learning content and by the instructional format. Before that time, cognitive load was used to describe the cognitive load caused by poor instructional formats only (cf. Sweller, in press). However, Sweller and colleagues found out that poor instructions were only harmful, if the intrinsic cognitive load was high (e.g., Chandler & Sweller; Sweller & Chandler, 1994). In line with the asymmetry assumption of the cognitive load types, Leahy and Sweller (2005) state that “If the intrinsic cognitive load of the materials as determined by element interactivity is low, extraneous cognitive load may not be critical because of total cognitive load may not exceed working memory capacity” (p. 268). Ginns (2006) showed that the effect size of the influence of spatial contiguity on learning outcomes is medium to high (d = 0.78), if the intrinsic cognitive load of the learning content is high, whereas the effect size is low (d = 0.28), if the intrinsic cognitive load is low. Thus, the effect of spatial contiguity on learning outcomes decreases or disappears, if the learning content is rather easy.

Extraneous cognitive load. All learning contents are presented in an instructional format that determines how a learner processes it (see the above mentioned strict view of instructional design, cognitive load, and learning outcome). If an instruction is designed in a poor way, learners have to exert irrelevant processes. These learning irrelevant processes impose extraneous cognitive load on learners’ working memory. The construct of extraneous cognitive load has been the first and core construct of CLT. It is assumed to be directly caused by a poor instructional design, thereby being under the direct control of the instructor. Thus, the main aim of CLT has been to reduce extraneous cognitive load. Eight out of twelve design principles derived by CLT aim at reducing extraneous cognitive. Only two principles aim at reducing intrinsic cognitive load, and two principles aim at increasing germane cognitive load (Sweller, 2009a). Assuming a constant amount of intrinsic cognitive load, the higher the extraneous cognitive load gets during learning, the lower are the learning outcomes, because the faster working memory is overloaded. Depending on the learning material and task, extraneous cognitive load is caused by different processes. Whereas in conventional problem solving instructions extraneous cognitive load is assumed to be caused for instance by processes of a means-ends heuristic (Renkl, 2005), in multimedia learning extraneous cognitive load is assumed to be caused for instance by mentally integrating information of physically separated text-picture formats (Ayres & Sweller, 2005). Furthermore, it is assumed that the processing of redundant information causes extraneous cognitive load (Sweller, 2005). These assumptions will be explained in more detail further below. CLT has emphasized the opinion that getting rid of irrelevant (extraneous) or inhibiting processes leads to better learning outcomes, because learners are no longer overloaded. The strong focus on reducing extraneous cognitive load is characteristic for CLT and has been described by Rey (2009) as a “less is more” approach and is described in this thesis as the approach with a focus on inhibiting processes. In line with this approach was the introduction of instructional efficiency. Paas and Van Merriënboer (1993) suggested combining cognitive load ratings and outcome measures to test which instructional design resulted in the highest learning outcomes with the lowest cognitive load (see Chapter 3). However, with introducing germane cognitive load into the framework (Sweller et al. 1998), this simple construct of instructional efficiency does no longer fit the more elaborated framework.

Germane cognitive load. Beyond causing irrelevant processes an instructional design can also foster learning. An instruction that directs learners’ attention towards processes relevant for more elaborated learning is said to cause germane cognitive load (Sweller et al. 1998). The higher germane cognitive load is during learning, the higher the learning outcomes, as least as long overall load (all three load types summed up) does not exceed working memory capacity. For example, actively imagining procedures or concepts (compared to studying them only) are thought to increase germane cognitive load, thereby enhancing learning outcomes (Leahy & Sweller, 2004). Moreover, increases in effort or motivation that promote learning by making learners apply more elaborated strategies are also possibilities to increase germane cognitive load (Paas et al. 2003). Considering all three cognitive load types, CLT states that appropriate instructional design formats keep intrinsic load as low as possible, reduce extraneous cognitive load, and increase germane cognitive load. Van Gog and Paas (2008) reconsidered the construct of instructional efficiency in line with the more elaborated framework by acknowledging that the general argumentation changed from facilitating learning by reducing cognitive load into enhancing learning by balancing cognitive load, that is, reducing extraneous and increasing germane cognitive load. Sweller (2005a) concluded that “The aim of instruction should be to reduce extraneous cognitive load caused by inappropriate instructional procedures.

Reducing extraneous cognitive load frees working memory capacity and so may permit an increase in germane cognitive load" (p. 28). Hence, the focus on inhibiting processes has been attenuated by including the construct of germane cognitive load and the aim to optimize (in contrast to reduce) cognitive load. According to the triarchic model of cognitive load, the above mentioned "less is more" approach (Rey, 2009) is no longer valid. However, as also mentioned above, most instructional design effects are still explained by referring to extraneous cognitive load only (Sweller, 2009a).

2.1.3 Explaining the Split-Attention and Expertise Reversal Effect by CLT

In explaining instructional design effects like the split-attention and the expertise reversal effect CLT researchers assume cognitive load to be the underlying cause that mediates or transmits the effects. After describing the three types of cognitive load above, the following sections present first, how CLT assumes that the split-attention effect caused by a lack of spatial contiguity is mediated by cognitive load and second, how CLT assumes that prior knowledge influences or moderates the effect of spatial contiguity on cognitive load in such a way that the split-attention effect can be turned up-side down with more knowledgeable learners resulting in an expertise reversal effect. The explanations of both effects concentrate on extraneous cognitive load, and thus, represent the above mentioned focus on inhibiting processes assumed to underlie these effects.

2.1.3.1 CLT and the Explanation of the Split-Attention Effect

When Tarmizi and Sweller (1988) conducted the first experiment on the spatial split-attention effect, CLT did not yet distinguish between the three load types but focused on overall or extraneous cognitive load. Tarmizi et al. (1988) investigated the effectiveness of worked-examples on learning geometry. They showed that only worked examples in which the solution steps were integrated in the diagram enhanced learning outcomes compared to a conventional separated worked-example format and to a conventional problem solving task. The authors argued that a separated format demands learners to split their attention between text and graphic. Because of this physical split between the information sources in a separated format, learners are forced to mentally integrate both types of information. This process of mental integration is said to cause high extraneous cognitive load. In contrast, integrated formats do not demand learners to mentally integrate disparate information sources, because the information is already integrated physically. Thus, by reducing the need to mentally integrate disparate information extraneous cognitive load should be reduced and learning should be facilitated. Because learners with the integrated worked-examples needed significantly less time to solve the problems in the acquisition phase than learners with the separated formats, Tarmizi and Sweller (1988) concluded that the reduced cognitive load freed cognitive resources and facilitated performance. Accordingly, Ward and Sweller (1990) argued that the critical feature of presentation formats is to “impose a relatively light cognitive load” (p. 4). Hence, mental integration was thought to cause extraneous processing. Later on, Sweller and Chandler (1994) introduced visual search in multimedia learning by explaining that the act “… of mental integration involves finding relations among elements associated with the diagram and statements. Unless the relevant relations among the elements are found, the instruction will be unintelligible. Finding relations among elements requires cognitive resources that must be expended…” (p. 192-193). According to this search-and-match processes as behavioral correlate of a mental integration processes, several authors assumed that extraneous cognitive load or mental integration might be reflected in learners’ viewing behavior. For instance, Erhel and Jamet (2006) and Tabbers, Martens, and Van Merriënboer (2000) assume that learners with a separated format switch frequently back and forth between text and illustration to integrate the disparate information mentally. However, these authors did not measure learners’ viewing behavior to test their assumptions. To sum up, according to CLT the high extraneous cognitive load due to learning with a separated format mediates the split-attention effect. The high extraneous cognitive load is caused by mental integration processes that may be reflected in visual search processes (Sweller, 2010).

Although extraneous cognitive load was traditionally assumed by CLT researchers to be the mediator of the split attention effect, it has also been argued within a CLT framework that besides the difference in extraneous cognitive load germane cognitive load might be enhanced when learning with an integrated format and reduced when learning with a separated format. (Kester, Kirschner, & Van Merriënboer, 2005; Tabbers et al., 2000). Considering germane cognitive load as a second factor mediating the split-attention effect is in line with Mayer’s first explanation of the split-attention effect derived by the CTML framework (see further below). However, the reason why CLT researchers included germane cognitive load into the explanation can be mainly traced back to the fact that the above mentioned authors did not find any differences in overall subjective cognitive load ratings. Hence, the argument why germane cognitive load is included into the explanation of the split-attention effect by some CLT researchers is not completely convincing. A more detailed review on whether and which empirical evidence is in line with a focus on extraneous cognitive load only or with the inclusion of germane cognitive load in explaining the split-attention effect will be provided in Chapter 4 subsequent to discussing the measurement of cognitive load in Chapter 3.

2.1.3.2 CLT and the Explanation of the Expertise Reversal Effect

As already mentioned, prior knowledge is a learner characteristic that is important to consider when designing instruction. According to Kalyuga (2005, 2007, 2009; Kalyuga et al. 1998, 2003) high knowledge learners do not benefit or even suffer from high-structured materials like integrated formats because they result in the processing of *redundant* information that causes extraneous cognitive load. Learners with high prior knowledge suffer from redundancy, because prior knowledge influences whether additional information is really essential for them or whether it is just redundant. To understand this explanation in greater detail, it is necessary to introduce the redundancy effect.

Redundancy. In contrast to essential information sources or complementary information sources that enhance or elaborate each other (e.g., an illustration is added to a text to convey all spatial information that is important to understand the content to be learned but which is not contained in the text), redundant information does not convey new information and does not elaborate information but is just equivalent information from another information source (e.g., the words “Blood flows from the body into this structure” next to an arrow depicting the way of blood flow from the body to a specific structure). Chandler and Sweller (1991) demonstrated that students with an integrated format about the blood flow in the heart (picture of the heart with text that just described easily to understand graphical information like arrows indicating the blood flow) learned less than students without such textual redundant information next to arrows. Similar results were demonstrated by Bobis, Sweller, and Cooper (1993) in a paper-folding task. Moreover, Chandler and Sweller (1991) demonstrated that students with a separated format of an electrical circuit who were instructed to explicitly read the redundant text information and relate it to the diagram studied longer and performed worse than students with a separated format who were not explicitly instructed to process the text. This phenomenon is called redundancy effect in CLT (for an overview see Sweller, 2005b). The redundancy principle recommends that verbal information should be avoided, if a diagram is intelligible in itself. Empirical evidence of the redundancy effect was also demonstrated in multimedia instructions with written text that was accompanied by the same text spoken by a narrator (e.g., Kalyuga, Chandler, & Sweller, 1999; Mayer et al., 2001). Thus, the redundancy effect is thought to appear whenever redundant instead of essential information is presented in instructions. According to Sweller (2005b), CLT provides the framework for the phenomenon that redundant information has a detrimental effect on learning. Whenever identical information is presented in multiple ways, it must be processed, coordinated with each other and with new essential information. When these coordination processes require working memory resources but do not contribute to learning (cf. Ainsworth, 2006), they cause extraneous cognitive load. Hence, students processing redundant information suffer from extraneous cognitive load which inhibits successful knowledge acquisition.

Prior knowledge. Concerning redundancy and prior knowledge the following conclusion is made by CLT: The higher learners' prior knowledge, the more complex are their schemas. The more complex and elaborated the learners' schemas, the higher is the probability that one information source (e.g., illustration) is enough for learners to understand the topic and the higher is the probability that any additional information source (e.g., text) is no longer essential but becomes redundant. Thus, whether information is redundant or not is not an objective characteristic of information but a matter of learners' prior knowledge. Kalyuga et al. (2003) assume that learners with high prior knowledge and for instance integrated formats "must integrate and cross-reference redundant information with their available knowledge schemas" (p. 29). In learning with integrated text-graphic formats, more knowledgeable learners might prefer to ignore the text but may have difficulty in doing so. Hence, they have to process the graphic and the redundant text (Kalyuga et al., 1998). Processing the integrated text and cross-referencing it with prior knowledge, however, is not necessary for learning but needs working memory capacity. According to these CLT assumptions, integrated formats increase extraneous cognitive load for more knowledgeable learners and thereby reduce their learning outcomes. This extraneous cognitive load explanation clearly differs from Mayer's (2005c) assumption that learners with higher prior knowledge can actively compensate poor multimedia instructions (e.g., separated formats) by means of applying processes relevant for learning.

Moreover, Kalyuga and colleagues (1998, 2003) also stated that separated text-picture formats and low-cohesive texts only formats share a similar low-guidance pattern, whereas integrated formats and high-cohesive texts share a similar high-guidance pattern, and thus, these material types are comparable on a cognitive load level. Thus, Kalyuga and colleagues explain the finding that learners with high prior knowledge suffer from high-cohesive texts and benefit from low-cohesive texts whereas learners with low prior knowledge benefit from high-cohesive texts and suffer from low-cohesive texts (McNamara et al., 1996) by referring to redundant processing causing high extraneous cognitive load in high- knowledge learners with high-cohesive texts. However, text comprehension researchers (e.g., McNamara et al., 1996) explain this type of expertise reversal effect within text comprehension research by referring to compensatory processing of high-knowledge learners with low-cohesive texts.

Because their explanation is very similar to the one provided by Mayer and colleagues within multimedia research, it will be described in more detail, when Mayer's explanation of the expertise reversal effect within multimedia leaning is presented. To understand Mayer's explanations in its details, the CTML framework will be described first in the next subchapter.

2.2 Cognitive Theory of Multimedia Learning

In contrast to the general framework of CLT, the CTML is a specific framework on multimedia learning. The CTML has been developed by Richard Mayer and his colleagues since the late eighties of the last century (Mayer, 1989; Mayer, Heiser, & Lonn, 2001). At the beginning, Mayer (1989) focused on different processes like selecting, organizing, and integrating information. Later on, Mayer and Sims (1994) stressed the distinction between verbal and pictorial mental channels in working memory. Mayer, Steinhoff, Bower, and Mars (1995) emphasized the active role of learners. Since 2001, the framework has been presented in its current version with a clear description of constructive learning processes, mental representations as well as cognitive structures including a central executive (Mayer et al., 2001). Moreover, Mayer and colleagues adopted the triarchic model of cognitive load as suggested by CLT, although they used somewhat different terms (cf. Mayer & Moreno, 2003; Mayer, Hegarty, Mayer, & Campbell, 2005; Moreno & Mayer, 1999). Before the explanations of the split-attention and expertise reversal effect proposed by the CTML are presented, the core aspects of the framework will be described. The CTML comprises three important assumptions. The dual-channel and limited capacity assumptions concern the cognitive architecture, whereas the active-processing assumption concerns cognitive processes of meaningful multimedia learning. First, the cognitive architecture will be outlined before relevant multimedia learning processes within this cognitive architecture are described. Subsequent to these core aspects, the adoption of the triarchic model of cognitive load will be outlined. This section will end with the explanation of the split-attention and expertise reversal effect according to the CTML.

2.2.1 Cognitive Architecture

Mayer (2001, 2003, 2005c) distinguishes three memory stores: (1) a sensory working memory for visual and auditory information perceived via the ear or eye, respectively, (2) a working memory with a dual-channel system, limited capacity, and central executive, and (3), a limitless long-term memory. Figure 3 depicts a model of the CTML framework.

Sensory working memory. Information enters sensory working memory via the eyes (visual information) or ears (auditory information). The sensory working memory is very time-limited and can hold a literal representation of the information only for a few milliseconds. Therefore, information must be selected into working memory very quickly.

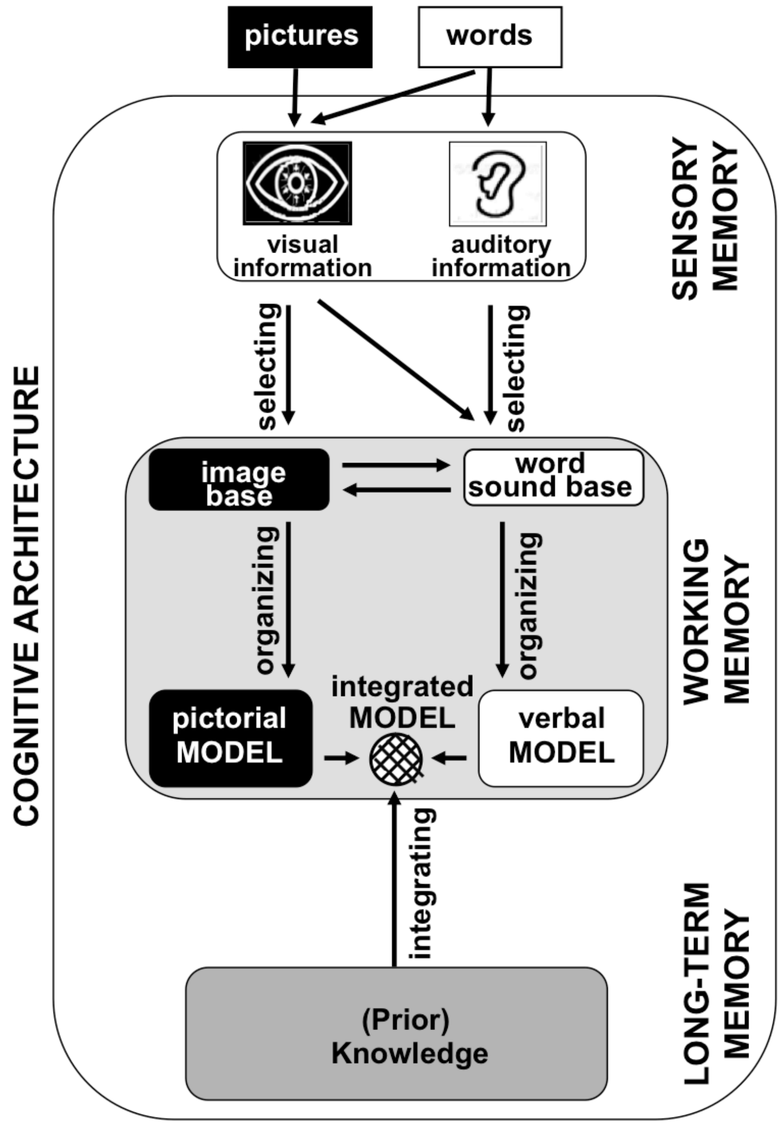


Figure 3. Flowchart of the CTML (modified version of Mayer, 2005a; cf. Rummer, Schweppe, Scheiter, & Gerjets, 2008)

Working memory. The central processes of multimedia learning take place in working memory that consists of substructures to process information of different presentation modes or codalities as well as of different sensory modalities. To integrate presentation modes or codalities and sensory modalities the dual-channel assumption refers to two theories, namely to Paivio's dual-coding theory (1986) as well as to

Baddeley's (1996, 2000) working memory model. By referring to these theories, Mayer (2001) first assumes that verbal material like text is processed in a so called auditory / verbal channel, whereas pictorial materials like illustrations are processed in a visual / pictorial channel. Second, he assumes that auditory information is processed in the auditory / verbal channel, whereas visual information is processed in the visual / pictorial channel. Thus, words presented as text are first processed in the visual / pictorial channel and then transferred into the auditory / verbal channel, whereas words presented as narrations are processed in the auditory / verbal channel from the very beginning. Notably, these assumptions are not identical to Baddeley's (1996, 2000) assumption within his working memory model including a phonological loop which processes verbal information no matter whether it is presented in auditory or visual mode and a visuo-spatial sketchpad which processes spatial and visual information no matter whether it is presented in pictorial or verbal codality. However, Mayer (2001, 2005c) does not elaborate these differences (for a more elaborated distinction see Rummer et al., 2008). According to the limited capacity assumption, the information that can be processed in one channel at one time is limited. This assumption goes back to Miller's memory span tests (1956). The amount of information can only be increased by chunking informational elements (Gobet, 2005). Due to this process, the cognitive capacity may remain the same, but more elements can be remembered within a chunk. Because of the limited capacity a learner has to decide which information (s)he should attend to and between which new informational elements and between which new and old informational elements (s)he should build connections. Mayer (2005c) refers to the cognitive subsystem which is used to allocate, coordinate, and control the limited capacity as the central executive according to Baddeley (1996). Although Baddeley (2000) introduced an episodic buffer as a further structure into his model, where the processes of integration can take place, such a structure has not been included in Mayer's CTML so far.

Long-term memory. The long-term memory contains a person's (prior) knowledge. Whereas CLT (Sweller et al.,1998) refers to schema theory to describe the structure of knowledge in more detail, Mayer (2005c) refers to different mental representations like for example a word base or a verbal model (see below) that seem to be similar on the one hand to the different mental representations described by the model of text comprehension and production proposed by Kintsch and van Dijk (1978) that was adapted into the ITPC model by Schnotz and Bannert (2003) and on the other hand to the mental model approach suggested by Johnson-Laird (1983). However, Mayer does not specify the mental representations in his model in such detail and does not refer to the above mentioned authors. However, similar to CLT, the CTML states that to actively process or think about already established knowledge and to incorporate new information or change old knowledge structures, a person has to bring the established knowledge first back to working memory where it must be actively processed.

2.2.2 Learning Processes and Cognitive Representations

Mayer (2004) describes learning as a cognitively constructive process as introduced by Wittrock (1989). According to the active processing assumption, learners cannot passively acquire knowledge but have to actively engage in cognitive processing to construct a coherent mental representation that makes sense in relation to their prior knowledge base. Mayer (2005c) distinguishes five processes concerning paying attention, organizing incoming information, and integrating information with existing knowledge. The five processes do not necessarily occur in linear order, rather a learner has to coordinate and monitor these processes by means of the central executive. The five processes take place in working memory and result in five mental representations.

Selecting relevant words. For example, a learner pays attention to the verbal information of the learning material and selects this information as a *word base* into the auditory / verbal channel of his / her working memory. If verbal information is presented auditorily (narration), the information is directly encoded in the auditory / verbal channel. If verbal information is presented visually (text), the information is first encoded in the visual / pictorial channel and in a second step encoded in the auditory / verbal channel. Because of the limited capacity the learner determines which words are most relevant and then select these words. One may assume, although Mayer does not state this explicitly for this level of processing, that learners with sufficient prior knowledge are more successful in selecting the relevant verbal information than learners without prior knowledge.

Selecting relevant images. For example, a learner pays attention to the pictorial information (illustration) of the learning material and selects this information as an *image base* into the pictorial channel of his / her working memory. Because of the limited capacity the learner must determine which parts of the illustration are most relevant and then select these parts. One may also assume (also Mayer does not state this explicitly) that learners with sufficient prior knowledge are more successful in selecting the relevant pictorial information than learners without prior knowledge.

Organizing selected words. By building connections among pieces of verbal knowledge the learner organizes the word base into a coherent mental representation, the so called *verbal model*. Because of the limited capacity the learner must focus on building a simple structure which often represents a cause-and-effect chain. Also Mayer does not state this explicitly, one may assume that making inferences may be an exemplary process of organizing. Again, learners with sufficient prior knowledge should be more successful in organizing verbal models than learners without prior knowledge.

Organizing selected images. By building connections among pieces of pictorial knowledge the learner organizes the image base into a coherent mental representation, the so called *pictorial model*. Because of the limited capacity the learner must again focus on building a simple structure which often represents a cause-and-effect chain. As mentioned above, learners with sufficient prior knowledge should be more successful in organizing pictorial models than learners without prior knowledge.

Integrating word-based and image-based representations. The most crucial step in multimedia learning is to make connections between word-based and image-based representations. In this *integrated model*, corresponding elements and relations from the verbal and pictorial model are mapped onto each other. To successfully map corresponding visual and verbal representations, learners have to hold both types of representations at the same time in working memory. Moreover, the integrated model also includes connections to prior knowledge (cf. the comments above). Building an integrated model is the most demanding process in the CTML that imposes heavy loads onto working memory. Although constructing an integrated model is the overall aim in successful multimedia learning, the CTML does not specify its representation mode or codality in greater detail. A specification concerning the mode of the different mental representations is presented within the ITPC model by Schnotz and Bannert (2003). These authors differentiate between symbolic propositions based on text and analogue mental models based on pictures. According to this model, a successful integrated model is not a single representation but successful translation processes of model construction and model inspection between a propositional representation and a mental model (Schnotz, 2005).

According to CTML, the more successfully learners construct knowledge representations, the more successfully they will be later able to solve retention and transfer tasks. In contrast to retention tasks demanding the recall of factual knowledge that represents knowledge at the level of the verbal and image models, transfer tasks demand inferences and thus require knowledge at the level of the integrated model.

2.2.3 CTML and Cognitive Load

In investigating cognitive mechanisms underlying instructional design effects, the three types of cognitive load are useful constructs, because different cognitive processes that share the same effect on learning outcomes can be subsumed under each type. This enables research to compare explanations of cognitive mechanisms on the same dimensions, even though these explanations are based on different learning activities which can be observed directly (e.g., reading text, inspecting graphics, or solving mathematical equations) and which should elicit cognitive processes not directly observable (e.g., integrating verbal and pictorial information mentally or drawing inferences). Hence, before the explanations of the split-attention and expertise reversal effect made by the CTML are presented, it will be outlined which processes of multimedia learning are subsumed under the three types of cognitive load by the CTML.

The framework of the CTML stresses the limited capacity of working memory like CLT does. In contrast to CLT, however, the CTML concentrates on describing and specifying which constructive learning processes should take place for multimedia learning to be effective. According to Mayer (2001, 2005c, 2009) learners have to mentally integrate verbal information, pictorial information, and prior knowledge to benefit from multimedia materials. Despite this focus on knowledge enhancing processes during multimedia learning Mayer and colleagues adopted the constructs of intrinsic, extraneous, and germane cognitive load, even though they use different terms (cf. Mayer & Moreno, 2003; Mayer, Hegarty, Mayer, & Campbell, 2005; Moreno & Mayer, 1999). Mayer and Moreno (2003) described essential processing to be equivalent to germane cognitive load. Essential processing was said to refer to the five learning processes (see above). Because Mayer specifies relevant multimedia learning processes, he focuses on germane cognitive load. Today, Mayer (2009) terms the five essential multimedia learning processes generative processing. Although the term was changed, the five learning processes can still be interpreted as germane cognitive load. Furthermore, Mayer's (2009) term essential processing is now used in the sense of intrinsic cognitive load. Some years ago, the term incidental processing was described to be equivalent to extraneous cognitive load and Mayer (2001) emphasized that extraneous cognitive load is caused by the instructional design. In line with CLT, incidental processing was said to refer to irrelevant processes like visual search in learning with disparate text-graphic formats (Mayer & Moreno, 2003). Today, Mayer (2009) uses the term extraneous processing to describe visual search. The term representational holding was first described as intrinsic cognitive load and said to refer to holding a representation in mind over a period of time (Mayer & Moreno, 2003).

Later on, however, Mayer et al. (2005) described representational holding "…as an example of extraneous processing" (p. 258). Thus, visual search and representational holding during visual search are processes specified as extraneous cognitive load. Integrating word-based and image-based representations with each other and with prior knowledge, however, is specified as processes of germane cognitive load. Notably, the process of integrating verbal and pictorial information into an integrated model is the most elaborated and important one that is necessary for successful multimedia learning according to the CTML. This focus on germane cognitive load is mirrored in the explanations of the split-attention and expertise reversal effect that will be presented next. Note, however, that within the CLT framework and literature the so called process of mental integration (that should be only necessary in learning with separated formats) is interpreted as extraneous cognitive load (e.g., Sweller et al., 1998).

2.2.4 Explaining the Split-Attention and Expertise Reversal Effect by CTML

As stated above, the processes of multimedia learning described in Mayer's CTML can be translated in terms of cognitive load types introduced by CLT (Sweller et al., 1998). Hence, the following explanations of the split-attention and its corresponding form of expertise reversal suggested by the CTML will be translated into the corresponding cognitive load types. Parallel to the descriptions of the explanations provided by CLT, the explanation of the split-attention effect is described first. Subsequently, the explanation of how the split-attention effect reverses under the influence of high prior knowledge is provided. The explanations of both effects concentrate on germane cognitive load, and thus, represent the above mentioned focus of CTML on promoting processes assumed to underlie these effects.

2.2.4.1 CTML and the Explanation of the Split-Attention Effect

When Mayer (1989) began to investigate the split-attention effect (or spatial contiguity effect), he had not yet developed the CTML nor distinguished between different types of cognitive load but focused on assimilation processes like guiding attention, fostering internal connections between ideas, and fostering connections between ideas from the material and learners' prior knowledge. Based on these assumptions, he argued that students with a labeled graphic (integrated format) of hydraulic drum breaks should engage in meaningful learning and outperform students with either a separated format of text and picture or with text only, because the labeled instructional format should focus learners' attention on the explanative information and help them to organize the information into a coherent mental representation. Some years later, Mayer et al. (1995) specified the cognitive processes of multimedia learning within the generative theory of textbook design (the precursor to the CTML). Mayer et al. (1995) argued that these constructive processes can be exerted when learners actively held text and image representations in working memory at the same time. They argued that learners with integrated formats engage in these processes, especially in integrating all mental representations, because the high spatial contiguity makes them hold both representations in working memory simultaneously, whereas learners with separated formats process disparate information sources in a more isolated way without engaging in the relevant integration processes. Hence, low-knowledge learners with integrated formats are stimulated to engage in generative processing resulting in germane cognitive load. The more generative processes these learners exert, the higher their germane cognitive load, and the better their learning outcomes. Learners with separated formats do not engage in generative processing, and thus have lower germane cognitive load which, in turn, results in lower learning outcomes (Mayer, 2009). This explanation is well in line with the active-processing assumption. Over the years, Mayer and colleagues (cf. Mayer, 2001, 2005, 2009; Mayer & Moreno, 2003; Moreno & Mayer, 1999) changed the focus of their argumentation by including the CLT argument of increased extraneous cognitive load in learners with separated formats. In addition to visual search processes which should cause extraneous cognitive load according to CLT researchers, Mayer and Moreno (2003) added representational holding as further processes causing extraneous cognitive load. Although Mayer, Moreno, Boire, and Vagge (1999, p. 639) stated that they "intended to vary learners' opportunities for building the referential connections needed for constructivist learning" by varying the extraneous cognitive load by means of the instructional format, it stays obvious that processes like "building referential connections" can or even should be interpreted as germane cognitive load within the CTML. Summarily, according to Mayer and colleagues integrated formats elicited germane cognitive load as described in the CTML, whereas separated formats cause extraneous cognitive load as proposed originally by CLT.

2.2.4.2 CTML and the Explanations of the Expertise Reversal Effect

The finding that high-knowledge learners did not benefit from integrated formats but low-knowledge learners did was explained by Mayer and colleagues in the tradition of the active-processing assumption (Mayer, 2001; Mayer & Gallini, 1990). Hence, similar to the explanation of the split-attention effect, they refer to germane cognitive load to explain the expertise reversal effect. High-knowledge learners use their prior knowledge for compensatory processing. Compensatory processing is interpreted as germane cognitive load. As mentioned above, Kalyuga et al. (1998) stated that the processing of separated text-picture formats and the processing of low-cohesive texts cause extraneous cognitive load in learners with low prior knowledge, whereas the processing of integrated formats and high-cohesive texts cause extraneous cognitive load in learners with high prior knowledge. However, text comprehension researchers assume a similar compensatory mechanism, as Mayer does for multimedia learning, in high-knowledge learners who benefit from low-cohesive texts but suffer from high-cohesive texts (McNamara et al., 1996). The assumptions of the compensatory mechanism will be first outlined with regard to multimedia learning and the CTML and subsequently with regard to text comprehension and Kintsch's Construction-Integration Model (CIM; Kintsch, 1988; 1998; McNamara, 2009). Whereas the CTML describes processes during multimedia learning, the CIM describes processes during text comprehension only. Despite this difference, both frameworks have in common that they are more precise than CLT in describing constructive cognitive processes during learning with specific materials.

Multimedia. Mayer and colleagues argue (e.g., Mayer & Gallini, 1990; Mayer et al., 1995) that high-knowledge learners do not suffer from separated formats, because they are able to apply imagery strategies while reading the text information, and thus, do not depend on pictorial information (cf. Alexander & Judy, 1988). Hence, learners' isolated processing approach to separated formats (text reading or picture reading without integrating information) does not prevent high-knowledge learners from exerting generative processes or germane cognitive load, because they are able to integrate word-based and imagined picture-based information with prior knowledge. Such active learning processes involved in imagery strategies help high-knowledge learners to focus on relevant information, so that specific illustrations are not necessary to ensure successful learning. Mayer (2001) states that this approach to the CTML is "based on the idea that high-knowledge compensates for poor instructions" (p. 167). Hence, according to the CTML high-knowledge learners compensate for a lack in instructional guidance by being able to use their prior knowledge, whereas low-knowledge learners cannot compensate poor instructional designs. To sum up, high-knowledge learners with separated formats engage in high germane cognitive load, and thus, can benefit from separated formats, whereas low-knowledge learners are not able to engage in germane cognitive load because they lack the necessary prior knowledge. According to Mayer, one can assume that high-knowledge learners do not switch very frequently between text and corresponding pictorial information to build a coherent mental representation. Rather, they focus on textual information and actively use their domain knowledge that already includes pictorial information.

Text comprehension. The explanation with regard to learners with high prior knowledge during multimedia learning is in line with the assumptions in text comprehension research proposed by McNamara and colleagues (McNamara & Kintsch, 1996; McNamara et al., 1996). In a series of studies they demonstrated that low-knowledge learners benefited from high-cohesive texts (e.g., many anaphoric referents, sentence connectives, background information, as well as meaningful headings and paragraphs), whereas high-knowledge learners benefited from low-cohesive texts. Analogue to Mayer, McNamara and colleagues argue that low-cohesive texts force high-knowledge learners to engage in compensatory processing to infer unstated relations in these texts, whereas high-cohesive texts seduce high-knowledge learners to pursue a more passive processing instead of activating relevant prior knowledge. Compensatory processing in text comprehension can be interpreted as imposing germane cognitive load. Because low-knowledge learners do not possess relevant prior knowledge to compensate low-cohesive texts, they need high cohesive texts that guide them to engage in relevant and explicit information processing, again related to germane cognitive load, to learn successfully.

2.3 Summary of Cognitive Load Explanations

In explaining the split-attention and its respective expertise reversal effect CLT focuses on extraneous cognitive load to explain why less knowledgeable learners suffer from separated and more knowledgeable learners suffer from integrated text-picture formats (inhibiting mechanisms), whereas the CTML focuses on the role of germane cognitive load imposed by processes that are relevant for successful multimedia learning (promoting mechanisms). The crucial differences concerning the suggested cognitive load mechanisms of both effects will be briefly summarized against the background of both frameworks.

2.3.1 Complementary Explanations of the Split-Attention Effect

The above mentioned different foci on extraneous and germane cognitive load within the CTML and CLT resulted in two explanations of the split-attention effect. Whereas CLT emphasizes an extraneous cognitive load mechanism to explain low learning outcomes of low knowledgeable learners with separated formats due to visual search and unnecessarily difficult mental integration processes of physically disparate information, the CTML stresses a germane cognitive load mechanism to explain high learning outcomes of low knowledgeable learners with integrated formats. This mechanism is based on a facilitation of holding word-based and image-based representations simultaneously in working memory supporting the active construction of an integrated model. Notably, CLT specifies not only visual search but also mental integration of information as extraneous cognitive load. If information sources are presented close to each other, they are physically integrated, and according to CLT, there is no need for further resource intensive mental integration processes. The CTML, however, specifies the mental integration of different mental representations as germane cognitive load, no matter whether information sources are presented close to or far away from each other. In both cases, learners have to mentally integrate word-based and picture-based representations to actively construct an integrated model. If information sources are presented close to each other, the probability is high that learners hold both representations in mind and integrate them. If information sources are presented far away from each other, the probability is low that learners hold both representations in mind simultaneously and mentally integrate them. Holding the information of one source in mind until the information of a disparate source is found during visual search (extraneous cognitive load) and generative mental integration processes (germane cognitive load) can be regarded as complementary mechanisms underlying the split-attention effect. That is, the less extraneous cognitive load is caused during learning, the more working memory resources can be used for germane cognitive load. Hence, when combining both explanations one can assume that the split-attention effect is mediated through extraneous as well as germane cognitive load.

Figure 4 and 5 show flowcharts in which the mediating roles of the three cognitive load types are depicted. Figure 4 depicts the extraneous cognitive load explanation. The extraneous cognitive load mechanism is represented by solid arrows or paths. In a first step, the instructional formats influence extraneous cognitive load (solid path) due to spatial contiguity (separated or integrated). In a second step, extraneous cognitive load influences learning outcomes negatively (solid path). Assuming that intrinsic cognitive load is the same in both formats (same content) and that germane cognitive load does not play a role, no additional influences are assumed to be exerted by intrinsic and germane cognitive load (dashed paths).

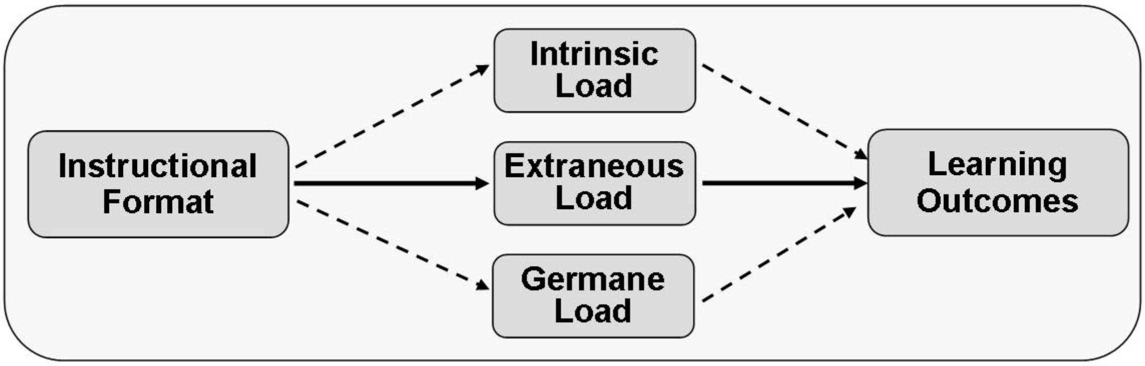


Figure 4. Flowchart of the extraneous cognitive load mechanism explaining the split-attention effect

Figure 5 depicts the germane cognitive load explanation. The germane cognitive load mechanism is represented by solid arrows or paths. In first steps, the instructional formats influence extraneous as well as germane cognitive load (solid paths) due to spatial contiguity (separated or integrated). In a second step, extraneous cognitive load influences learning outcomes negatively (solid path), whereas germane cognitive load influences learning outcomes positively (solid path). Assuming that intrinsic cognitive load is the same in both formats (same content), no special influence is assumed to be exerted by intrinsic cognitive load (dashed path).

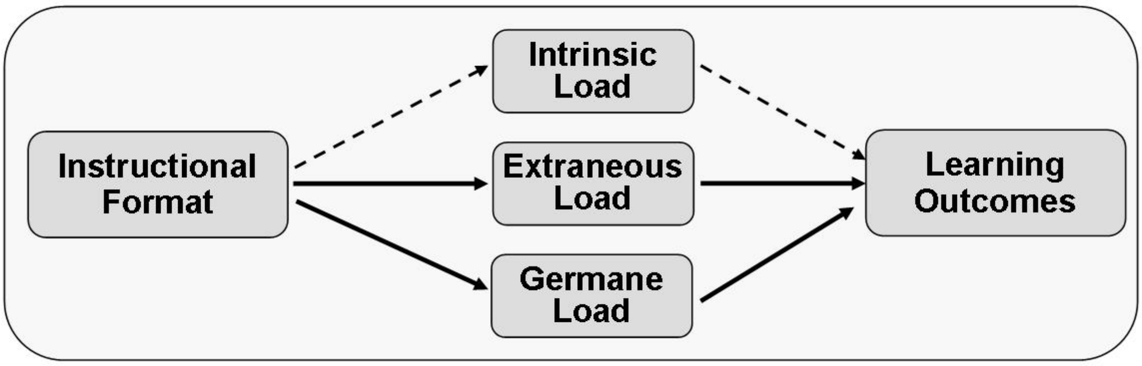


Figure 5. Flowchart of the germane cognitive load mechanism explaining the split-attention effect

To demonstrate whether empirical research supports the extraneous cognitive load explanation only or whether the germane cognitive load explanation has also been supported so far, the existing split-attention literature will be reviewed in chapter 4 and analyzed with respect to the empirical evidence in favor of both cognitive load explanations.

2.3.2 Contradictory Explanations of the Expertise Reversal Effect

The CTML and CLT have become more and more similar in explaining the split-attention effect by including the extraneous cognitive load explanation into CTML assumptions and the germane cognitive load explanation into CLT assumptions. However, the frameworks assume apparently contradictory mechanisms in explaining the expertise reversal effect. Within the CTML Mayer and Gallini (1990) argue that high-knowledge learners with separated formats are cognitively loaded, because these learners actively use their prior knowledge during reading the text to compensate for information not included in the text but in the illustration only. According to this explanation high-knowledge learners with separated formats should focus on the text and experience high germane cognitive load while constructing elaborated integrated models, whereas high-knowledge learners with integrated formats do not need to actively use their prior knowledge, and thus, do not experience high germane cognitive load, but therefore, do not construct elaborated integrated models. This germane cognitive load explanation derived by the CTML (and CIM) has been challenged by CLT researchers, especially by Slava Kalyuga, since the late 1990s (e.g., Kalyuga et al., 1998; Yeung, Jin & Sweller, 1997). In a paper on the expertise reversal effect concerning worked examples and problem solving, Kalyuga, Chandler, Tuovinen, and Sweller (2001, p. 580) stated that:

> Mayer (2001) refers to this finding as the "individual differences principle" and notes that high-knowledge learners tend to use their prior knowledge to compensate for poor instructional formats. Other evidence more closely supports the hypothesis that more knowledgeable learners may benefit more from problem solving than from worked examples because of redundancy.

Moreover, in a paper about research on the expertise reversal effect concerning multimedia instructions Kalyuga et al. (1998, p. 15) stated that:

> Although the findings of the present studies and those of McNamara et al. (1996) are similar, our interpretations are quite different. We suggest that eliminating material for high-knowledge learners is advantageous because it reduces the cognitive load associated with processing redundant information, whereas McNamara et al. suggested it is advantageous because it forces more active processing.

Kalyuga and colleagues assume that high-knowledge learners with integrated text-picture formats do not need verbal explanations to learn from illustrations. Rather, these learners are overloaded by extraneous cognitive load caused by unnecessarily cross-referencing redundant verbal information with intelligible pictorial information and with prior knowledge. Kalyuga et al. (1998) claim that an illustration should be enough to learn from for high-knowledge learners. Accordingly, separated formats should be better for high-knowledge learners, because they allow ignoring the text, whereas ignoring text is not possible in learning with integrated formats. Thus, high-knowledge learners with separated formats should focus on the illustration and should be cognitively low loaded, whereas high-knowledge learners with the integrated format should switch more often between text and illustration and should suffer from extraneous cognitive load.

The assumingly contradictory explanations are shown in Figure 6 and 7 depicting mediated moderations in flowcharts. According to the germane cognitive load explanation (Figure 6), prior knowledge moderates the influence of instructional format on germane cognitive load (solid path). This moderation effect is mediated through the effect of germane cognitive load on learning outcomes (solid path). Because no additional influences are assumed in this explanation, the remaining paths are dashed.

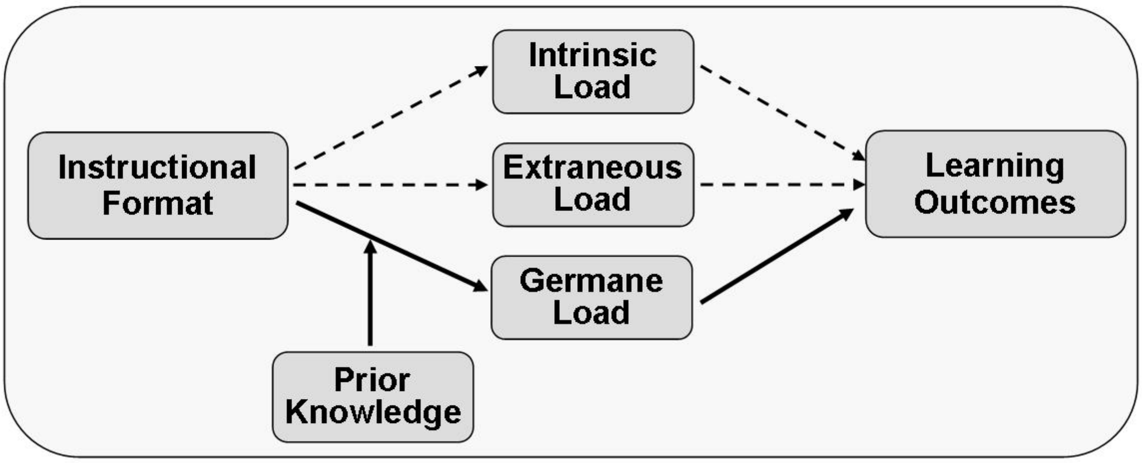


Figure 6. Flowchart of the germane cognitive load explanation for the expertise reversal effect

According to the extraneous cognitive load explanation (Figure 7), prior knowledge moderates the influence of instructional format on extraneous cognitive load (solid path). This moderation effect is mediated through the effect of extraneous cognitive load on learning outcomes (solid path). Again, no further assumptions are made (dashed paths).

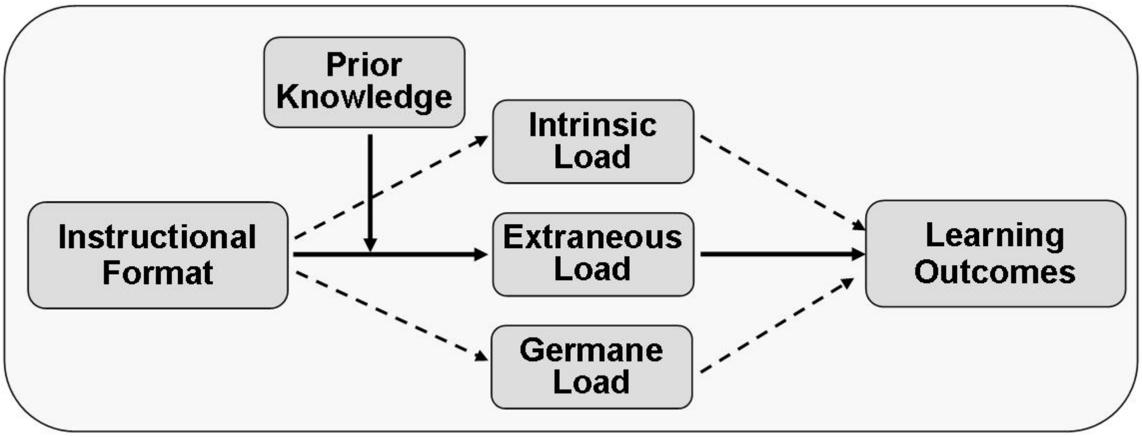


Figure 7. Flowchart of the extraneous cognitive load explanation for the expertise reversal effect

Although Kalyuga et al. (1998) assume that both explanations are contradictory, it is possible that both mechanisms work in parallel. That is, high-knowledge learners with integrated formats might not only be rather passive but might suffer from redundancy and extraneous cognitive load, whereas high-knowledge learners with separated formats might benefit not only from reduced extraneous cognitive load but also from compensatory processing and germane cognitive load. If both mechanisms work in parallel, there need not to be differences in overall cognitive load between high-knowledge learners. To demonstrate whether only one or both of the cognitive load explanations do apply, a literature review of expertise reversal effect studies with respect to the empirical evidence in favor of both cognitive load explanations will be presented in Chapter 6. However, in order to interpret the empirical evidence of the different cognitive load explanations of the expertise reversal effect and the split-attention effect competently, a deeper understanding of the measurement of cognitive load is necessary, in particular with respect to the distinction between the three cognitive load types. Hence, the measurement of cognitive load will be elaborated in the next chapter.

3 Cognitive Load Measurement

The measurement of cognitive load is an important issue when deciding on the cognitive load explanations of the split-attention and expertise reversal effect. Before specific cognitive load measures are described more closely, several general measurement aspects are briefly introduced.

3.1 General Measurement Aspects

In accordance with other models of cognitive load, the triarchic model of cognitive load defines cognitive load as the mental work that results from information processing in working memory (Hogg, 2007). Hence, cognitive load is a complex function of task demands and learner characteristics like working memory capacity and learners' cognitive reactions to the task demands (e.g., Kahneman, 1973; Reed, Burton, & Kelly, 1985; Wickens, 1991). Contrary to other models, the triarchic model specifically refers to learning (instead of task performance) and distinguishes three load types resulting from different sources (instructional content (intrinsic cognitive load) and instructional design (extraneous and germane cognitive load)) and differing in effects on knowledge acquisition (inhibiting (extraneous cognitive load and too high intrinsic cognitive load) versus promoting (germane cognitive load)). Concerning the measurement of cognitive load in general, two aspects can be distinguished: *what* should be measured and *how* should it be measured. The what-aspect surely concerns the three load types, but it also concerns the dimension of time as well as the distinction of task demands and learners' reactions to task demands. The how-aspect concerns methodological techniques that also differ with respect to what they can measure. Before specific methodological techniques used in CLT research are introduced in greater detail, some general distinctions and classifications concerning the above mentioned aspects are presented.

By referring to the temporal dimension of cognitive load expended during performing a task, Xie and Salvendy (2000) distinguish among instantaneous workload (measurement of cognitive load at any point in time), average workload (mean degree or intensity of cognitive load expended during learning), accumulated workload (the total amount of cognitive load expended during learning), peak workload (the maximum of cognitive load expended during learning), and overall workload (learners' subjective experience of cognitive load). Whereas instantaneous and peak workloads demonstrate the dynamic character of cognitive load during a complex process like learning, the other measures represent cognitive load as a more or less static construct. Measurement techniques differ in how much dynamism of cognitive load they can represent.

Another distinction is made by Paas and Van Merriënboer (1994). They distinguish between two cognitive load aspects: mental load and mental effort. Mental load is defined as expected cognitive capacity demands that can be considered an a priori estimate of the cognitive load. Analytical methods like task analysis, mathematical models, or experts' opinions can be used to determine mental load (Cook, Zheng, & Blaz, 2009). The early work in determining cognitive load by CLT used such analytical methods (cf. Sweller, 1988). Although an estimation of the task demands' complexity is important, the mediating role of cognitive load is assumed to go back to the learner's mental effort. Mental effort is defined as the cognitive load that refers to the working memory capacity that is actually expended during learning or performing a task. The three different cognitive load types as well as the different types of workload suggested by Xie and Salvendy (2000) relate to this aspect. That is, how much capacity does the learner really expend during learning for which type of cognitive load? To test the assumptions postulated by CLT and the CTML to explain instructional design effects, it is important to measure cognitive load with techniques that allow capturing the mental effort caused by the three cognitive load types separately. However, as will become obvious in the next sections measuring the three cognitive load types separately is the primary challenge in testing these assumptions.

In their overview of methodological techniques, Brünken, Plass, and Leutner (2003) classify the techniques for measuring cognitive load along the two dimensions objectivity (objective vs. subjective) and causal relationship (direct vs. indirect). Table 2 presents a selection of the measures included in the classification scheme by Brünken et al. (2003). This classification will be discussed in more detail in the following because it suggests definite characteristics of the cognitive load measures classified that do not exist in the way suggested by Brünken and colleagues.

Table 2

Classification of Cognitive Load Measures (cf. Brünken et al., 2003)

Causal Relationship	Objectivity	
	Objective	**Subjective**
Direct	Secondary task performance	Perceived difficulty ratings
Indirect	Learning outcome measures, Behavioral measures (e.g., studying time, viewing behavior)	Mental effort ratings

According to Brünken et al. (2003) direct measures of cognitive load are secondary task performance and perceived difficulty. Whereas secondary task performance can be measured objectively by measuring for example persons' reaction time to a stimulus during learning, ratings of perceived difficulty are assumed to reflect learners' subjective impression of the intrinsic and extraneous cognitive load expended during a task.

According to the classification scheme, indirect measures of cognitive load are behavioral measures like learning outcomes, studying time, viewing behavior, and self-reported ratings of invested mental effort. These measures are thought to vary according to changes in cognitive load but not to represent cognitive load directly. However, why should a behavioral activity like learners' reaction times in a secondary task be more direct than their viewing behavior during learning? Moreover, Brünken et al. (2003) as well as Paas and Van Merriënboer (1994) categorize learning outcomes as an indirect measure of cognitive load. However, learning outcomes are the result of the pattern of cognitive load types during learning – not the mediating factor. Thus, by declaring learning outcomes as a cognitive load measure the authors mix cause (cognitive load) and effect (learning outcomes) thereby contributing to a circular definition of cognitive load that has been criticized by several researchers (cf. de Jong, 2010; Gerjets et al., 2009). This imprecision makes other researchers like Cook et al. (2009) criticize the classification scheme as rather arbitrary. Despite these objections to the value of the scheme, it contains not only techniques often used but also those who are not yet very well established. Studying times, for example, as indirect measure can be measured easily and objectively. However, what learners exactly do during this time mostly remains unclear. Nevertheless, studying times have often been used as empirical evidence to support different cognitive load argumentations. In contrast, viewing behavior as behavioral process measure can also be objectively measured by using the eye tracking methodology but nevertheless has not been often used to measure cognitive load in multimedia research. In contrast to studying times that tell relatively little about what learners are doing, however, viewing behavior is a process measure that reveals in detail when and for how long learners are looking at different information sources. Despite this information richness it has not been clarified so far which measures of viewing behavior are related to which cognitive load types.

Another indirect measure is the subjectively reported amount of invested mental effort. According to Brünken et al. (2003) mental effort reflects learners' impression of how much load they invested but without exactly knowing what type of cognitive load the effort was used for. Brünken et al. (2003) argue that perceived difficulty ratings are a more direct measure of intrinsic and extraneous cognitive load, whereas mental effort ratings might measure total cognitive load, and therefore can or should also include germane cognitive load. However, not all researchers share this assumption. For example, DeLeeuw and Mayer (2008) even claim that mental effort ratings are most appropriate to measure intrinsic cognitive load, whereas perceived difficulty is claimed to be most appropriate to measure germane cognitive load, and secondary task performance should be most appropriate to measure extraneous cognitive load. Such different interpretations of the same cognitive load measures have severe effects when evaluating the empirical evidence in favor of different cognitive load explanations of instructional design effects.

3.2 Specific Cognitive Load Measures

To be able to evaluate the validity of cognitive load measures, and thereby the empirical evidence against or in favor of suggested cognitive load explanations, the specific strengths and weaknesses of the most common cognitive load measures – namely (1) studying times, (2) secondary task performance, (3) subjective ratings, and (4) behavioral activities of learning (e.g., viewing behavior) – will be discussed in the next sections. This discussion should help to clarify how different cognitive load measures might be best interpreted and which cognitive load measures seem to be most promising as measures of different cognitive load types.

3.2.1 Studying Times

Although studying time has been used very often in CLT experiments as cognitive load measure (see review on the explanation of the split-attention effect in Chapter 4), its value as a measure is very limited, because studying times are not based on a well-defined rationale. Rather, there are at least two basic possibilities how studying times can be interpreted within the triarchic model of cognitive load. Besides this qualitative ambiguity, there are also limitations in using it as a quantitative cognitive load measure of working memory demands.

3.2.1.1 Qualitative ambiguity

There are two typical interpretations with regard to which type of cognitive load is represented by studying times. To decide whether studying times measure extraneous or germane cognitive load one usually has to rely on learning outcomes. On the one hand, when learners have low learning outcomes despite long studying times, one can assume that studying times represent extraneous cognitive load. That is, learners do lots of unnecessary processing that inhibits knowledge acquisition. In line with this rationale, Sweller, Chandler, Tierney, and Cooper (1990) argue that learners suffer from separated formats, because they suffer from high extraneous cognitive load which was reflected in longer studying times compared to learners with integrated formats. On the other hand, when learners have high learning outcomes with long studying times, one can argue that studying times represent germane cognitive load. That is, learners exert important processing that supports knowledge acquisition. In line with this rationale, McNamara and Kintsch (1996) argue that high-knowledge learners benefit from incohesive texts, because they engage in active processing or germane cognitive load which was reflected in longer studying times compared to high-knowledge learners with cohesive texts. Moreover, although Martin-Michiellot and Mendelsohn (2000) found longer studying times of learners with the separated format, they did not find the split-attention effect on learning outcomes. In this case, one might infer that these students used longer studying times to compensate the demands caused by the separated format. If so, the definition of extraneous cognitive load would not fit properly to longer studying times and the question arises whether such successful compensation processing should not be better defined as germane cognitive load. The possibility that studying times might be used for germane cognitive load is also favored by Rose and Wolfe (2000). These authors showed that especially inexperienced learners had higher learning outcomes, when they invested more time in studying. Thus, whether learners expend high extraneous or high germane cognitive load is not determined by time per se but can only be decided in the light of learning outcomes. Only in relation to learning outcomes as the result of the learning process, one can interpret whether long studying times represent either germane or extraneous cognitive load (under the condition that intrinsic cognitive load is held constant). Furthermore, as studying times always include intrinsic cognitive load one can assume that there is a positive correlation between intrinsic cognitive load and studying times. Thus, if intrinsic cognitive load also varies between instructional conditions, studying times might not be usable to measure extraneous or germane cognitive load.

3.2.1.2 Quantitative ambiguity

Besides the aforementioned ambiguity in qualitative respects, there are also two objections why studying times are difficult to use as a quantitative measure of intrinsic, extraneous or germane cognitive load. First, learners might seem to study intensively because they are sitting in front of the learning materials for a long time. However, it is possible that they think about different issues and do not engage in knowledge acquisition at all, and therefore, do not succeed in learning. In this case, it would be misleading to infer that learners were overloaded due to high extraneous cognitive load just because they showed long studying times and low learning outcomes. Second, if learners are allowed to study as long as they want to, they might be tempted to reduce the processing speed, that is, they might also reduce their instantaneous cognitive load (Xie & Salvendy, 2000). Hence, although learners study for a long time, they do not use their complete working memory capacity, and thus, do not overload their working memory (Kerr, 1973). With such a strategy they might study longer without that being an indication of a high level of cognitive. Thus, these learners do not experience a high level of cognitive load. To sum up, studying times are a cognitive load measure that is very difficult to interpret with regard to the three cognitive load types. Accordingly, the interpretation of studying times as cognitive load measure is subject to severe limitations so that this measure should only be used carefully.

3.2.2 Secondary Task Performance

Measuring cognitive load via secondary task performance is based on the dual-task paradigm. Thus, the dual-task paradigm and implementation characteristics that need to be controlled when using this paradigm are explained in the following two sections. Subsequently, studies on instructional design principles that used secondary task performance as cognitive load measure are critically discussed.

3.2.2.1 General rationale

As mentioned above, studying times are a rather critical measure with regard to their interpretability in terms of the load imposed onto a limited working memory capacity. For instance, studying times do not reveal whether learners really use their full working memory capacity or not (Kerr, 1973). The dual-task paradigm provides a general means for separating processing or studying times from demands on working memory capacity. This may be the reason why Brünken et al. (2003) describe it as a direct measure of cognitive load. The paradigm requires persons to perform two tasks simultaneously: A primary and a secondary task. The basic assumption of the paradigm is that two tasks sharing the same limited cognitive resources in working memory may interfere with each other, thereby reducing the performance of the secondary task (Baddeley, 1996; Barrouilet et al., 2007; Bourke, Duncan, & Nimmo-Smith, 1996; Eysenck, 1982). Welch (1898), one of the pioneers of this method, showed that the performance of a physical hand grip task was impaired when persons engaged in mental activities like calculation or reading. Her argument was that one could use the degree of interference of two processes in working memory represented by the impairment of the secondary task (e.g., hand grip) as an index of how much cognitive capacity is needed by a primary task (e.g., calculation). The more instantaneous cognitive capacity a primary task needs, the worse gets the secondary task performance. Reaction times to a visual or auditory stimulus have been often used as secondary task (Kahneman, 1973; Kerr, 1973). The increase in reaction times above baseline performance or the comparison of reaction times related to two different primary tasks are used to indicate the relative cognitive demands of the primary tasks (e.g., Darley, Klatzky, & Atkinson, 1972). Secondary task performance (e.g., reaction times or error rates) is a measure of overall cognitive load that can be obtained more or less instantaneously but that does not allow to differentiate between the three different load types. In order to use secondary task performance as overall cognitive load measure, several implementation demands must be met.

3.2.2.2 Implementation demands

Implementing the dual task paradigm requires that researchers are particularly attentive with regard to two criteria: Intrusion in primary task and level of interference. Because of these two criteria Howe and Rabinowitz (1989) are rather critical with regard to secondary task performance and argue that if the two preconditions are violated, any interpretations of secondary task performance are impossible.

The first precondition in interpreting secondary task performance as cognitive load measure is that the secondary task is designed in such a way that it does not affect the primary task performance (Fisk, Derrick, & Schneider, 1986; Kerr, 1973). Therefore, the secondary task must be simple enough and students should be aware that they are to respond to the second task without allowing the response to interfere with the primary task (e.g., learning). Whether performing a secondary task impaired learning can be tested by comparing learning outcomes of students with and without secondary task. The learning outcomes should not differ between the groups. Whelan (2007) argues that complex tasks like solving mathematical problems are less suitable as dual tasks, whereas monitoring tasks to which learners must respond as quickly as possible (e.g., reaction times) seem to be more suitable tasks.

The second precondition is that the interference between primary and secondary task occurs on the cognitive level but not on the perceptual or response level. Whereas the first type of interference is termed capacity interference, the latter two types are termed structural interference (Kahneman, 1973; Kerr, 1973). That is, processing structures needed for the perception of the secondary task or for the response task must be distinct from those structures needed for the primary task, namely learning. For example, persons cannot perceptually process two widely separated visual signals at the same time because of the limitations of our visual system. Hence, if the primary task requires a substantial amount of visual processing, the secondary task should be designed in such a way that the secondary task stimulus can nevertheless be perceived easily. Otherwise the secondary task performance might be impaired due to perceptual limitations that affect the perceptual processing of the secondary task stimulus and not due to the cognitive processing demands imposed by the primary task. The following section provides an overview on how secondary tasks were implemented with respect to these implementation demands in CLT research.

3.2.2.3 Secondary task implementations in CLT research

Brünken et al. (2003) emphasize that secondary task performance is a promising method for CLT, because it is a direct measure of cognitive load. Moreover, Paas, Tuovinen, Tabbers, and Van Gerven (2003) claim that it is a highly sensitive and reliable technique. Despite these advantages, it has not been used very often in multimedia or CLT research. Exceptions can be found in studies by Brünken, Steinbacher, Plass, and Leutner (2002), Brünken, Plass, and Leutner (2004), Chandler and Sweller (1996), DeLeeuw and Mayer (2008), Marcus, Cooper, and Sweller (1996), Sweller (1988), Van Gerven, Paas, Van Merriënboer, and Schmidt (2002) as well as Whelan (2006). The rare use of the method may be attributed to the aforementioned preconditions that have to be met in order to obtain a valid measurement. Three of the aforementioned studies will be discussed in this section to illustrate exemplarily how the difficulties with the secondary task method resulting from these preconditions may challenge the interpretability of these data.

In investigating the split-attention effect, Chandler and Sweller (1996) used a rather difficult secondary task. They asked three groups of students during problem solving to encode letters under time pressure and to remember them some time later. The students had to solve programming problems by either studying with a conventional separated manual or a manual in integrated format, or both manuals plus a computer (split-attention conditions), or they had to study with an integrated manual without computer (fully integrated condition). The secondary task stimuli, the letters that were to remember, were presented on an additional computer screen. Whenever a tone appeared a letter was shortly presented on this computer screen and the students had to encode this letter and to recall the letter previously presented. The study demonstrated that learners with a split-attention format (manuals plus computer) not only solved less complex programming problems but also made more errors in recalling letters than the group without a computer. The findings of the learners in the split-attention conditions were interpreted as clear evidence that these learners were cognitively more loaded by extraneous cognitive load than learners with an integrated manual only. However, how valid are these results in the light of the two aforementioned criteria? First, a replication of the study by Martin-Michiellot and Mendelsohn (2000) provides interesting information concerning learning intrusion as “every trial learner was unable to do even the most basic primary tasks and reported to be highly disturbed and even annoyed by the experiment” (p. 289). Thus, to not disturb students’ learning (primary task) too much, the authors gave up using the secondary task. Although Martin-Michiellot and Mendelsohn (2000) did not use the secondary task any longer, because it seemed to intrude learning, neither they nor Chandler and Sweller (1996) themselves tested whether the secondary task affected learning outcomes by using a control condition. Second, concerning capacity and structural interference, one has to consider that both groups that learned with a manual and a computer had to visually processes three information sources (manual, computer screen to learn with, and computer screen with the secondary task stimuli), whereas the integrated format group only had to visually process two sources (manual and computer screen with the secondary task stimuli). Thus, according to Howe and Rabinowitz (1989) the secondary task results of the study by Chandler and Sweller (1996) are not interpretable, because both criteria of designing secondary tasks seemed to be violated. Similar interpretational problems can be found in the study by Marcus et al. (1996).

Brünken et al. (2002) as well as DeLeeuw and Mayer (2008) used easier secondary tasks to avoid intrusion into learning. They simply used learners’ response times to visual secondary task stimuli presented within the learning materials to measure learners’ cognitive load during learning with a computer-based learning environment. Learners only had to press the space bar as fast as possible when the color of the respective visual stimuli changed. In the Brünken et al. (2002) study, learners had to respond to the color change of a letter placed on the upper part of the screen, whereas in the DeLeeuw and Mayer (2008) study, learners had to react to the color change of the whole background of the materials (the background of the text space, however, did not change its color). This secondary task seems to be easy enough to not disturb learning compared to the one used in the study by Chandler and Sweller (1996). However, this assumption was also not tested explicitly. Moreover, in both studies learners had either to learn with a picture and narration or with a picture and visual text. Learners with picture and narration not only learned better but also responded more quickly to the color changes than learners with picture and visual text. The authors infer from these findings that learners with picture and visual text suffered from higher extraneous cognitive load than learners with picture and narration. Unfortunately, it is also possible that the differences in secondary task performance were caused by structural interference, because in both studies learners with slower reaction times also had to process more visual information (they had to inspect pictures and read text) than learners with better secondary task performance (they had to inspect pictures only). The fact that the condition with the better learning outcomes also had less visual stimuli to process (interference on a perceptual level) makes it difficult to interpret the secondary task performance as a pure cognitive load measure.

The above presented examples show how difficult it might be to design an appropriate secondary task. But even if the challenge is met to generate a secondary task that fulfills both conditions, secondary task performance is a measure of overall cognitive load that does not provide information about the level of intrinsic, extraneous, and germane cognitive load separately. When learners perform worse in a secondary task than others, this can be the result of higher intrinsic or higher extraneous but also of higher germane cognitive load. The type of cognitive load cannot be inferred from secondary task performance, but only the overall amount of cognitive load as compared to another condition. And if no differences are found in secondary task performance (e.g., there were no differences between the factors of instructional format in the study by Van Gerven et al., 2002), the result may be caused by different patterns of the three cognitive load types. Whereas extraneous cognitive load may increase in one condition, germane cognitive load may increase in the other condition. In effect the total amount of cognitive load is the same in both conditions, however, in the first condition there is more extraneous cognitive load besides intrinsic and germane cognitive load, whereas in the second condition there is more germane cognitive load besides intrinsic and extraneous cognitive load. Secondary task performance itself provides no information about the mixture of different cognitive load types. To demonstrate different patterns of cognitive load types, methods are needed that measure intrinsic, extraneous, and germane cognitive load individually. Currently, subjective ratings seem to offer a potential way to assess the different load types separately. Thus, several subjective rating scales are discussed in the next sections.

3.2.3 Subjective Ratings

Subjective ratings are one of the most often applied techniques to measure cognitive load in instructional design research. Several scales, unidimensional and multidimensional, have been developed and have been used to measure cognitive load. However, the scales applied so far were rarely used to differentiate between the three load types carefully. Nevertheless, contrary to secondary task performance subjective ratings have the potential to differentiate between cognitive load types. Before the specific characteristics of different scales used for cognitive load measurement are outlined in detail, the general rationale of subjective ratings scales will be introduced.

3.2.3.1 General rationale

The basic assumption that provides the basis of subjective measures is that persons are able to introspect and retrospect certain characteristics of their cognitive processes. Several studies on subjective ratings scales demonstrated that subjective ratings can represent a person’s amount of processing resources that he/she invested to meet the task demands (e.g., Borg, Bratfish, & Dornic, 1971; Gopher & Braune, 1984; Gopher, Chillag, & Arzi, 1985). Moreover, some researchers even suggested that the type of scale used is not critical, that is, the choice of category scales, magnitude estimation, and the presence or absence of verbal labels should make little difference (Borg, 1978; Borg, Bratfish, & Dornic, 1971; Paas, 1992). The advantages of subjective rating scales are their high face validity, the fact that they are easy to administer, and that they do not disrupt the learning process, if they are administered afterwards. Concerning cognitive load types, it can be concluded, however, that these studies varying task complexity only showed that differences in intrinsic cognitive load were reflected in subjective ratings. Because the instructional design of presenting materials was not varied, the other cognitive load types were not addressed in these studies. Thus, with regard to the measurement of the three cognitive load types, some researchers doubt that learners are able to intro- or better retrospect on the distinct cognitive processes caused by the three cognitive load types (e.g., Schnotz & Kürschner, 2007). Another critique is that subjective rating scales cannot measure instantaneous load (see Xie & Salvendy, 2000), at least if they should remain non-intrusive. Otherwise one would need to ask persons permanently or at least very often during learning to report about their cognitive load levels. This would obviously interrupt a normal learning process. Moreover, subjective measures may be subject to social desirability bias (Nunnally, 1978) so that learners might report higher or lower level of cognitive load than they actually experienced. The following sections introduce some of the unidimensional and multidimensional scales that have been used in CLT research.

3.2.3.2 Unidimensional scales

The usage of subjective ratings in CLT research can be traced back to the work of Paas (1992) who developed a unidimensional scale to measure cognitive load by asking learners how much *mental effort* they invested in solving a problem (in the sense of mental effort in general). This scale is similar to the one developed by Zijlstra and Van Doorn (1985). Subsequently, Kalyuga, Chandler, and Sweller, (1998) introduced a different unidimensional scale asking learners to rate their *perceived difficulty*. This scale is based on a difficulty scale used by Borg et al. (1971). As already discussed above, CLT emphasizes the extraneous cognitive load aspects of instructional designs. Hence, it is not astonishing that researchers emphasized the measurement of difficulty as is also exemplified by the titels of two papers: “Why some material is difficult to learn?” by Sweller and Chandler (1994) and “Cognitive load theory, learning difficulty, and instructional design” by Sweller (1994). In line with this focus some CLT researchers asked their participants to rate their perceived difficulty, whereas other CLT researchers asked their participants to rate their mental effort to measure cognitive load. Thus, although all CL researchers share the view that mental effort is the amount of cognitive load invested during learning, they have used different scales. Moreover, researchers from other theoretical backgrounds like Salomon (1984) defined the construct of mental effort more specific and used the subjective ratings of the level of concentration to measure it. It can be assumed, however, that asking whether persons to rate their mental effort, their perceived difficulty, or their level of concentration might make a difference in what is actually measured with respect to the three cognitive load types.

3.2.3.2.1 Mental effort ratings

According to Paas (1992) and Brünken et al. (2003) the question “How much mental effort did you invest?” does not aim specifically at relevant or irrelevant learning processes, but rather comprises all processes expended during learning. Therefore, the mental effort scale can be said to measure overall cognitive load that comprises intrinsic, extraneous, and germane cognitive load (Brünken et al., 2003). Notably, in the first study on invested mental effort in CLT research, Paas (1992) did not find any differences on the amount of mental effort reported after each practice problem between learners with conventional problems, worked examples, or completion examples. Nevertheless, the groups differed on transfer test performance – learners with worked and completion examples outperformed the conventional problem solving group in a far-transfer test. Because there were no differences between groups in mental effort ratings during learning but differences in learning outcomes, one may argue that different patterns of cognitive load type were responsible for the different learning outcomes across the groups, although mental effort as an overall cognitive load measure was the same across the groups. Later on, such interpretations were indeed made by researches who did neither find any group differences on the mental effort scale but group differences on learning outcomes (e.g., Kester, Kirschner, & Van Merriënboer, 2005; Wouters, Paas, & Van Merriënboer, 2010). The remaining insecurity, though, about which cognitive load type is exactly measured and about why learners expended mental effort or why not, made Brünken et al. (2003) classify mental effort ratings as indirect measure.

Other researchers like DeLeeuw and Mayer (2008), however, concluded on an empirical basis that mental effort ratings measure intrinsic cognitive load only. In their study, students were asked to learn how an electrical motor works. They studied either with graphic and narration only or with graphic, narration, and visual text (manipulation of extraneous cognitive load due to redundancy of text and narration). During studying the materials, the students were asked eight times to rate the level of mental effort they had invested directly before the question was asked. The question was put four times after sentences with low and four times after sentences with high complexity. The results indicated that the mental effort ratings after low-complexity sentences were lower than after high-complexity sentences (although not significantly). According to this result, one can infer that the mental effort ratings of the study by DeLeeuw and Mayer (2008) indicated variations in intrinsic cognitive load. Because there was no difference found in mental effort ratings between students with the graphic and narration only and students with the graphic, narration, and visual text, although a redundancy effect was hypothesized, the authors argued that the mental effort ratings did not measure extraneous cognitive load. It is, however, questionable how valid this conclusion, based on a non-significant result, is. The authors did not report any differences in learning outcomes between both instructional format groups. Therefore, it remains questionable whether there were any group differences in learning outcomes at all that were caused by extraneous cognitive load. If there were no differences in learning outcomes due to extraneous cognitive load, there should also be no differences in the mental effort ratings between the groups. Noteably, other studies already applied mental effort ratings successfully to demonstrate differences in extraneous cognitive load. For example, there are studies that manipulated the instructional design by comparing worked examples with conventional problems. These studies showed that mental effort ratings differed between groups that were thought to differ in extraneous cognitive load (e.g., Paas & Van Merrienboer, 1993; Tuovinen & Sweller, 1999). In another study on how to present information (piece-by-piece or all at once), the phrasing of the mental effort rating scale was varied. Kester, Kirschner, and Van Merriënboer (2006) asked students after each problem on electrical circuits to rate how much mental effort it required to find a solution. Furthermore, at the end of the learning phase they also asked students to rate how much effort it required to understand the whole subject matter. Whereas there were group differences with regard to the mental effort invested in finding problem solutions, there were no group differences with regard to the mental effort invested in understanding the whole subject matter. Whereas the subject matter was the same for all students, the finding of solution was influenced by the manipulated format of presenting information piece-by-piece or all at once. For this study, one might argue that the amount of mental effort required for understanding aimed at measuring intrinsic or even intrinsic plus germane cognitive load, whereas the mental effort required for finding a problem solution aimed at measuring extraneous cognitive load.

According to the above cited results, a general conclusion that mental effort ratings do always measure intrinsic cognitive load only (cf. DeLeeuw & Mayer, 2008) seems not to be reasonable. Rather, whether mental effort ratings indicate only intrinsic or only extraneous (cf. Paas & Van Merrienboer, 1993; Tuovinen & Sweller, 1999) or all three types of cognitive load (cf. Brünken et al., 2003) depends to a large extent on (1) the specific processes learners should rate (e.g., understanding vs. finding solution), (2) the experimental manipulations of the instructional design, and (3) whether these manipulations indeed influence learning outcomes. Hence, an easy interpretation of reported ratings of mental effort does not seem to be possible.

3.2.3.2.2 Perceived difficulty ratings

In contrast to the amount of reported mental effort, perceived difficulty ratings seem to have higher face validity with regard to intrinsic and extraneous cognitive load, because they obviously aim at inhibiting aspects during learning. Indeed, difficulty measures successfully demonstrated group differences in experiments that carefully manipulated either extraneous cognitive load only or intrinsic cognitive load only. For example, Kalyuga et al. (1998) manipulated the instructional design by changing the spatial contiguity of text and graphic. They demonstrated first, that students with a physically integrated text-graphic format explaining electrical circuits rated the difficulty to understand such electrical circuits lower than students with a separated text-graphic format or students with no text at all. Second, they demonstrated that students with the physically integrated format had better learning outcomes than the other two groups (see the split-attention effect). Because only the instructional format but not the learning content was manipulated, the authors suggested that one can conclude that extraneous cognitive load was reduced by physically integrated text-graphic formats as shown by the low difficulty ratings. Perceived difficulty ratings also differed between groups in other studies that aimed at manipulating extraneous cognitive load, thereby probably indicating extraneous cognitive load (e.g., Kalyuga, Chandler, & Sweller, 2000; Pawley, Ayres, Cooper, & Sweller, 2005). However, there are also studies that aimed at manipulating extraneous cognitive load, but that did not find the suggested differences between groups (e.g., Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

Instead of manipulating extraneous cognitive load, Ayres (2006a) manipulated the intrinsic cognitive load of algebraic bracket-expansion problems by changing the plus and minus signs in the problems. Because a minus sign before a bracket demands further cognitive manipulation steps (a plus inside the bracket becomes a minus and vice versa) by the student, problem parts with minus sign before the bracket have higher intrinsic cognitive load than those without one. The instructional format itself was kept constant. Ayres (2006a) demonstrated that students rated the difficulty of bracket-expansions within a problem higher, if brackets had higher element interactivity (minus sign before it) and demanded more steps than brackets with low element interactivity and less steps (without minus sign). Furthermore, he showed a positive correlation between perceived difficulty and error rates, indicating that bracket-expansions that were rated to be more difficult were solved less successfully than bracket-expansions that were rated less difficult. Other studies aimed at manipulating intrinsic cognitive load by using differently knowledgeable learners. These studies also showed differences between groups on perceived difficulty ratings (e.g., Clarke, Ayres, & Sweller, 2005; Pollock, Chandler, & Sweller, 2002). More knowledgeable learners rated the learning tasks less difficult than less knowledgeable learners. But similar to studies aiming at manipulating extraneous cognitive load, there were also studies that did not find the suggested differences on perceived difficulty (e.g., Kalyuga, Chandler, & Sweller, 2001).

Thus, perceived difficulty ratings seem to be easier to interpret than mental effort ratings. Nevertheless, it still seems to be relevant to know whether the researchers aimed at manipulating intrinsic or extraneous cognitive load to clearly interpret whether perceived difficulty measured intrinsic or extraneous cognitive load. To summarize, depending on the experimental manipulations, perceived difficulty ratings can be used to measure intrinsic or extraneous cognitive load. Whether they can always successfully indicate the assumed cognitive load type, however, is not clear. With these results of many studies in mind, it is quite astonishing that DeLeeuw and Mayer (2008) are the only ones by now who concluded that perceived difficulty should be most appropriate to measure germane cognitive load. Why did they draw this conclusion? DeLeeuw and Mayer (2008) asked students (condition 1: text and diagram; condition 2: narration, text, and diagram) to rate their perceived difficulty after the learning phase. The authors divided their students in one group with high learning outcomes and one group with low learning outcomes by means of a median split, and therefore, independently of the instructional manipulations. They showed that the group with higher learning outcomes had lower difficulty ratings than the group with lower learning outcomes. This result was interpreted by DeLeeuw and Mayer (2008) as indication that perceived difficulty measures germane cognitive load appropriately. However, their conclusion seems to be inappropriate, because two issues complicate their interpretation. First, germane cognitive load should be elicited by the instructional format. Obviously, this was not the case in that study. Rather, the authors had to make a median split to get two groups with different levels of learning outcomes independently of the instructional format. Second, germane cognitive load should be positively related with learning outcomes. However, the difficulty measures were related negatively to learning outcomes. Thus, one has to make an intermediary step in their argumentation and interpret that learners with low difficulty ratings used their remaining working memory capacity for learning relevant processes (germane cognitive load), whereas learners with high difficulty ratings had not enough free working memory capacity, and thus, could not use many working memory resources for learning relevant processes. Because the cognitive load differences between the two groups were not caused by the instructional design format but by a median-split with regard to students learning outcomes, the conclusion of DeLeeuw and Mayer (2008) does not seem to be well substantiated. Rather, perceived difficulty also seems to be a more appropriate measure of intrinsic and/or extraneous cognitive load in their study.

According to the above mentioned studies it can be concluded that perceived difficulty is more appropriate to measure intrinsic and extraneous cognitive load than mental effort ratings are. Nevertheless, the difficulty scales used so far lack specificity in their phrasing with regard to (1) the content complexity (intrinsic cognitive load) or (2) the design of the learning material (extraneous cognitive load). If one wants to measure both intrinsic and extraneous cognitive load in one experiment, the learners must know whether they should rate the difficulty of the learning content or the difficulty of the instructional design. Implementing two scales with specific phrasings may solve these interpretation problems.

3.2.3.2.3 Level of concentration ratings

As outlined above, intrinsic and extraneous cognitive load seem to be measurable independently by subjective difficulty ratings, if the design of the study and the phrasing of the scale is definite enough. But what is about a specific measurement of germane cognitive load? To give an answer and to suggest a possible solution, the construct of mental effort will be reconsidered from a theoretical perspective outside CL research. Notably, mental effort was not only defined as working memory capacity expended during learning by CLT researchers but was already defined by Salomon (1983, 1984) as conscious, non-automated elaborations that support learning. According to Salomon’s definition of mental effort it can be said that the more effort is invested during learning, the better should be the learning outcomes. Thus, Salomon’s construct of invested mental effort is more specific and well in line with the construct of germane cognitive load. Salomon (1984) further claims that learners themselves decide about their amount of invested mental effort based on the perceived instructional task demands. For example, if learners assume that the instructional task demands are high (e.g., reading a text), they invest more mental effort. However, if learners assume that the instructional task demands are rather low (e.g., watching TV), they do not invest much mental effort. Interestingly, to measure the amount of invested mental effort, Salomon (1984) did not ask learners to rate mental effort directly but asked students among other questions to rate their level of concentration during learning. As predicted, the higher the level of concentration was, the better were the learning outcomes. Contrary to the scale asking “How much mental effort did you invest?” that was negatively related with learning outcomes in CLT research, the scale asking “How much did you concentrate while reading (watching TV)?” was positively related with learning outcomes in Salomon’s research. Therefore, one can conclude that the level of concentration was interpreted by the students in the sense of germane cognitive load, that is, learners’ attention was directed to learning relevant processes. Hence, contrary to the term *concentration* that seems to be interpreted by students as something that fosters their knowledge acquisition, the term *effort* (as well as *difficulty*) seems to be interpreted by most students as something that makes learning more difficult. Thus, measuring mental effort by subjective mental effort or difficulty ratings might represent other cognitive load types than subjective ratings of the level of concentration. However, whether for example difficulty ratings measure a different cognitive load type than concentration ratings has not yet been tested in a study in multimedia or CLT research.

3.2.3.3 Multidimensional scales

So far, only unidimensional scales were discussed, although the construct of cognitive load suggested by CLT and the CTML is triarchic. However, there are also studies which used multidimensional scales to measure cognitive load. There are two types of multidimensional scales in the literature on cognitive load measurement. Whereas one type uses a computational approach combining the aforementioned unidimensional scales as one dimension with learning outcomes as the other dimension into efficiency measures, the other type uses multiple rating scales to measure different aspects, types or qualities of cognitive load.

3.2.3.3.1 Efficiency measures

According to the beginnings of CLT, in which reducing overall cognitive load was thought to be the main aim of instructional designs, Paas and Van Merriënboer (1993, 1994) developed a computational approach to estimate instructional efficiency. This approach was based on the knowledge test phase, thereby ignoring the learning phase as the critical phase of cognitive load during knowledge acquisition. In the approach, high learning outcomes associated with low ratings of mental effort during solving the knowledge test (mental test effort) is termed high instructional efficiency, whereas low learning outcomes associated with high ratings of mental test effort is termed low instructional efficiency. To get a comparable measure, relative condition efficiency is defined as the observed relation between the amount of mental test effort and learning outcomes in a particular condition relative to a hypothetical baseline condition, in which each z-score unit of invested mental effort equals one z-score unit of learning outcomes. This can be visualized by a two-dimensional coordinate system (see Figure 8) with mental effort as one dimension (x-axis) performance as the second dimension (y-axis). The hypothetical baseline condition E is a line through the zero point with a slope value of plus one (straight diagonal through zero).

The relative condition efficiency is calculated as the perpendicular distance from a point representing a particular experimental condition (P: z-score learning outcomes / R: z-score test effort) in the coordinate system to the efficiency baseline E, represented by the formula $E = (R - P) / \sqrt{2}$. This measure was thought to inform about the efficiency of instructions. However, mental test effort (effort invested during the test phase) instead of learning effort (effort invested during learning) was suggested originally. Because test effort does not (necessarily) inform about learning effort, no conclusions with regard to cognitive load as a mediator of instructional design effects can be derived from instructional efficiency. Therefore, most of the researchers applying the efficiency measure substituted test effort for learning effort (e.g.,. Camp, Paas, Rikers, & Van Merriënboer, 2001; Carlson, Chandler, & Sweller, 2003). This led to a revised version of the original efficiency formula (Van Gog & Paas, 2008).

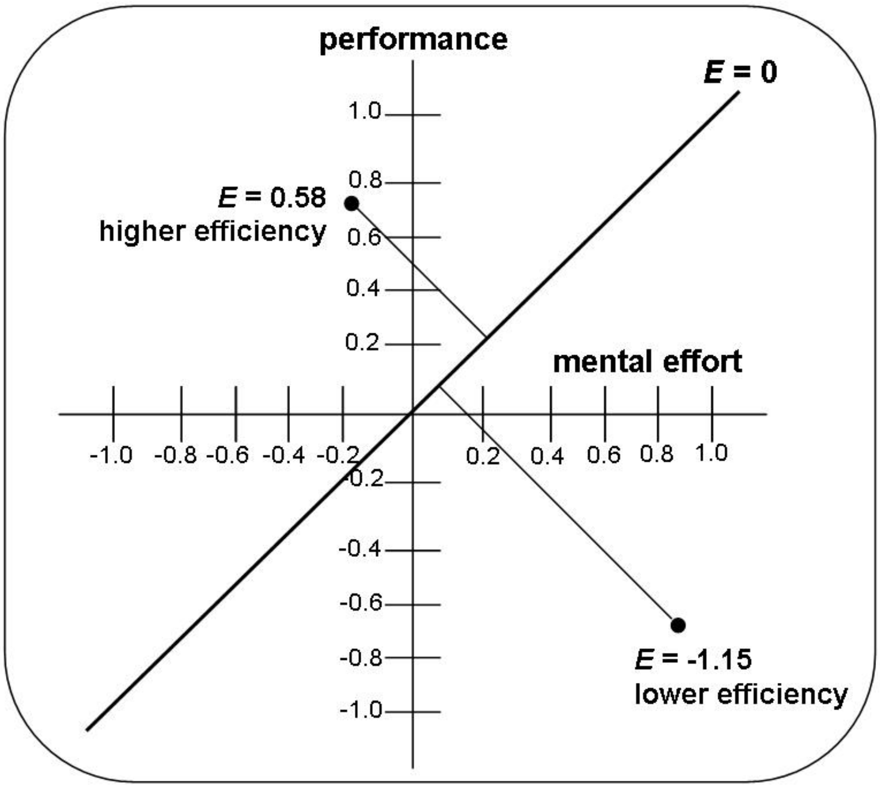


Figure 8. Representation of relative condition efficiency (E) for two groups (cf. Paas & Van Merriënboer, 1993)

Moreover, many researchers applying the efficiency formula for did not use mental effort ratings but ratings of perceived difficulty (e.g., Kalyuga, 2006; Marcus et al., 1996). According to Van Gog and Paas (2008), however, it is possible that "…the outcomes of the effort and difficulty questions that are inserted in the efficiency formula are completely opposite" (p. 23). They argue that mental effort ratings might measure different cognitive load types than perceived difficulty ratings. According to them, it is even possible that mental effort ratings might sometimes measure germane cognitive load, whereas perceived difficulty ratings might represent extraneous cognitive load. Thus, even if the revised version is used, it remains unclear, what the efficiency measure exactly indicates (for further complications of how the efficiency formula was used see de Jong, 2010). For example, consider two groups that report the same amount of mental effort, whereby the second group has higher learning outcomes than the first group. According to the measure, the second group would be classified as being more efficient. However, this result does not explain why the second group performed better. Efficiency measures do not inform about the three load types. Interpretations like for example, that the first group might suffer from higher extraneous cognitive load, whereas the second group might benefit from higher germane cognitive load are only post-hoc and do rely on empirical evidence.

Tuovinen and Paas (2004) developed a three-dimensional measure that combines mental effort ratings during the learning phase and mental effort ratings during the test phase with learning outcomes. Whereas the two-dimensional efficiency measure was applied by many CLT researchers, the three-dimensional measure has not been applied that much so far. The main reason probably is that the efficiency results become more and more difficult to interpret without gaining further insights into the pattern of cognitive load types. Hence, instructional efficiency, no matter whether it is based on the two-dimensional or the three-dimensional formula, might be seen as a rather unsuitable method with regard to the triarchic model of cognitive load. Especially, since the construct of germane cognitive load was introduced into CLT and since cognitive load research has started to shift its focus towards finding instructional techniques that elicit germane cognitive load by stimulating the allocation of working memory resources to relevant processes for learning (cf. Bannert, 2002; Paas & Van Gog, 2006) these problems became very salient. However, the efficiency approach treated cognitive load more or less as a homogeneous construct instead of one that comprises qualitatively different load types. Thus, it has been suggested to develop rating scales for the three cognitive load types separately.

3.2.3.3.2 Multiple rating scales

Even ten years after the introduction of germane cognitive load into CLT, there are only few researchers who attempted to capture the multi-dimensionality of the cognitive load construct directly (e.g., Corbalan, Kester, and Van Merriënboer, 2008; Fischer, Lowe, & Schwan, 2008; Gerjets, Scheiter, & Catrambone, 2004; Gerjets, Scheiter, & Catrambone, 2006; Gerjets, Scheiter, Opfermann, Hesse, & Eysink, 2009; Kester, Lehnen, Van Gerven, & Kirschner, 2006; Whelan, 2006; Windel & Wiebe, 2007). Many of these researchers referred to an already existing multidimensional scale measuring mental workload, the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The CLT literature on this scale is discussed in the following. The NASA-TLX distinguishes among six subscales: (1) mental demands (How much mental and perceptual activity was demanded?), (2) physical demands (How much physical activity was required?), (3) temporal demands (How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?), (4) performance (How successful do you think you were in accomplishing the goals of the task set by the experimenter?), (5) effort (How hard did you have to work to accomplish your level of performance?), and (6) frustration level (How insecure, discouraged, irritated, stressed versus secure, content, and relaxed did you feel during the task?). According to Hart and Staveland (1988) the NASA-TLX scale is more sensitive to experimental manipulations than other unidimensional rating scales. Windell and

Wiebe (2007) further concluded that the global NASA-TLX measure (a weighted combination of all six subscales) was more sensitive to indicate differences in instructional design manipulations thought to represent extraneous and intrinsic cognitive load than a one-dimensional rating scale asking "How difficult was it for you to understand this learning module and correctly answer the questions that followed?" (p. 9). The one-dimensional rating scale, however, was also sensitive in indicating differences in content difficulty manipulations thought to represent intrinsic cognitive load. Because the phrasing of the difficulty scale used by Windell and Wiebe (2007) focused on understanding and answering test items, it is also possible to infer that participants rated content difficulty only without bothering about the instructional design. This would explain why the rating scale did not indicate differences in extraneous cognitive load thought to represent the different instructional formats. Moreover, it seems to be disputable why one ratings scale should be enough to measure both intrinsic as well as extraneous cognitive load. Hence, it seems to be too premature to infer that a global overall-value of the NASA-TLX is more appropriate to measure both extraneous and intrinsic cognitive load (without being intrigued by physical and motivational factors) than other multiple rating scales with specific phrasings.

Whether the NASA-TLX really provides a sensitive and valid measure for the three different cognitive load types was also questioned by other researchers. Some of the researchers, who considered using the TLX, were not satisfied with it, and therefore, modified the test by not using all questions and/or rephrasing the questions (e.g., Fischer et al., 2008; Gerjets et al. 2004, 2006). Gerjets et al. (2004, 2006, 2009) were the only ones who attempted to map the scales they used onto the three load types. For example, Gerjets et al. (2006, p. 110-111) asked "how much mental and physical activity did you invest to accomplish the learning task, e.g., thinking, deciding, calculating, remembering, looking, searching etc." (cf. mental demands and effort) to measure intrinsic cognitive load. To measure extraneous cognitive load they asked "how much effort did you invest to navigate the learning environment" (cf. effort), and to measure germane cognitive load they asked "how hard was it for you to understand the contents of the learning environment" (cf. effort). Although the results were in line with their predictions, one should ask whether the phrasings used really represented the three cognitive load types. For example, one can ask why processes needed to accomplish the task like calculating and searching should not be influenced by the instructional format. Furthermore, it seems questionable why only navigating but not searching for information should be extraneous. Finally, it seems questionable why "working hard to understand something" should only indicate germane processes but not also include inhibiting extraneous processes. Hence, after analyzing the existing applications of multiple rating scales, it seems questionable whether the scales used so far are already optimal to measure the three cognitive load types separately.

3.2.4 Behavioral Activities as Cognitive Process Measures

The above mentioned techniques to measure cognitive load do not provide any information about which specific activities and which specific cognitive processes learners actually exert during learning. Neither studying times, nor secondary task performance, nor subjective ratings tell *how* learners actually process learning materials. However, CLT as well as CTML assume that cognitive load is caused by specific cognitive processes which should be related with more or less observable learning behavior. To overcome this lack of information in cognitive load measures, one way to measure cognitive load is to capture students' learning behavior. Students' behavioral activity data describe specific activities which can be observed directly. Hence, behavioral learning activities can be measured objectively. Subsequent to explaining the basic rationale of behavioral learning activities as measures of cognitive load, some learning activities which were reported in studies about the split-attention effect are shortly discussed before the main focus is set on learners' viewing behavior. Because information about learners' viewing behavior seems to be a promising way to find out more about learners' cognitive processing without disturbing learners during knowledge acquisition, the eye-tracking methodology has recently gained more and more attention by CLT and multimedia researchers (see for instance the special issues by Scheiter & Van Gog, 2009; Van Gog & Scheiter, 2010).

3.2.4.1 General rationale

Behavioral activities of learning, like for instance reading sequences, are thought to provide crucial online information about the learning process. However, knowledge acquisition (of concepts) is a cognitive process, and thus, behavioral activities are not equivalent with cognitive activities (Mayer, 2001). Although there is no direct relation between observable learning activities and the cognitive load types, the basic assumption is that behavioral learning activities may provide rich information about cognitive load. Before one can use behavioral activities as measures of intrinsic, extraneous, or germane cognitive load, however, two steps have to be taken. First, one must infer or determine theoretically which behavioral activities are related with which cognitive processes. Second, the cognitive processes must be classified as one of the three cognitive load types. These two steps are critical and need to be based not only on theoretical arguments but also on empirical evidence (Gerjets et al., 2009), if behavioral activities during learning should be used as measures of cognitive load. Moreover, these steps are rather difficult because behavioral activities might be ambiguous. For example, if learners switch between reading a text and inspecting a picture, it is not obvious per se whether the learner constructs an integrated model as suggested by the CTML, and thus, engages in generative processing or germane cognitive load or whether the learner is just distracted by one of the information sources or is searching for the corresponding information, and thus, engages in extraneous processing or extraneous cognitive load as suggested by CLT.

3.2.4.2 Different behavioral activities

There are many behavioral activities that can be potentially measured to investigate cognitive processing. Cognitive researchers try to capture activities that provide as much information as possible about the cognitive processes during learning by means of different methods. A possible method is to ask learners to self-explain (Chi et al., 1989) or think-aloud during learning (Ericsson & Simon, 1993). The utterances produced by learners are analyzed and classified. In investigating the split-attention effect, for example, Mwangi and Sweller (1998) asked learners to self-explain and analyzed the verbal utterances in eight categories: (1) rereads, (2) paraphrases, (3) operations, (4) subgoals, (5) goals, (6) metacognitive statements, (7) inferences, (8) incorrect inferences. Although Mwangi and Sweller (1998, p. 180) argued that self-explanations “provide a qualitative measure of the relative cognitive load associated with studying integrated and split-source example formats”, the authors did not define the relations between the behavioral activity measures (e.g., rereads, inferences) and the three cognitive load types explicitly. Moreover, self-explaining is thought to be a learning strategy (Chi et al., 1989) and thus influences the way how learners process materials. Therefore, other methods which are less intriguing seem to be better suited to measure learning activities and thus cognitive processing during learning. Another method used in investigating the split-attention effect was used by Martin-Michielott and Mendelsohn (2000). These authors recorded information about students’ navigation behavior during learning a computer program by means of log file protocols which can be automatically recorded by respective computer programs. Recording log file protocols does not influence learning behavior. The authors showed that only students with separated format explored commands of the program that were not part of the manual and thus not relevant for learning in that case. This information about learners’ processing behavior of the program that was revealed by log file data was helpful in explaining the results of the study by Martin-Michielott and Mendelsohn (2000). However, the method of log file analysis is limited with respect to the information it provides when one wants to know how learners process the information presented on one page. Log files do not reveal information about which information learners actually process when they open a specific page of a learning program that contains lots of information like text and graphics. It remains unknown whether learners read only the text, or inspect only the graphic or maybe neither of both information sources presented on a specific page. To find out more on how learners actually process the information presented on one page, the eye-tracking methodology seems to be more appropriate because it provides rich information about learners' viewing behavior. Viewing behavior as one type of behavioral activities during learning is outlined in more detail in the following sections.

3.2.4.3 Viewing behavior

Analyzing learners' viewing behavior that was collected during studying seems to be a promising way to find out more about multimedia learning and its underlying cognitive mechanisms because learners' viewing behavior provides direct information about the way different learners processed instructional formats. Before the results concerning measures of viewing behavior in the most relevant eye-tracking studies with regard to the influence of instructional design characteristics and of learners characteristics are presented, the basic assumptions of viewing behavior and its analyzing methods are presented.

3.2.4.3.1 General rationale

To investigate *how* learners process learning materials visually, eye tracking is a suitable method, because it can provide information on the distribution of learners' visual attention throughout the whole learning process (Mayer, 2010). Nevertheless, visual attention allocation is a behavioral activity, not a cognitive one. Thus, eye-tracking data are in many cases not self-explaining and require the researcher's interpretation with regard to cognitive processes (Scheiter & Van Gog, 2009). An important assumption in interpreting eye tracking data is the so called eye-mind hypothesis (Just & Carpenter, 1980). According to this hypothesis, information that is fixated by the person's eye is processed in the person's mind. In general, when we read or inspect pictorial scenes, we move our eyes to perceive all information. The eyes' movements are generally called saccades. These movements are very fast, about 500° per second. Between two saccades, however, the eyes remain relatively still and fixate visual information for about 200 – 300 ms (Rayner, 1998). Information is assumed to be encoded only during these fixations, because information is passed too rapidly during a saccade so that it cannot be encoded by the visual system. Although we can move our (cognitive) attention without moving our eyes (Posner, 1980), it is more efficient to move our eyes rather than our attention alone when we process complex information (He & Kowler, 1992). There is evidence suggesting that attentional movements and saccades are obligatory coupled (Deubel & Schneider, 1996). These findings support the eye-mind hypothesis and qualify viewing behavior as a potential behavioral correlate of cognitive processing.

Eye tracking data provide rich data about which information learners look at and when and for how long they look at this information during reading instructional materials. This spatial (content information) and temporal information contained in the fixations and saccades generated by eye-tracking systems can be analyzed into differently complex measures combining both information types by means of the following analyzing methods (Holmqvist et al., 2011): (1) Temporal analysis provides information about the duration of learners' fixations (measure: fixation duration) and about how long learners processed (or at least looked at) a specific information (measure: dwell time). (2) Frequency analysis provides information about how often a learner looked at specific information (measure: rereads of verbal or pictorial information). (3) Transition analysis provides information about whether and how often a learner switched between two different information units (measure: i.e., frequency of switches or spatial density of switches). (4) Sequence analysis provides information about when a learner processed which information during learning (measure: i.e. learning sequence). Depending on the interest of research one or more of these measures are suitable to describe learners' viewing behavior and the cognitive processing assumed to be associated with it.

3.2.4.3.2 Viewing behavior in multimedia learning

So far, eye tracking studies have provided thorough insights into the processes involved in word and sentence reading and scene perception (for a detailed review see Rayner, 1998). The analysis of processes involved in learning from complex multimedia materials, however, is still in its beginnings. Although research in this area mainly concentrates on analyzing the overall dwell/fixation time on text vs. graphic (Hyönä, 2009), authors increasingly present different measures of viewing behavior. Whether and how multimedia materials influence learners' viewing behavior is presented in the next sections summarizing the findings from eye-tracking studies on how learners process multimedia instructions. Besides studies which manipulated instructions with regard to modality of text (text presented either in written or auditory mode) and color coding (corresponding information of text and graphic is either presented in the same color or not), a first study investigated the influence of spatial contiguity between text and graphic.

Early research. In the 1990s, several studies recorded learners' viewing behavior during processing multimedia material and described it as precisely as possible but did not test two or more design formats against each other. Hannus and Hyönä (1999) studied the processing behavior of elementary school children who studied authentic textbook materials with illustrations. According to their results, learning is heavily driven by the text, whereas illustrations are only inspected minimally. Similar results were already demonstrated by Hegarty (1992a, 1992b) and Hegarty and Just (1993). In these studies students were asked to learn how pulley systems work. The pulley systems were represented by separated text-graphic formats. Most of the students read the sentence describing the special part of the depicted pulley systems, before they inspected the part on the graphic. Hegarty concluded that the construction of a mental model is heavily driven by the text. This behavior seems to be a rather general processing behavior, because this finding was also demonstrated in research on the processing of print advertisements with very little text (Carroll, Young, & Guertin, 1992; Rayner et al., 2001). Although students sometimes switched between text and graphic, especially when information was difficult, it remained unclear whether learners with integrated format would switch more or less often than learners with separated format, because these studies did not compare different instructional formats.

Text modality. Schmidt-Weigandt, Kohnert, and Glowalla (2009a, 2009b) compared different instructional formats of the same learning content (formation of lightning). They investigated in three experiments how instructions consisting of animations that were either explained by written text or by narrations (cf. modality and temporal contiguity principle in Mayer, 2009) were visually processed by learners. Schmidt-Weigandt and colleagues showed that the processing of the materials was highly text driven, because learners in the written text conditions first read the texts before they switched to the animations and spent more time reading the text than inspecting the animations. Furthermore, the authors showed that learners in the auditory text condition spent more time studying the animations. The long studying times of animations in the auditory text condition were interpreted as causal factor for the modality effect that is the effect that learners with spoken text and animation learn better than learners with written text and animation (Low & Sweller, 2005). These studies on multimedia instructions suggest that processing time on graphical information that was presented in animations might be an indicator of germane processing because learners who watched the animations in average longer had higher learning outcomes than learners who watched the animations in average shorter. These studies showed the primacy of the written text, but they also demonstrated that putting the animation into learners' visual attention might lead to more elaborated cognitive processes. However, the comparison of the viewing behavior during processing instructional formats presenting animations and written text with the viewing behavior during processing instructional formats presenting only animations and no further visual information is seen very critical. Therefore, these results should not be generalized so far. Moreover, animations convey transient information and therefore it is difficult to say, whether the processing of static pictures accompanied by different text modes is related with a similar viewing behavior.

Color coding. Further information about whether the instructional format influences how learners process the instructional materials was provided by two studies on the effect of color coding. When separated formats of text-graphic instructions highlight corresponding textual and pictorial information in the same color, then these instructions are said to be color-coded. Kalyuga et al. (1999) assume that the color coding of the corresponding elements of the text and graphic reduce visual search processes in separated formats and their associated extraneous cognitive load. Ozcelik, Karakus, Kursun, and Cagiltay (2009) investigated this assumption in an eye-tracking study. The authors demonstrated that students with the color-coded format not only had higher learning outcomes but also longer average fixation durations than students without color-coded format. Learners' average fixation durations were positively related to their performance in the transfer test. Furthermore, learners with the color-coded format needed less time to find corresponding text and graphic elements than learners without color coding. Moreover, Ozcelik et al. (2009) also measured cognitive load by means of learners' perceived difficulty of understanding the instruction. Learners with the color-coded format reported the same level of perceived difficulty like learners without the color-coded format. Learners' perceived difficulty was not related with their average fixation durations. The authors suggested that color coding induced deeper information processing. This assumption is in line with former research that has shown that longer fixations in text-graphic materials are indicative of a deeper and more integrative information processing (Carroll et al., 1992; Underwood, Jebbett, & Roberts, 2004). Thus, according to the triarchic model of cognitive load, the results of the study suggest that color coding stimulated the learners to increase germane cognitive load as represented in longer fixation durations and did not reduce learners' extraneous cognitive load because the groups did neither differ in their perceived difficulty nor were these ratings related with the average fixation durations. Whereas Ozcelik and colleagues showed interesting results with regard to learners' average fixation durations and learning outcomes, Folker, Ritter, and Sichelschmidt (2005) did not find differences in learning outcomes between learners with and without a color-coded format in their eye-tracking study on color coding. However, they showed that learners with the color-coded format answered the test items faster than learners without color coding. The authors argued that the faster the test items were answered the better the mental representations of the information to be learned. Moreover, Folker and colleagues showed that learners with the color-coded format processed the graphic shorter than learners without color coding, whereas there were no differences between the format conditions with regard to the text. A further result was that learners with the color-coded format made more switches per second between text and graphic than learners without color coding. The authors argued that these results showed that learners with the color-coded format not only invested less time in visual search processes as indicated by shorter dwell times on the graphic but also engaged in more integrative processing as indicated by the higher number of switches per second.

To sum up, both studies showed that the instructional format influenced learners' viewing behavior and obviously related cognitive processes. Although Folker and colleagues did not find differences between the instructional format conditions in the learning outcomes, the results of learners' viewing behavior showed that learners processed the information differently depending on the instructional format. The authors of both studies argued that not only visual search (extraneous cognitive load) was reduced by color coding as suggested by Kalyuga et al. (1999) but also that color coding increased the elaborated processing of the learning materials (germane cognitive load). The last assumption was especially corroborated by the findings that average fixation durations were related with performance in a transfer test but not with perceived difficulty in the study of Ozcelik et al. (2009). By testing the relations between different types of measures, namely performance data, measures of viewing behavior, and subjective cognitive load ratings (perceived difficulty), Ozcelik and colleagues helped to disambiguate the meaning of the viewing behavior measures.

Spatial contiguity. Although the aforementioned studies suggest dwell time on graphic (and text) as well as switching behavior to be possible important indicators of cognitive processes and thus of cognitive load, a study of Holsanova, Holmberg, and Holmqvist (2009) is especially relevant for this thesis. These authors explicitly compared the eye movements of readers with integrated vs. separated information formats presented in a regular newspaper format. Hence, this study is directly related to the spatial contiguity principle. According to Holsanova et al. (2009) readers with integrated format switched more frequently between text and corresponding graphic, and thereby are assumed to show a better integrative processing of text and picture than readers with separated format. This finding shows that spatial contiguity between text and picture also influences learners' viewing behavior. Moreover, the findings might be interpreted as support of the active processing assumption of Mayer's CTML (e.g., Mayer, 2001). However, the number of integrative switches (switches between corresponding elements of text and graphic) was not related with comprehension performance. Thus, Holsanova et al. (2009) argued "that a few well-placed and well-timed movements between relevant parts of text and graphics are more important than making as many saccades as possible" (p.1224). They further argued that switches between text and graphic can be caused by different cognitive processes because switches may "reflect either a successful integration of the material or difficulties integrating the information" (p. 1225). The discussion of these authors leads to the following questions but provides no answers to them: Which learners know when it is the best time to switch and where to switch? Do specific learner characteristics like prior knowledge influence learners' switching behavior between text and picture? Whether and how learners' prior knowledge influences learners' viewing behavior is outlined in the next sections.

3.2.4.3.3 Viewing behavior and prior knowledge

The above outlined research investigated how characteristics of instructional materials influence viewing behavior and thus focused on so called bottom-up processes. Another line of eye tracking research investigated how characteristics of learners influence their viewing behavior and thus focused on so called top-down processes. The question behind this research is to investigate whether individual characteristics like prior knowledge already influence the early processes of information selection (visual attention/encoding) or whether such variables influence the information processing only at later cognitive stages after information selection/encoding. The classical study of Yarbus (1967) has already demonstrated that the way how persons inspected a painting was influenced by the task they had to perform when inspecting the painting. This finding shows that viewing behavior can be at least partially determined by so called top-down processes triggered by the characteristics of the cognitive system instead by perceptual characteristics of the materials to be processed. Recent research has shown similar results. For example, Wedel, Pieters, and Liechty (2008) showed that persons made more but shorter fixations on local parts of print advertisements when they were asked to learn than when they were asked to evaluate the advertisements. Information on whether prior knowledge also influences viewing behavior is presented in the next sections.

Viewing behavior on one instructional format. Researchers investigated whether prior knowledge influences the selection or perceptual encoding of information or whether it influences information processing only at a later stage after information encoding. For instance, Haider and Frensch (1999) showed that students who were trained in verifying artificial alphabetic strings containing task-relevant and task-irrelevant information actively ignored the task-irrelevant information and concentrated on task-relevant information. According to their information-reduction hypothesis, Haider and Frensch (1999) argued "that persons learn, with practice, to become selective in their use of information, that is, to distinguish between task-relevant and task-redundant information and limit their processing to task-relevant information" (p. 172). The influence of expertise on the perceptual processing of visual stimuli was also shown during car driving (Underwood et al., 2003) and in the domain of chess (Charness, Reingold, Pomplun, & Stampe, 2001). Charness et al. (2001) demonstrated that expert chess players had a higher proportion of fixations on relevant pieces than had intermediate players. The authors argued that expert chess players perceptually encoded chess configurations, instead of individual pieces. Hence, good performance in chess seems to depend on an efficient perception of the configurations, which is obviously linked with expertise. In a recent study on the comprehension of weather maps, Canham and Hegarty (2009) tested whether the information reduction-hypothesis can be generalized to domain prior knowledge in complex and realistic learning tasks. Canham and Hegarty (2009) showed that learners with low prior knowledge looked longer on irrelevant information on weather maps during a verification task, whereas after a tutorial on the pressure gradient and Coriolis phenomena these learners looked longer on relevant information than irrelevant information. This finding corroborates the information-reduction hypothesis that more knowledgeable learners ignore irrelevant information already on a perceptual level and concentrate on the encoding of relevant information.

Viewing behavior in comparing two instructional formats. Whereas the former studies showed a main effect of prior knowledge on information selection, a recent study on learning from non-linear concept maps either presented in a hierarchical or a network structure, showed not only an expertise reversal effect on learning outcomes but also the moderational influence of prior knowledge on perceptual information processing (Amadieu, Van Gog, et al., 2009). Amadieu, Van Gog et al. (2009) demonstrated that low-knowledge learners with the hierarchy map had longer fixation durations than low-knowledge learners with the network map, whereas there were no differences between high-knowledge learners. The authors suggested that low-knowledge participants with the hierarchy map but not high-knowledge participants processed the information more deeply as indicated by the longer fixation durations than those with the network map. This interpretation was supported by better performance on a conceptual knowledge test by low-knowledge students with the hierarchy map, whereas there were no performance differences between high-knowledge students on this test. The results of this study also indicate that learners with different levels of prior knowledge probably differ in their information selection depending on the instructional format, and this might influence their learning outcomes. Hence, the expertise reversal effect might be caused by different ways of processing integrated or separated formats depending on learners' prior knowledge. Different ways of processing might cause or represent different types of cognitive load. However, so far not much is known about the relations between learners way to process different formats of text and graphic and cognitive load, although Mayer and colleagues (e.g., Mayer & Gallini, 1990) as well as Sweller and colleagues (e.g., Kalyuga et al., 1998; Sweller & Chandler, 1994) claim that specific ways of processing depending on learners prior knowledge cause different cognitive load types and thereby influence learning outcomes.

3.3 Conclusions: Cognitive Load Explanations and Cognitive Load Measures

To investigate which cognitive load explanation is more suited to explain the split-attention and expertise reversal effect, it is necessary to measure cognitive load. However, the rationales and empirical studies discussed above show that there is neither an ideal instrument for measuring total cognitive load nor an ideal instrument for measuring the three cognitive load types separately, so far. Nevertheless, it was shown that the measures above have certain strengths and potentials. The conclusions with regard to testing cognitive load explanations with these measures are presented in the following sections.

In explaining the split-attention effect, the extraneous cognitive load explanation assumes that extraneous cognitive load is responsible for the split-attention effect, and more specific, that integrated formats reduce extraneous cognitive load. Hence, learners with integrated formats have lower overall cognitive load than learners with separated formats. To test this specific assumption a measure of overall cognitive load would be helpful. It was discussed above that studying times are not only a weak measure to make valid conclusions on the type of cognitive load but also a weak measure with regard to the quantitative aspect of cognitive load. The fact that studying times do not tell whether learners invest cognitive resources or not makes studying times a less useful measure. Moreover, if learning times are held constant across groups as is the fact in more controlled studies, there is not even the option that learners can differ at all in studying times. Thus, to measure the quantity of overall cognitive load, secondary task performance is the better option. When the rationale of the dual-task paradigm is accepted, then secondary task performance can be used as a relative (in comparison to other (primary) tasks) measure of overall cognitive load under the condition that the implementation demands are met (no primary task intrusion or structural interference). To meet both implementation demands but also to be as similar as possible to former studies in this field (see Chandler & Sweller, 1996) it is necessary to use a visual secondary task probe that is not too difficult. However, as learning with separated or integrated formats and performing a secondary task with visual stimulus is prone to structural interference, it is helpful to not only use a visual secondary task stimulus that is restricted to one specific area (see the color change of a letter placed in the upper part of the materials in the study of Brünken et al., 2002) but also to test a visual stimulus that is not restricted to one specific area in the learning materials. Such a visual stimulus might be the color change of the whole background of the learning materials. Implementing two different visual stimuli does allow investigating whether the type of secondary task stimulus influences secondary task performance. Moreover, to test that the secondary task does not cause primary task intrusion, in this case reducing learning outcomes, there should be also control groups without any secondary task.

Although secondary task performance is an objective measure of overall cognitive load, it does not differentiate between the three cognitive load types. However, if the germane cognitive load explanation is true, there might be no differences in overall cognitive load between integrated and separated format. Therefore, additional measurement techniques are needed that differentiate between the three cognitive load types. Subjective rating scales seem to be a rather promising method to measure the different cognitive load types. So far, different scales were used to measure especially extraneous and/or intrinsic cognitive load. For example, some CLT researchers asked students to rate the perceived difficulty. Depending on the experimental design, the instructional manipulations, perceived difficulty seems to be a good measure for intrinsic and extraneous cognitive load. However, the formulations to distinguish between intrinsic and extraneous cognitive load are not that specific so far to distinguish between both types in one experiment. Thus, to differentiate between intrinsic and extraneous cognitive load in one experimental design, it is necessary to be precise in asking for the primary source of the perceived difficulty. As intrinsic cognitive load is caused by the complexity of the learning content, it seems to be necessary to ask learners after an experimental learning phase how difficult the learning content was for them. One may ask *“How difficult was the learning content for you?”* And as extraneous cognitive load is caused by the format of instruction as inherent characteristic of the materials, learners might be asked how difficult the learning materials were for them. One may ask *“How difficult was it for you to learn with the materials?”* Others like Salomon (1984) asked students to rate their level of concentration. According to his construct this scale might measure germane cognitive load because the investment of concentration raised learners’ learning outcomes. Hence, to measure germane cognitive load one may ask learners *“How much did you concentrate during learning?”* Although it is controversial how thoroughly learners can retrospect on their own cognitive load during learning, the use of multiple rating scales measuring intrinsic, extraneous, and germane cognitive load in one single study may help to investigate whether the subjective multiple ratings are in line with the different cognitive load explanations of the split-attention effect or not.

In explaining the expertise reversal effect, the extraneous cognitive load explanation assumes that high-knowledge learners with the integrated format are loaded by extraneous cognitive load, whereas the germane cognitive load explanation assumes that high-knowledge learners with the separated format are loaded by germane cognitive load. In case that both explanations are true with regard to their assumption it would not be very helpful to measure learners overall cognitive load to differentiate between both explanations because high-knowledge learners of both instructional format conditions might have the same overall cognitive load despite different patterns of cognitive load types. Thus, secondary task performance would not be helpful as cognitive load measure. To distinguish between learners’ load type pattern the three cognitive load types should be measured separately. As mentioned above, this might be done by multiple subjective ratings scales that ask for the specific source of the cognitive load types (intrinsic: content difficulty; extraneous: difficulty of materials; germane: learners’ level of concentration).

Moreover, although not used very often so far, learners’ behavioural activities seem to be an interesting measure to investigate the expertise reversal (and/or split-attention) effect. As stated above, learners’ viewing behaviour seems to be a promising measure because one can assume in general that this information is processed in mind which learners look at (eye-mind hypothesis). Hence, recording and analyzing learners’ viewing behaviour provides direct evidence of the way differently knowledgeable learners process integrated and separated format. The eye-tracking methodology therefore helps to investigate whether Kalyuga et al.’s (1998) assumption that high-knowledge learners with separated format focus on graphics or whether Mayer's and Gallini's (1990) assumption that high-knowledge and low-knowledge learners focus on text is correct. Moreover, learners' viewing behaviour would help to investigate whether the assumption of Erhel and Jamet's (2006) is true that low-knowledge learners with separated format switch very often between text and graphic and therefore suffer from extraneous cognitive load or whether the assumption of Sweller and Chandler (1994) is true that low-knowledge learners with separated format not only have to switch between text and graphic but also have to visually search the right places on the graphic to mentally integrate text information with graphical information, a process which should cause high extraneous cognitive load.

If one wants to investigate the aforementioned assumptions about learners' processing behaviour, one has to decide which kind of viewing behaviour measures should be analyzed. The following three types of measures are thought to be important ones when investigating learners processing behaviour of text and graphic materials. As mentioned above, measures of viewing behavior contain differently complex information of time (fixation durations) and/or information of space/content (areas of interest, AOIs). (1) With regard to the question of how integrated and separated formats are processed by learners the simplest measure is learners' *average fixation duration* without any further differentiations. This measure was already successfully used by Ozcelik et al. (2009) in investigating color coding. Nevertheless, this measure is not very informative, if one wants to learn more about learners' processing behavior of multimedia materials. (2) A more complex but also a more informative measure combining temporal and spatial/content information is learners' *dwell time on text* and their *dwell time on graphical information* to test the different assumptions made by Kalyuga et al. (1998) and Mayer and Gallini (1990). This measure was already successfully used by Folker et al. (2005) investigating color coding. Although this measure contains more information than simple average fixation durations, it does not provide insights in how thoroughly learners actively integrate text and graphic information. (3) To learn more about behavioural and cognitive integration processes, learners' *switching behaviour* between different information units should be analyzed, especially because high-knowledge learners are not assumed to switch between text and graphic but either concentrate on text (germane cognitive load assumption) or on the graphic (extraneous cognitive load assumption). Folker et al. (2005) already investigated learners' switching rate per minute. However, they reported only switches between text and graphic without differentiating whether learners switched between corresponding or non-corresponding text-graphic information. Holsanova et al. (2009) were more specific and reported the proportion of switches between corresponding text and graphic information (so called integrative saccades). However, to investigate in more detail how learners process integrated and separated text-graphic formats, a comprehensive overview of different possible types of switches is needed. There are four possible types of switches: (a) *switches between different text units* representing different information, (b) *switches between different graphical units* representing different information, (c) *switches between non-corresponding text-graphic units/information* representing different information, and (d) *switches between corresponding text-graphic units/information.* To analyze these types of switches one might concentrate on their frequency that is the absolute number of switches (per minute) or the proportion of a switching type. The above mentioned authors concentrated on the frequency dimension. However, the frequency dimension does not provide information on whether learners switch very often between for example only one corresponding text-graphic information or whether they switch only once between many different corresponding text-graphic information units. In order to gain knowledge of how comprehensively learners process the whole learning material the spatial dimension of the switching types should be considered. Taking the spatial dimension into account has the advantage that information is provided on how many percentages of possible switching options a learner really exerts (see Goldberg & Kotval, 1999). This provides on the one hand information on how intensively one information representation (e.g., text) was processed in isolation without considering the other representation (e.g., graphic), and on the other hand it provides information on how integratively and (un)structured two types of representation (text and graphic) are processed in combination. Thus, switching measures that are based on the spatial dimension seem to be most informative for investigating how learners process integrated and separated formats. So far, no study is known which used these measures to analyse learners' switching behaviour between text and graphic. To investigate further, whether viewing behaviour measures represent a specific cognitive processing or load type, Hyönö (2009) recommends a complementary approach. According to the complementary approach, it is helpful to investigate how learners' viewing behaviour correlates with subjective cognitive load ratings as well as with learning outcomes.

II. EMPIRICAL WORK

Trust, but verify.

(Old Russian proverb assigned to Vladimir Lenin)

4 The Split-Attention Effect: Cognitive Load Explanations and their Empirical Evidence

It's black, it's white.
(From Michael Jackson's Black or White, 1991)

Mayer's general multimedia principle states that adding graphics to text enhances learning. However, care has to be taken on how the text and picture are arranged. With regard to interpretational pictures depicting scientific or technical contents an important specification was that the spatial distance or contiguity between textual information and corresponding graphical information has to be considered. The phenomenon that students with integrated formats (high spatial contiguity) outperform students with separated formats (low spatial contiguity) is termed the split-attention effect[1] (for an overview see Ayres & Sweller, 2005; Mayer, 2001). Hence, the corresponding spatial contiguity principle (Mayer, 2001, 2009) or split-attention principle (Ayres & Sweller, 2005) recommends that words and corresponding graphics should be placed as near to each other as possible.

The importance of spatial contiguity was first discovered by CLT researchers investigating mathematical problem-solving in geometry (Tarmizi & Sweller, 1988). Shortly afterwards, Mayer (1989) discovered independently from CLT research the importance of integrated text-graphic formats in multimedia learning. Since then, there were published about 50 experiments in international peer-reviewed journals demonstrating the split-attention effect. It can be argued that the spatial contiguity effect is currently one of the best documented multimedia effects. The majority of these studies compared the spatial contiguity of text and corresponding graphic, but there are also a few exceptions which demonstrated that the effect does hold true for other types of representations. For example, Ward and Sweller (1990) showed that the integration of mathematical equations into word problems enhanced learning. Concerning the learning domain, there was a great variety among the studies. There were experiments in which students had to fold paper-discs into triangles (Bobis, Sweller, & Cooper, 1993), interpret time signatures in music (Owens & Sweller, 2008),

[1] Whereas Ayres and Sweller (2005) subsumed spatial and temporal contiguity effects under the split-attention effect, this thesis concentrates on spatial contiguity only.

compute tax liabilities (Rose, 2002), understand physics laws (Ward & Sweller, 1990), learn programming codes (e.g., Cerpa, Chandler, & Sweller; Chandler & Sweller, 1992, 1996), check electrical circuits (e.g., Kalyuga, Chandler, & Sweller, 1998), learn the development of lightning storms (Mayer, Steinhof, Mars, & Bower, 1995), and to master biological domains like the blood circulatory system (Chandler & Sweller, 1991) or the biochemical processes at the synapses (Florax & Plötzner, 2009).

In a meta-analysis on the spatial and temporal contiguity effects, Ginns (2006) analyzed 37 experiments on spatial contiguity and 13 experiments on the temporal contiguity effect. The analysis of these experiments revealed a weighted mean effect size of d = 0.85, indicating a large effect. Ginns' (2006) meta-analysis corroborates the effectiveness of spatial contiguity statistically. Despite the impressive overview of the effect sizes, Ginns (2006) did neither analyze the explanations offered by the researchers nor did he analyze the empirical evidence in favor of these explanations. However, as already described in Chapter 2, there are complementary explanations of the split-attention effect. Whereas most CLT researchers (e.g., Ayres & Sweller, 2005) favor an extraneous cognitive load explanation (learners with separated formats suffer from high extraneous cognitive load by being forced to search corresponding information and integrate it mentally), CTML researchers (e.g., Mayer, 2001) suggest a germane cognitive load explanation (learners with integrated formats benefit from high germane cognitive load by being supported to construct an integrated mental model).

To evaluate the evidence of each explanation, the following review of 47 experiments on the split-attention effect particularly focuses on the cognitive load measures used to corroborate the respective cognitive load explanation. Thirty-six experiments[2] reported in Ginns' (2006) meta-analysis served as basis of the review. Eleven additional experiments were included. Five of them were already published in papers cited by Ginns (2006) but not included in his meta-analysis and six of them were published after Ginns' (2006) meta-analysis. The experiments were categorized with regard to the cognitive load explanation suggested by the respective authors. This procedure resulted in two groups arguing either for

[2] The work of White (1993) was not included, because this dissertation thesis concentrates on articles in English peer-reviewed journals only. Experiment 3 in Sweller and Chandler (1994) was neither included, because it investigated the redundancy effect.

(1) the extraneous cognitive load explanation or (2) for the germane cognitive load explanation (see Table 3). Table 3 summarizes the experiments, the learning domain, the theoretical framework, the cognitive load measures, and whether the cognitive load measures could be interpreted in support of the respective explanation. The next sections present a qualitative review of the split-attention effect explanation. Afterwards, there follows a quantitative summary of the empirical evidence of both cognitive load explanations and recommendations for further research.

Table 3
Summary of the reviewed papers about the split-attention effect

Study	Domain	Theoretical framework	CL measure	CL measure supports explanation
ECL Explanation				
Bobis et al. (1993) *Experiment 4*	Practical folding task	CLT	Studying times	Yes
Bodemer, Plötzner, Feuerlein, & Spada (2004) *Experiment 1*	Mechanical system of tire pump	CLT	—	—
Bodemer, Plötzner, Feuerlein, & Spada (2004) [a] *Experiment 2*	Univariate variance analysis	CLT	—	—
Cerpa et al. (1996) *Experiment 1*	Computer software package	CLT	Studying times	No
Cerpa et al. (1996) *Experiment 2*	Computer software package	CLT	Studying times	No
			Efficiency based on perceived difficulty	Possibly

Table 3 *(continued)*	Domain	Theoretical framework	CL measure	CL measure supports explanation
Chandler & Sweller (1991) *Experiment 1*	Electrical installation testing	CLT	—	—
Chandler & Sweller (1991) *Experiment 6*	Blood circulation (heart and lung)	CLT	Studying times	Yes
Chandler & Sweller (1992) *Experiment 1*	Numerical control machine programming	CLT	—	—
Chandler & Sweller (1992) [b] *Experiment 2*	Scientific reports	CLT	Studying times	Yes
Chandler & Sweller (1996)	Computer software package	CLT	Secondary task performance	Yes
Kablan & Erden (2008) [c]	Location and Movement	CLT, CTML	Studying times	No
			Difficulty ratings	Yes
Kalyuga et al. (1998)	Electrical circuits	CLT	Studying times	No
			Difficulty ratings	Yes
Kalyuga, Chandler, & Sweller (1999) [d]	Electrical circuits	CLT	Efficiency based on perceived difficulty	Possibly
Martin-Michiellot & Mendelsohn (2000) [a]	Computer software package	CLT	Studying times	Yes
Moreno & Mayer (1999) *Experiment 1*	Developing of lightning storms	CTML	—	—
Mwangi & Sweller (1998) *Experiment 3*	Mathematics: inconsistent language two-step compare problems	CLT	Rereads, inferences	

Table 3 *(continued)*	Domain	Theoretical framework	CL measure	CL measure supports explanation
Owens & Sweller (2008) [c]	Time signatures in music	CLT	—	—
Purnell, Solman, & Sweller (1991) *Experiment 1*	Reading geographical diagrams	CLT	—	—
Purnell, Solman, & Sweller (1991) *Experiment 2*	Reading geographical diagrams	CLT	—	—
Purnell, Solman, & Sweller (1991) *Experiment 3*	Reading geographical diagrams	CLT	—	—
Purnell, Solman, & Sweller (1991) *Experiment 4*	Reading geographical diagrams	CLT	—	—
Rose (2002) *Experiment 1*	Tax liability	CLT	—	—
Rose (2002) *Experiment 2*	Tax liability	CLT	—	—
Rose & Wolfe (2000)	Tax liability	CLT	Studying times	No
Sweller & Chandler (1994) *Experiment 1*	Computer software package	CLT	Studying times	Yes
Sweller & Chandler (1994) *Experiment 2*	Computer software package	CLT	Studying times	Yes
Sweller & Chandler (1994) *Experiment 4*	Electrical appliances (electrical kettle + megger meter)	CLT	Studying times	Yes
Sweller, Chandler, Tierney, & Cooper (1990) [c] *Experiment 1*	Coordinate geometry	CLT	Studying times	Yes
Sweller, Chandler, Tierney, & Cooper (1990) [c] *Experiment 2*	Coordinate geometry	CLT	Studying times	Yes
Sweller, Chandler, Tierney, & Cooper (1990) [c] *Experiment 3*	Coordinate geometry	CLT	Studying times	Yes

Table 3 *(continued)*	Domain	Theoretical framework	CL measure	CL measure supports explanation
Sweller, Chandler, Tierney, & Cooper (1990) *Experiment 4*	Numerical control machine programming	CLT	Studying times	Yes
Sweller, Chandler, Tierney, & Cooper (1990) [c] *Experiment 5*	Numerical control machine programming	CLT	Studying times	Yes
Sweller, Chandler, Tierney, & Cooper (1990) [c] *Experiment 6*	Numerical control machine programming	CLT	Studying times	Yes
Tarmizi & Sweller (1988) *Experiment 4*	Circle geometry	CLT	Studying times	Yes
Tindall-Ford, Chandler, & Sweller (1997) *Experiment 1*	Electrical appliances (electrical kettle + megger meter)	CLT	—	—
Ward & Sweller (1990) [b] *Experiment 4*	Word problems on acceleration	CLT	—	—
Ward & Sweller (1990) [c] *Experiment 5*	Physics of concave mirrors and convex lenses	CLT	—	—
GCL Explanation				
Erhel & Jamet (2006)	Functioning of the heart / reproduction of AIDS virus	CLT, CTML	—	—
Florax & Plötzner (2009)	Biochemical processes in synapses	CLT, CTML	—	—
Holsanova et al. (2009) *Graphic 1*	Report on diving accident	CLT, CTML	Text-pictures switches	*n.d.**

Table 3 *(continued)*	Domain	Theoretical framework	CL measure	CL measure supports explanation
Kester et al. (2005)	Electrical circuits	CLT	Studying times	Possibly
			Mental effort ratings	Possibly
Mayer (1989) *Experiment 2*	Mechanical system of brakes	CTML	—	—
Mayer, Steinhof, Mars, & Bower (1995) *Experiment 1*	Developing of lightning storms	CTML	—	—
Mayer, Steinhof, Mars, & Bower (1995) *Experiment 2*	Developing of lightning storms	CTML	—	—
Mayer, Steinhof, Mars, & Bower (1995) *Experiment 3*	Developing of lightning storms	CTML	—	—
Pociask & Morrison (2008) [c]	Localization testing in physical therapy	CLT	Studying times	Possibly
			Difficulty ratings	No (GCL) Yes (ECL)
Tabbers, Martens , & Van Merriënboer (2000) [d]	Developing a blueprint for educational training	CLT	Mental effort ratings	Possibly

Note. [a] Experiment included in Ginns' (2006): there was no statistically significant split-attention effect on learning outcomes reported.
[b] Experiment included in Ginns' (2006): the materials consisted of different text elements or text and formula instead of text and illustration.
[c] Experiment not included in Ginns' (2006) meta-analysis about contiguity effects.
[d] Experiment investigated color coding that is assumed to enhance spatial contiguity.
CL = Cognitive load, GCL = germane cognitive load, ECL = extraneous cognitive load.
CLT = Cognitive Load Theory, CTML = Cognitive Theory of Multimedia Learning
n.d. = not defined: Holsanova et al. (2009) did not report statistics of learning outcomes.

4.1 Literature Review of Split-Attention Studies

The following sections summarize the researchers' argumentations and give a qualitative overview of the empirical evidence of the respective cognitive load explanation concerning different cognitive load measures used. The following review starts with the studies in which the extraneous cognitive load explanation was postulated because this explanation is the most prevalent one. Subsequently, the studies in which the germane cognitive load explanation was postulated are presented.

4.1.1 Extraneous Cognitive Load Explanation: Inhibiting Mental Integration Processing

Many split-attention effect studies argued in favor of the extraneous cognitive load explanation, but not all these studies provided empirical evidence supporting this explanation. The next sections group the studies according to the cognitive load measurement used.

Learning outcomes. In several studies, especially in some of the first ones (e.g., Chandler & Sweller, 1991; Purnell, Solman, & Sweller, 1991; Ward & Sweller, 1990) cognitive load was not measured at all. Thus, any assumption made in these studies was just based on learning outcomes which represent the effect and not the cause (see Chapter 3). In line with the extraneous cognitive load explanation, the authors of these studies, for example Ward and Sweller (1990), argued that the critical feature of presentation formats was to "impose a relatively light cognitive load" (p. 4). The findings that learners with integrated formats outperformed learners with separated formats were taken as evidence that integrated formats impose a relative light cognitive load on learners compared to separated formats. Because the CLT did not distinguish between the three different cognitive load types during the early 1990s, no other explanations were considered.

Studying times. As mentioned above, CLT did not yet distinguish between different load types, when Tarmizi and Sweller (1988) conducted the first experiment on the spatial split-attention effect. They investigated the effectiveness of worked-examples on learning geometry and showed that only worked examples in which the solution steps were integrated in the diagram enhanced learning outcomes compared to conventional separated worked-example formats and to conventional problem solving tasks. The authors argued that a separated format demands learners to split their attention between text and graphic and to mentally integrate both types of information. In contrast, integrated formats do not demand the mental integration of disparate information sources. Thus, by reducing the need to mentally integrate physically disparate information overall cognitive load should be reduced and learning facilitated. This argument was supported by the finding that learners with integrated worked-examples needed significantly less time to solve the problems in the acquisition phase. The results with regard to studying times thus supported the extraneous cognitive load explanation. Although similar results were demonstrated by several other studies (e.g., Bobis et al., 1993; Sweller & Chandler, 1993), there were also studies (e.g., Cerpa et al., 1996; Kalyuga et al., 1998) that did not find that learners with integrated formats studied significantly shorter than learners with separated formats. In their study on the split-attention effect, Rose and Wolfe (2000) even showed that longer studying times were correlated with higher learning outcomes for learners with higher problem solving effectiveness. For learners with lower problem solving effectiveness there was no correlation between studying times and learning outcomes. A positive correlation is against the interpretation of studying times as extraneous cognitive load but fits the interpretation of germane cognitive load. These diverse results underline that studying times are a rather critical cognitive load measure (see Chapter 3.2).

Secondary task performance. The first 21 experiments conducted to investigate spatial contiguity between text and picture did either not measure cognitive load or measured it only indirectly by means of studying times. This changed, when Chandler and Sweller (1996) applied the dual-task paradigm to measure learners' cognitive load directly. Learners' secondary task performance was a direct measure of learners' overall cognitive load. In this study students had to learn how to use a computer program. During studying, students had to remember letters which were presented on a separate screen. The secondary task was to recall the letter previously presented on the screen while encoding the new letter. The students of the split-attention conditions who learned with a manual plus a computer recalled fewer letters correctly than students who learned with the integrated manual only. According to this result, Chandler and Sweller (1996) corroborated the assumption that overall cognitive load is higher in learners with separated format than with integrated format. This higher overall cognitive load was interpreted as extraneous cognitive load, because learners with separated format also had lower learning outcomes than learners with integrated format (integrated diagram only). However, as already noted in chapter 3.2, the results of the secondary task performance in this study are hard to interpret because of learning intrusion and structural interference (see also Martin-Michiellot and Mendelsohn (2000) who gave up using this secondary task in their replication study).

Subjective ratings. Cerpa et al. (1996) not only measured studying times (see above), but also asked students after learning to rate how difficult it was to understand the materials. Unfortunately, only efficiency measures (the combination of difficulty ratings and learning outcomes) were reported. These measures showed that students with the integrated manual on the screen had higher efficiency than students with a conventional manual plus computer. It stays, however, unclear whether this difference in the efficiency measures was also reflected in the simple difficulty ratings. A difference in perceived difficulty would have been a rather direct measure of extraneous cognitive load, whereas the efficiency measure is also influenced by learning outcomes, and thus, cannot be taken as clear evidence in favor of the extraneous cognitive load explanation. In another study on electrical circuits presented in separated or integrated format or without text (Experiment 1), Kalyuga, Chandler, and Sweller (1998) measured students' perceived difficulty after each problem solving task during the acquisition phase. Students with separated format rated the problems more difficult than students with integrated format (students with diagram only had the highest difficulty ratings). Kablan and Erden (2008) yielded similar results. Although these results corroborate the extraneous cognitive load explanation, they are no evidence against the germane cognitive load explanation, because perceived difficulty is thought to measure extraneous but not germane cognitive load separately.

Behavioral activities. Some years after the first split-attention experiments, Sweller and Chandler (1994) extended their argumentation of mental integration by referring to perceptual processes of visual search. They assumed that the act "… of mental integration involves finding relations among elements associated with the diagram and statements. Unless the relevant relations among the elements are found, the instruction will be unintelligible. Finding relations among elements requires cognitive resources that must be expended…" (p. 192-193). Later on, several authors, although favoring the germane cognitive load explanation, elaborated this assumption and argued that finding relations or extraneous cognitive load is reflected in learners' viewing behavior. Erhel and Jamet (2006) as well as Tabbers et al. (2000) assumed that learners with separated format are forced to switch very frequently back and forth between text and illustration what causes high extraneous cognitive load. However, these authors did not measure learners' viewing behavior to test these assumptions.

An approach to investigate the mechanism of the split-attention effect on a process level was used by Mwangi and Sweller (1998). Mwangi and Sweller (1998) asked young students to self-explain during solving arithmetic word problems. The authors did not only find more wrong inferences from students with separated format but also more correct inferences from students with integrated format. Furthermore, they demonstrated that students with separated format made more simple rereads without connecting the information to solutions. On the other side, students with integrated format linked their rereads to solution steps more often. This result may indicate that these learners processed the material more deeply. Mwangi and Sweller (1998), however, did not argue that more correct inferences and relating read information with problem solutions represented higher germane cognitive load, although the construct of germane cognitive load was introduced in that year (see Sweller, et al., 1998). Today, Sweller (in press) still states that extraneous cognitive load is the main source of the split-attention effect. Because the relations between the behavioral activities measured (e.g., rereads, inferences) and the three cognitive load types were not explicitly defined, the results by Mwangi and Sweller (1998) stay rather ambiguous.

More information based on learners' behavioral activities was provided by Martin-Michiellot and Mendelsohn (2000) in their replication study of the one by Chandler and Sweller (1996). Interestingly, Martin-Michiellot and Mendelsohn collected information about learners' computer-based navigation behavior during learning a computer program by means of log file data. They found that students with separated format (manual plus computer) explored commands of the program that were not part of the manual, thereby distracting themselves from the original learning tasks. The authors did not discuss whether such additional processes that are not directly linked to the learning task but nevertheless informed learners about other functions of the program should be defined as extraneous cognitive load or not. Unfortunately, it is not known whether the participants in the original study by Chandler and Sweller (1996) also explored commands that were not part of the learning task and knowledge tests respectively.

Additional evidence by alternative instructional formats. The extraneous cognitive load explanation was also favored by Bodemer, Plötzner, Feuerlein, and Spada (2004). These authors did not measure cognitive load, but used another instructional format that should underline their extraneous cognitive load argumentation. They argued that integrated formats have the potential to reduce extraneous cognitive load but that integrated formats, nevertheless, do not support meaningful learning processes or germane cognitive load. Bodemer et al. (2004) even assumed that learners with integrated formats stay passive, and therefore, do not invest germane cognitive load. Thus, they constructed an alternative instructional design to foster germane processing. They asked students with a separated format of a computer-based learning environment to actively integrate text into graphics by dragging and dropping text elements, an activity which should increase germane cognitive load. In Experiment 1, they showed that students with both the integrated format of the material and the active integration format outperformed students with conventional separated format. In some knowledge tests, students with active integration format even outperformed students with the integrated version. According to the authors, this result shows that integrated formats improve knowledge by reducing extraneous cognitive load and that active integration formats push knowledge acquisition even further by activating germane cognitive load. In Experiment 2, students with integrated format did not outperform students with separated format. Hence, there was no split-attention effect, but students with the active integration format outperformed students with integrated and separated format. The finding that the authors did not find the split-attention effect of separated and integrated format is at odds with all other experiments demonstrating the split-attention effect (cf. Martin-Michiellot & Mendelsohn, 2000). Nevertheless, the authors argued that learners with integrated format processed the materials rather passively with reduced extraneous cognitive load and without or reduced germane cognitive load. Because this experiment did not show the split-attention effect, it seems difficult to use it as clear evidence that learners with integrated formats benefit from reduced extraneous cognitive load only. The idea that learners with integrated formats stay rather passive is compatible with the extraneous cognitive load explanation, but is in direct contrast to Mayer's (cognitively) active-processing assumption stating that learners with integrated formats engage in germane processing.

4.1.2 Germane Cognitive Load Explanation: Promoting Mental Integration Processing

There are only few experiments arguing that germane cognitive load plays an important role in causing the split-attention effect and most of these studies did not collect any cognitive load data.

Learning outcomes. As outlined in Chapter 2.2, Mayer (1989; Mayer et al., 1995) first explained the split-attention effect by referring to the active-processing assumption, that is, to constructive cognitive processing (mentally organizing words and images, and integrating mental representations of words and images in working memory). These processes can be interpreted as germane cognitive load. Later on, he adopted the extraneous cognitive load explanation (e.g., Moreno & Mayer, 1999). No matter which explanation Mayer favored, he did not collect any cognitive load or cognitive process measures in his experiments on the spatial split-attention effect. Therefore, he offers no empirical cognitive load measure for neither of both explanations. Mayer (2001) based his arguments on group differences in learning outcomes of different knowledge tests: (1) retention tests to measure remembering and (2) transfer tests to measure understanding as empirical evidence for both cognitive load explanations. If learners with integrated formats outperformed learners with separated formats on problem solving tasks that ask learners to apply the information to novel situations (transfer test), this was taken as evidence that learners with integrated formats constructed a more elaborated integrated model or a better understanding of the content to be learned by the investment of germane cognitive load.

Erhel and Jamet (2006) argued that their split-attention effects might not be caused by a lack of split-attention in integrated formats but by an explicit reference between text and picture which obviously refers to Mayer's active-processing assumption (2001, 2005a). They further argued that the mapping process of students with the separated format "…could have led to several errors and could have been carried out at the cost of the elaboration of relevant mental representations" (p. 144). Hence, Erhel and Jamet (2006) favored the extraneous in combination with the germane cognitive load explanation. Like Mayer they did not measure cognitive load but also relied only on different knowledge scales (e.g., paraphrases test for remembering, inferences test for understanding). Florax and Plötzner (2009) argued explicitly that students with an integrated format invested more germane cognitive load besides a reduced extraneous cognitive load. However, they neither measured any cognitive load type and relied on different knowledge tests only (retention test for remembering and comprehension test for understanding).

Studying times. Pociask and Morrison (2008) also favored the germane cognitive load explanation and summarized that their "…results suggest that designers can increase the germane cognitive load by reducing the extraneous cognitive load through good instructional and message design practices" (p. 379). Nevertheless, they hypothesized that learners with separated format would have longer studying times than learners with integrated format. This hypothesis refers to the extraneous cognitive load explanation only. They did not find any difference in studying times between the groups and argued that the learners with separated format just did not invest enough effort. Actually, this interpretation is not in line with their extraneous cognitive load hypothesis but it is a good example of the difficulties in interpreting studying times. An alternative interpretation might be that learners with integrated formats used the time for investing germane cognitive load, whereas learners with the separated format needed their time for extraneous cognitive load. This interpretation might also hold true for the non-significant difference in studying times of practice problems in the study by

Kester et al. (2005). These results are only weak evidence in favor of the germane cognitive load explanation, but they can be interpreted as possible evidence supporting the germane cognitive load hypothesis (see Table 3) or as evidence against the extraneous cognitive load hypothesis.

Subjective ratings. Tabbers, Martens, and Van Merriënboer (2000) were the first who explicitly mentioned germane cognitive load within a CLT explanation of the split-attention effect and provided at least ambiguous empirical evidence for this explanation. They manipulated split-attention by color-coding text-graphic relations and modality in an instruction about an instructional design model. Moreover, they asked students to rate their mental effort. Concerning learning outcomes, Tabbers et al. (2000) found a split-attention effect (but no modality effect). However, the groups did not differ in their mental effort ratings. Tabbers et al. argued that the mental effort scale measured total cognitive load. Therefore, the authors interpreted that the ratings of students with a separated format represented high extraneous cognitive load but at the same time low germane cognitive load, whereas the ratings of students with an integrated format represented low extraneous cognitive load but at the same time high germane cognitive load. A similar pattern of result and argumentation was presented by Kester et al. (2005) who investigated spatial contiguity in the design of just-in-time presented information in problem solving tasks on electrical circuits. Whereas Tabbers et al.'s argumentation was made post-hoc, because they actually expected students with the separated format to have higher mental effort ratings, Kester et al. (2005) argued already a priori. The non-significant mental effort ratings might represent the different load type patterns. However, this non-significance of mental effort ratings is only weak evidence in favor of the germane cognitive load explanation, because mental effort ratings are a measure of total cognitive load but not of the individual load types. Pociask and Morrison (2008) did not only measure studying times (see above) but also asked students to rate how difficult the instruction was to understand. Students with the separated format rated it more difficult to understand the instruction than students with the integrated format. However, if difficulty ratings are a measure of extraneous cognitive load as suggested in Chapter 3.2, the result can be used as evidence in favor of the extraneous cognitive load explanation only but not as clear evidence supporting the germane cognitive load explanation.

4.2 Quantitative Summary of the Review

In contrast to Ginns' (2006) meta-analysis that concentrated on the quantitative aspect of the split-attention effect, the above presented review aimed at analyzing and evaluating the empirical evidence of different cognitive load explanations for the split-attention effect. Figure 9 depicts the quantitative summary of the review. This review yielded several findings concerning these explanations. First, there are two different explanations of why learners benefit from integrated formats. In 37 experiments (79%) the split-attention effect is explained by referring to extraneous cognitive load only (extraneous cognitive load explanation), whereas in only ten experiments (21%) the effect is explained by an increased germane cognitive load in addition to a reduced extraneous cognitive load (germane cognitive load explanation).

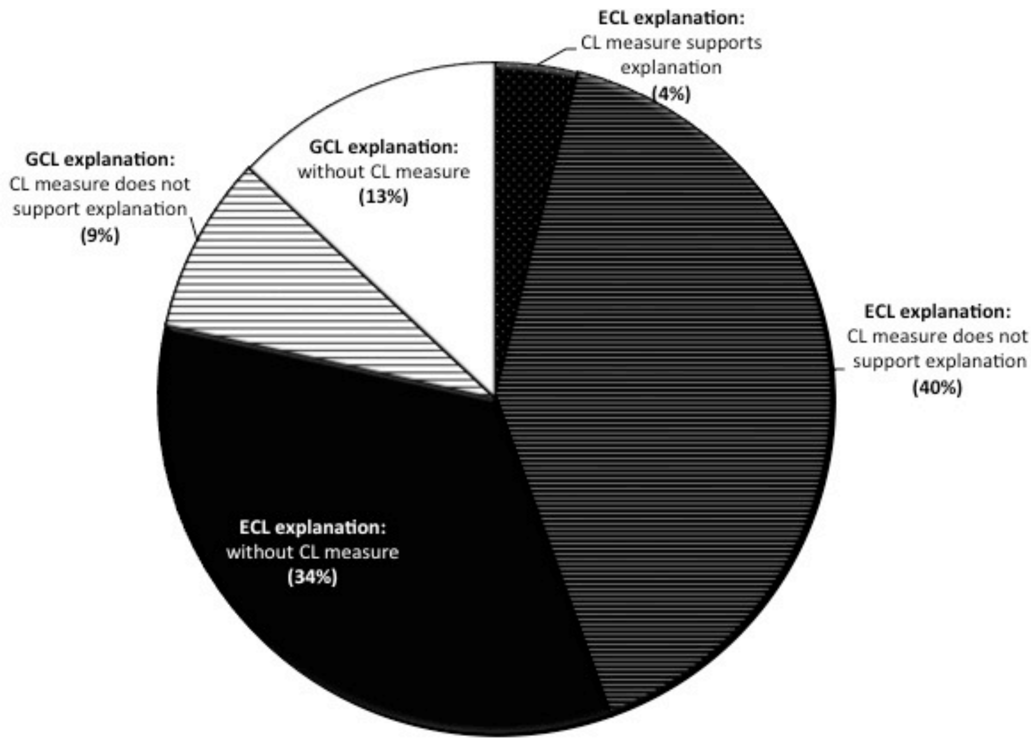


Figure 9. Quantitative summary of 47 experiments on the split-attention effect

Second, although cognitive load is assumed to be a direct mediator of spatial contiguity effects, cognitive load measures were only provided in 26 experiments (55%). Five of these experiments reported two cognitive load measures. Cognitive load measures were reported in 21 out of 37 experiments (57%) arguing for the extraneous cognitive load explanation, whereas cognitive load measures were only reported in four out of ten experiments (40%) arguing for the germane cognitive load explanation.

Third, five methods for cognitive load measurement were used. The cognitive load measures[3] were (1) behavioral activities (self-explanations and viewing behavior) in two experiments (8%), (2) secondary task performance in one experiment (4%), (3a) subjective mental effort ratings in two experiments (8%), (3b) subjectively perceived difficulty rating (or instructional efficiency based on perceived difficulty ratings) in five experiments (19%), and (3) learners' studying times in 21 experiments (81%). Some empirical evidence in favor of the extraneous cognitive load explanation was provided by 19 (79%) of the experiments providing cognitive load measures. Most of this evidence is provided by 14 experiments using studying times (58%). Notably, the more recent studies were less consistent with regard to finding longer studying times for learners with a separated format (e.g., Kablan & Erden, 2008; Rose & Wolfe, 2000). Empirical evidence in favor of the germane cognitive load explanation was more or less provided by all cognitive load measures reported in four experiments (except difficulty ratings). However, these measures provide only weak evidence because germane cognitive load was not measured explicitly.

Fourth, despite the assumption that cognitive load mediates the split-attention effect only one single study used path analyses (based on multiple regression analyses) to investigate whether cognitive load mediates the split-attention effect. Rose and Wolfe (2000) showed that the longer learners with higher problem effectiveness studied, the higher were their learning outcomes. However, they could not show that spatial contiguity influenced studying times. Despite this result which fits quite well with the assumption that studying times represented germane cognitive load, these authors assumed the extraneous cognitive load assumption. Hence, an important statistical precondition for mediation analyses was not fulfilled (for further explanations of mediation analyses see Chapter 5.2).

4.3 Conclusion and Research Outlook

According to the review, it is evident that the extraneous cognitive load explanation is meanwhile assumed by most researchers investigating the split-

[3] Because five experiments reported two cognitive load measures but the number of experiments is used as reference, the number of percentages reported does not sum up to 100%.

attention effect (cf. Mayer, 2001 vs. Mayer, 2009). Notably, early studies concentrated on higher overall cognitive load for the separated condition. Later on, this higher overall cognitive load was specified to result from higher extraneous cognitive load only. Despite this emphasize on extraneous cognitive load several researchers have also proposed the germane cognitive load explanation as complementary explanation since 2000 (e.g., Tabbers et al., 2000). This qualitative switch in the explanation also includes that overall cognitive load of learners with separated formats need not to be higher than the overall cognitive load of learners with integrated formats, because learners with integrated formats might use their working memory capacity for processes resulting in germane cognitive load.

After reviewing the empirical evidence in favor of the extraneous cognitive load explanation, it is also evident that there are only a few studies which provide some direct empirical evidence for this explanation. If one puts aside, first, the studies which used studying times as ambiguous cognitive load measure, second, studies which reported efficiency measures that are intrigued by learning outcomes, and third, the study which used secondary task performance as overall cognitive load measure but which seemed to violate important implementation demands (Chandler & Sweller, 1996), there are only two studies (out of 37) left that provided direct empirical evidence that learners with separated formats seem to experience a higher level of extraneous cognitive load than learners with integrated formats. These studies by Kalyuga et al. (1998) and by Kablan and Erden (2008) demonstrated that learners with separated formats had not only inferior learning outcomes but also higher perceived difficulty ratings than learners with integrated formats.

The analysis of the empirical evidence in favor of the germane cognitive load assumption even shows that there is no direct empirical evidence for it at all. Although in some experiments the studying times and mental effort ratings (supposed to measure total cognitive load) did not differ between learners with separated and integrated formats – which might be interpreted as indirect evidence that learners with separated formats had higher extraneous and lower germane cognitive load, whereas learners with integrated formats had lower extraneous and higher germane cognitive load – none of these ten studies used a specific germane cognitive load measure to provide direct evidence for these interpretations. Despite this lack of direct evidence for the germane cognitive load explanation, it seems, however, too premature to conclude that the split-attention effect is mediated by extraneous cognitive load only. As these are only two studies with cognitive load measures that clearly demonstrated evidence in favor of the extraneous cognitive load explanation. Moreover and even more important, there is no split-attention study so far which has tried to measure germane cognitive load separately. Until this is not done, subjective ratings of extraneous cognitive load are not sufficient to decide about the cognitive load explanations of the split-attention effect. Third, if the extraneous cognitive load explanation is correct (and germane cognitive load is not influenced by spatial contiguity manipulations (see for example, Bodemer et al., (2004)), then the overall cognitive load of learners with separated formats should be higher than that of learners with integrated formats. So far, however, there is only weak evidence for this claim, too (see Chandler & Sweller, 1996; Martin-Michiellot & Mendelsohn, 2000).

In order to decide whether the split-attention effect is mediated by extraneous cognitive load only or germane cognitive load additionally, an experiment would be needed that meets three criteria. First an interpretable measure of overall cognitive load is necessary, in order to test whether learners with separated formats suffer from higher overall cognitive load caused by extraneous cognitive load as suggested by the original extraneous cognitive load explanation.

Second, separate measures of extraneous and germane cognitive load are necessary and intrinsic cognitive load needs to be controlled. In case that the germane cognitive load explanation is correct, an overall cognitive load measure might not show differences between differently instructional formats but it does also not provide any direct information about the different load types. Furthermore, only demonstrating that learners with separated formats suffer from higher extraneous cognitive load is neither enough to corroborate the extraneous cognitive load explanation nor to falsify the germane cognitive load explanation. Rather, an individual germane cognitive load measure is needed to corroborate both explanations.

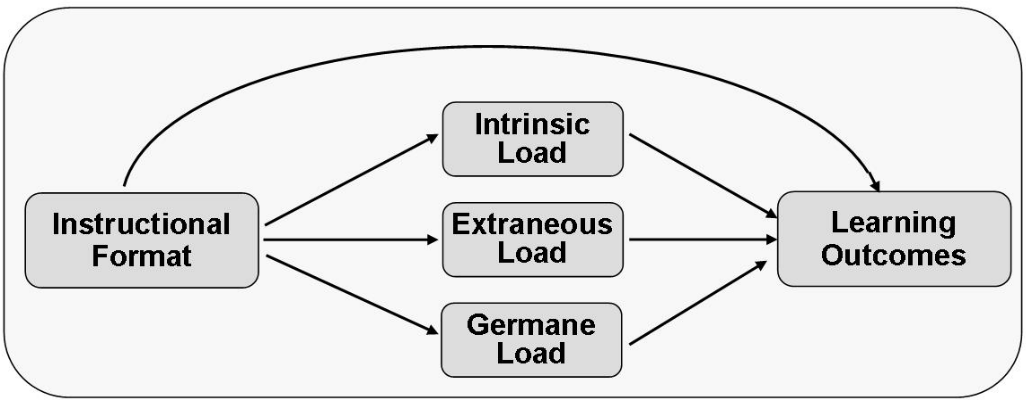


Figure 10. Path model to test the cognitive load explanations of the split-attention effect

Third, to test the assumption that cognitive load (either extraneous or germane or both) mediates the effect of the instructional design on learning outcomes, mediation analyses are needed. Figure 10 depicts the path model that has to be tested to answer the question whether the split-attention effect is mediated by extraneous cognitive load only or by an increase in germane cognitive load in addition.

So far, there has been no study that has fulfilled all these criteria to test different explanations of the split-attention effect. Therefore, Experiment 1 of this thesis was conducted in such a way that it met all three criteria. This experiment will be presented in the next chapter.

5 Experiment 1: What Explains the Split-Attention Effect?

To investigate the question how different cognitive load types mediate the split-attention effect, the following experiment was conducted. The experiment was intended to test two explanations: (1) The extraneous cognitive load explanation mainly suggested by Sweller and colleagues within the CLT (Sweller et al. 1998) and (2) the germane cognitive load explanation first suggested by Mayer (1989) within the CTML. To test both explanations explicitly, it was necessary to measure the three cognitive load types separately. Therefore, three subjective ratings scales were developed intended to measure intrinsic, extraneous, and germane cognitive load separately. To additionally provide objective cognitive load measures the dual task paradigm was applied. Beyond using multiple measures to track different cognitive load types, this experiment additionally differs from existing studies on the split-attention effect in that it applied statistical multiple mediation analyses to test the mediating role of each of the three cognitive load types. This chapter will first summarize the hypotheses derived from both explanations. Subsequently the design and experimental materials used in Experiment 1 will be described. Finally, the results of this experiment will be presented and discussed.

5.1 Hypotheses

The hypotheses tested refer to learning outcomes and cognitive load measured by secondary task performance as well as by subjective ratings for the three cognitive load types.

5.1.1 Hypotheses for Learning Outcomes

Concerning learning outcomes a split-attention effect is predicted, that is, learners with integrated format should outperform learners with separated format. This should especially be true for complex information and test items that require inferences from the learner but not for items asking for simple vocabularies ***(H 1.1)***.

5.1.2 Hypotheses for Cognitive Load

Assumptions about cognitive load are distinguished in assumptions referring to overall cognitive load and assumptions referring to each type of cognitive load. Table 4 summarizes the hypotheses that can be derived from the germane and extraneous cognitive load explanations.

Table 4

Overview of the hypotheses concerning the split-attention effect

	Germane CL Explanation	Extraneous CL Explanation
Learning Outcomes	IF > SF	IF > SF
Overall CL	IF < / = / > SF	IF < SF
ICL	IF = SF	IF = SF
ECL	IF < SF	IF < SF
GCL	IF > SF	IF = SF

Note. IF = integrated format, SF = separated format, ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

5.1.2.1 Secondary task performance: Overall cognitive load

Whereas the extraneous cognitive load explanation predicts that learners with separated format suffer from higher overall cognitive load, the germane cognitive load explanation assumes that learners with separated format need not differ in overall cognitive load from learners with integrated format, because learners with integrated format who do not suffer from extraneous cognitive load may invest more germane cognitive load. Therefore, according to the extraneous cognitive load explanation learners with separated format should perform worse in a secondary task than learners with integrated format ***(H 1.2.1a)***. According to the germane cognitive loads explanation, however, learners with separated format need not differ in secondary task performance from learners with integrated format ***(H 1.2.1b)***.

5.1.2.2 Subjective ratings: Cognitive load types

Concerning intrinsic and extraneous cognitive load the two explanations of the split-attention effect predict the same load pattern, however, concerning germane cognitive load both explanations lead to different predictions.

Intrinsic cognitive load. Both the germane and the extraneous cognitive load explanation predict that learners with separated format should experience the same amount of intrinsic cognitive load like learners with integrated format. Hence, there should be no differences in intrinsic cognitive load ratings ***(H 1.2.2a)***.

Extraneous cognitive load. The extraneous cognitive load explanation predicts that learners with separated format should suffer from higher extraneous cognitive load, and thus, should differ in extraneous cognitive load ratings from learners with integrated format ***(H 1.2.2b)***. Furthermore, the extraneous cognitive load ratings should mediate learning outcomes ***(H 1.2.2c)***. The germane cognitive load explanation shares these predictions.

Germane cognitive load. According to the germane cognitive load explanation, learners with integrated format should rate their germane cognitive load higher than learners with separated format ***(H 1.2.2d)***. Furthermore, the germane cognitive load ratings should mediate the learning outcomes ***(H 1.2.2e)***. Finally, germane cognitive load should be negatively correlated with extraneous cognitive load ***(H 1.2.2f).*** On the contrary, the extraneous cognitive load explanation assumes that there should be no differences in subjective ratings of germane cognitive load between learners with separated and integrated format ***(H 1.2.2g)***.

5.1.3 Hypotheses for Control Variables

Cognitive variables. Participants' cognitive ***learning prerequisites, domain,*** and ***topic prior knowledge*** were obtained to test whether they do not differ between groups to ensure the internal validity of the experiment.

Motivational variables. In addition to the cognitive variables, participants' ***perceived task demands*** and ***interest*** were explored to investigate whether motivational aspects that have not been considered so far by the cognitive load explanations play a role in the split-attention effect. Perceived task demands (Salomon, 1984) were also obtained to test that they do not differ between the groups to ensure the internal validity of the experiment. Moreover, it was explored whether learners with separated format reported the same interest in the topic after learning than learners with integrated format, because interest might be a central motivational variable for learning (Hidi, 1990; Hidi, Renninger, & Krapp, 2004). For instance, Alexander Kulikovich, and Schulze (1994) demonstrated that learners who rated a text to be more interesting gained higher comprehension scores than learners who rated the text to be less interesting. It was also explored whether one of the cognitive load types was related to these motivational variables to ensure internal validity.

5.2 Methods

The following sections describe the participants, materials and procedure of Experiment 1.

5.2.1 Participants and Design

One hundred and three university students participated in the study for either payment or course credit. Two participants had to be excluded from the study, because they were already familiar with the learning content. Three students were excluded because of technical problems with the computer software used. The remaining students were 63 females and 35 males with an average age of 22.65 (*SD* = 3.88) years. Participants studied diverse subjects: Mathematics / Informatics (27), Psychology (26), Language / Linguistics (19), Politics (6), Geography (6), Sports (4), History (3), Economics (3), Biology (2), Jura (2). None of the participants was color blind or had any difficulties in distinguishing between the red and green color used for the secondary task probes (see below). Participants were randomly assigned to one out of six experimental conditions which resulted from a 2 x 3 design with instructional format (integrated vs. separated) and secondary task type (letter vs. background vs. without) as independent variables.

5.2.2 Materials

In the following the learning and test materials developed for this study are described. All experimental materials were developed by the experimenter (author of this dissertation thesis). The learning topic was the physiological functioning of the kidney which is a complex topic with high clinical relevance. Because of its complexity the intrinsic cognitive load of the learning materials should be high enough to render the instructional design relevant for learning (Sweller, 1994). Furthermore, to understand how the physiology of the kidney works, it is important to not only know which processes take place but also in which physiological structure these processes take place, how the type of structure (e.g., membrane type) enables the process and how these structures are spatially related. The topic is usually described in textbooks with text and graphic outlined in separated format. As the learning topic was a realistic content that is taught in medical education, the learning and test materials developed were checked by two medical students independently to ensure that all information was correct. Both students had already successfully passed exams about the physiology of the kidney.

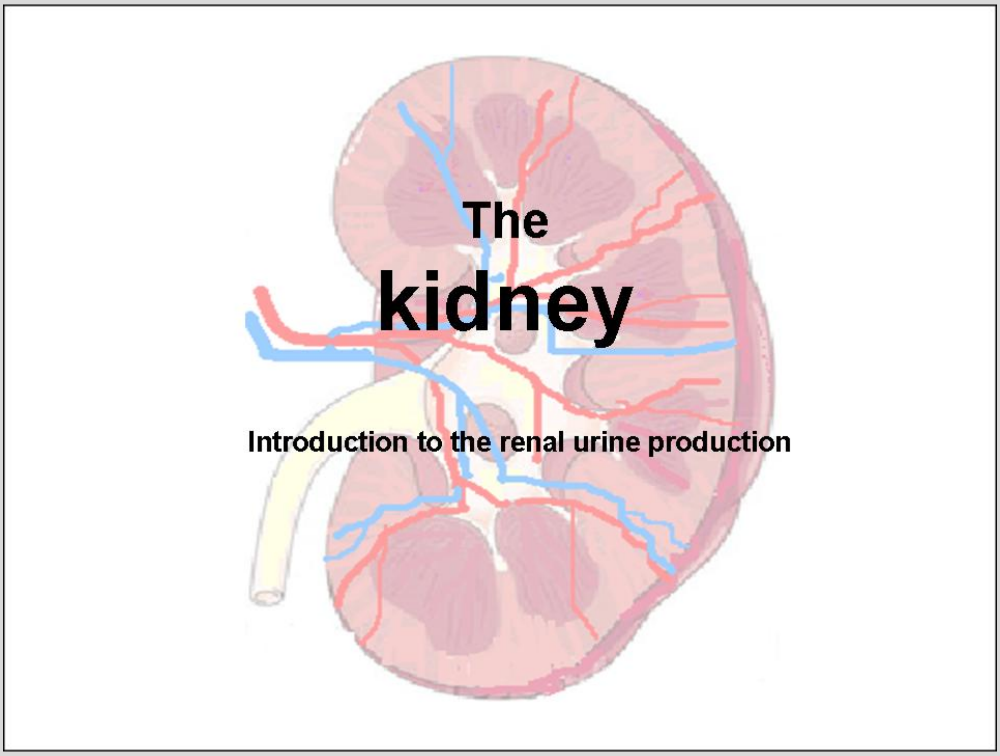
5.2.2.1 Independent Variables

Two variables were manipulated in the first experiment. First, the instructional format of the learning materials differed with respect to spatial contiguity. Second, participants had either to perform one out of two secondary tasks or not.

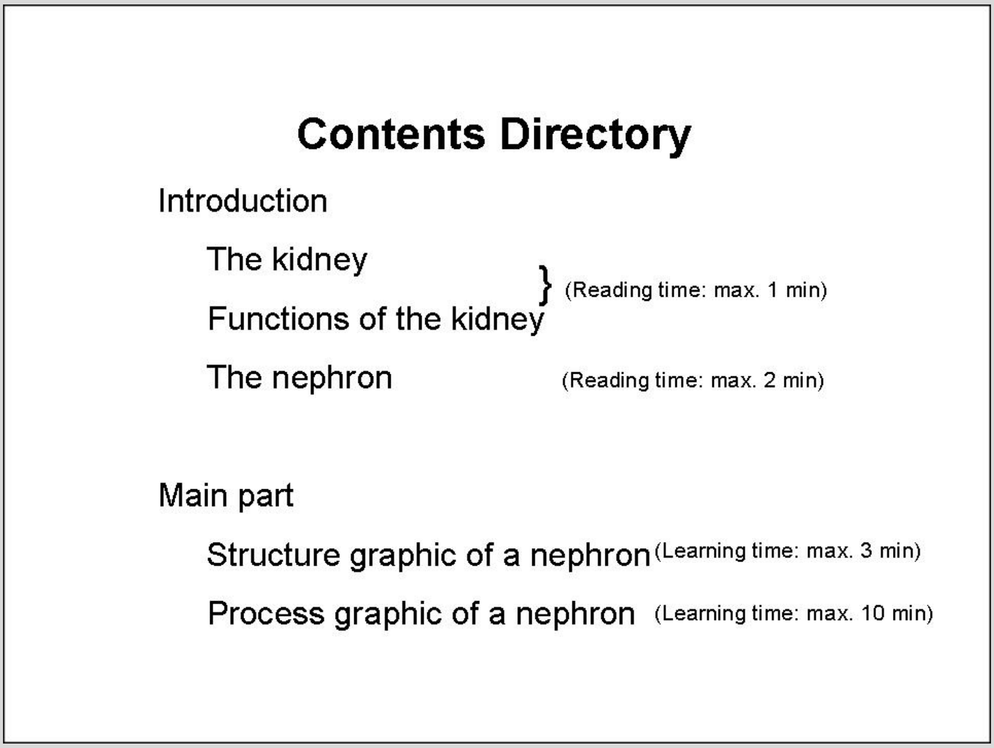
Instructional format and learning materials. The learning materials consisted of a computer-based learning environment about the physiology of the kidney developed in PowerPoint (Microsoft®) and presented with MediaLab (Empirisoft). The physiological processes taking place in the kidney were described according to the information presented by the respective chapters of four books by Golenhofen (1997), Hick and Hick (2000), Huppelsberg & Walter (2003), and Schmidt and Thews (1995). The environment was system paced and consisted of an introduction into the topic and two instructional graphics with accompanying text. The introduction was the same for all participants. It consisted of four slides presenting information about the general structures and functions of the kidney (313 words; font: Arial; size: 14). The introductory slides are presented in Figure 11 and Figure 12.

Subsequent to the introduction, the main part followed with a structure and process graphic according to Mayer and Gallini (1990) who distinguished between parts (structure) and steps (processes) in mechanical systems. The first experimental graphic consisted of a colored illustration of a nephron (functional unit of the kidney) with verbal information about its *structural parts* (46 words; font: Arial; size: 11). The terms could not be inferred from the graphic. The structure graphic was presented either in an integrated or a separated format as depicted in Figure 13. The process graphic consisted of a visualization of the *physiological processes* in the nephron accompanied by textual explanations (249 words; font: Arial; size 9). In general, the processes were visualized and explained verbally. The verbal explanation did in general not contain the name of the structure (eg., The information about the following processes was given “The hormone Adiuretin opens water channels.” without naming the physiological structure “tubulus reunions” where this process takes place). Hence, the text explaining the physiological processes alone was not self-explanatory enough to fully learn about the topic without the specific structures in mind. And because there is no standardized way to present different types of hormones and membranes, the process graphic was neither self-explanatory, although easy-to-understand symbol types were used, like arrows to indicate the flow direction, or solid or dashed lines in different colors to represent the membrane permeability, or chemical terms like

CH_4N_2O. The chemical-physiological laws of osmotic pressure and diffusion could not be visualized but only their consequences (e.g., water efflux, flow of CH_4N_2O) were represented visually and described verbally. Overall, the process graphic was not self-explanatory. It was presented either in integrated or separated format as depicted in Figure 14.



(a) Cover



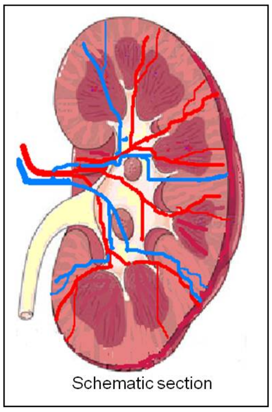
(b) Contents directory

Figure 11. Illustrations of the first introduction slides of the learning material (English version)

The kidney

The kidneys (*nephros*) are paired organs,
and each person has normally two kidneys (left and right kidney).

A kidney is bean-shaped and has a length of 10 - 12 cm.
The weight varies between 120 and 200 g.
The kidney is divided into the cortex of the kidney (external tissue)
and the renal medulla (internal tissue).



Main function of the kidney

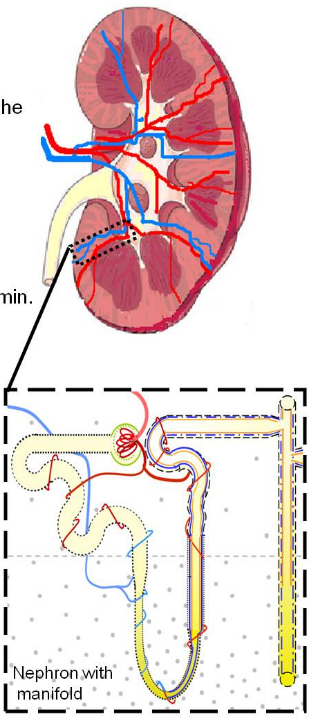
Besides the kidney`s influence on the blood pressure and
the skeletal structures, one main function is the excretion of
metabolic end-products (e.g. urea), impurities (e.g. drugs) and spare water.
A second main function is to retain essential substances
like electrolytes, amino acids and water.
The kidney is therefore responsible for an appropriate chemical composition of the blood.
This task is accomplished by means of the production of urine.

(a) General information about the kidney

The nephron

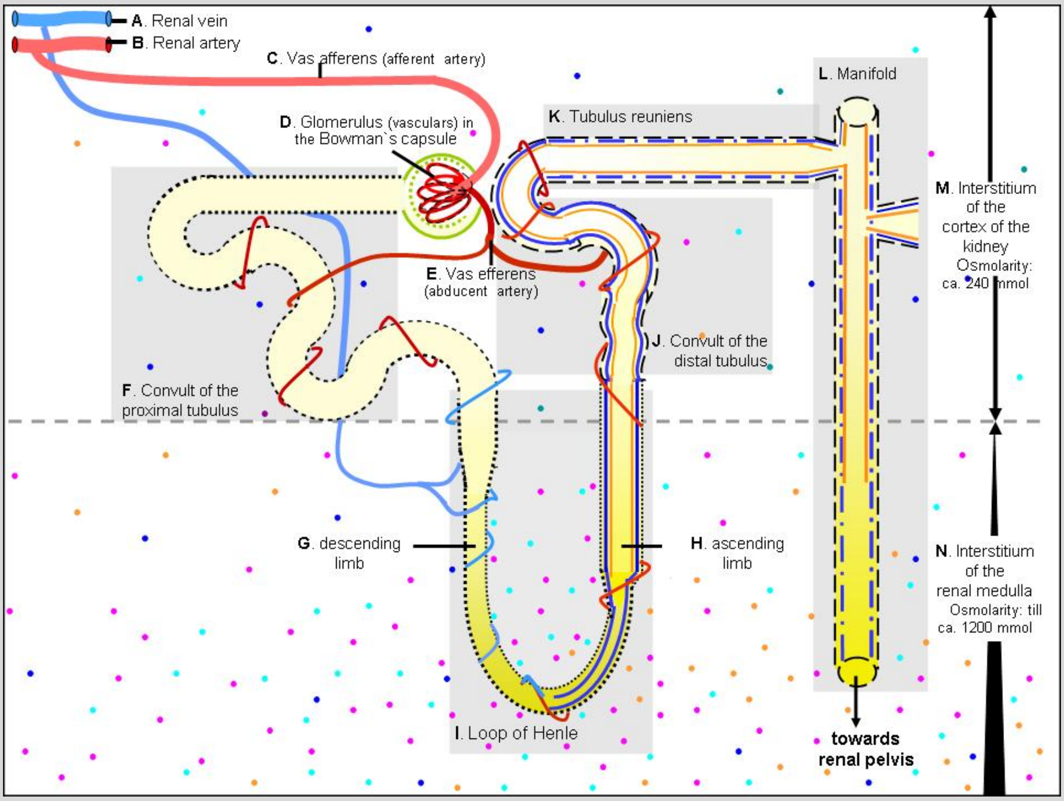
The nephron is the functional unit of the kidney which is responsible for the
production of urine.
Each kidney contains about 1 mio nephrons. All nephrons of a person
produce about 180 Liter primary urine per day.
This corresponds to a filtration rate of the blood through the nephrons of
about 120 ml/min.
To assure this amount, the blood flow through the kidneys is about 1,2 l/min.
The perfusion of the kidneys is therefore much higher than
the perfusion of the heart, the brain or the liver.
However, in the end, only 1% of the primary urine are excreted as urine.
This is about 1 - 2 l/day.
Depending on the ingestion and the imbibition the urine can be
more (<u>**hyper**</u>osmolar), equal (<u>**iso**</u>osmolar) or less (<u>**hypo**</u>osmolar)
concentrated than the blood plasma.

Each nephron consists of a glomerulus in the Bowman`s capsule
and tubulusapparatus(*tubulus,* lat. tube).
The tubulusapparatuses of several nephrons end in the manifold.
The urine flows from the manifold to the bladder.
The bladder collects the urine and finally excretes it.

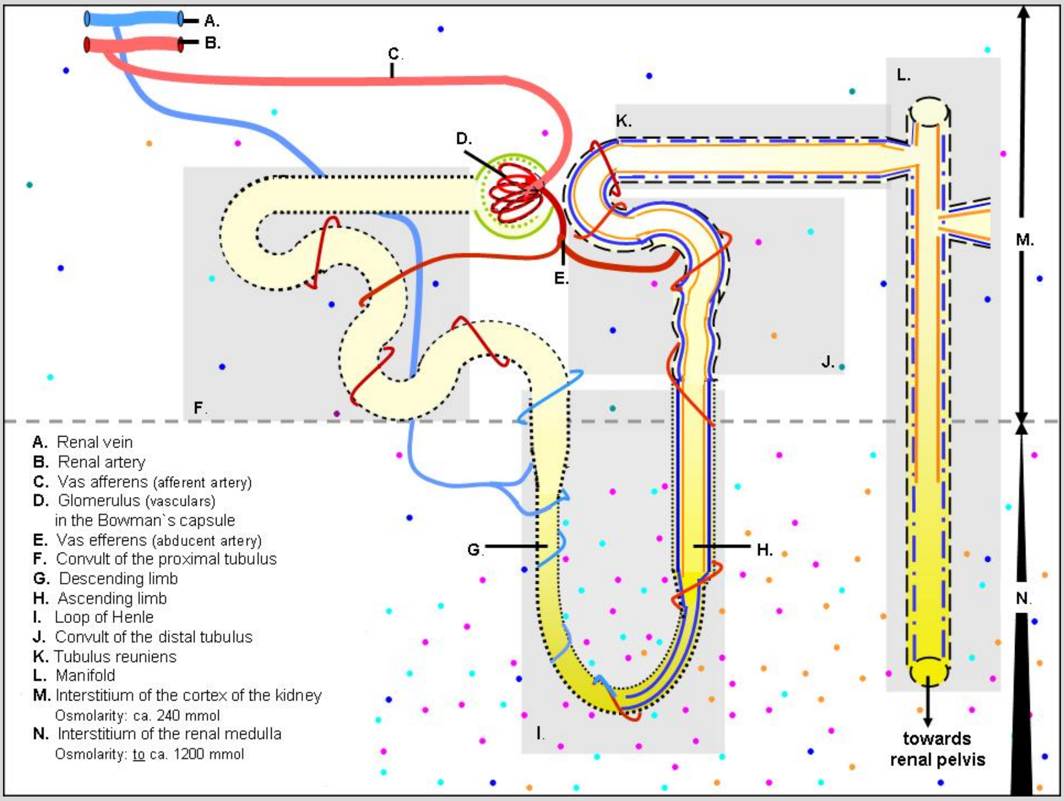


(b) General information about the nephron

Figure 12. Illustrations of the last introduction slides of the learning material (English version)

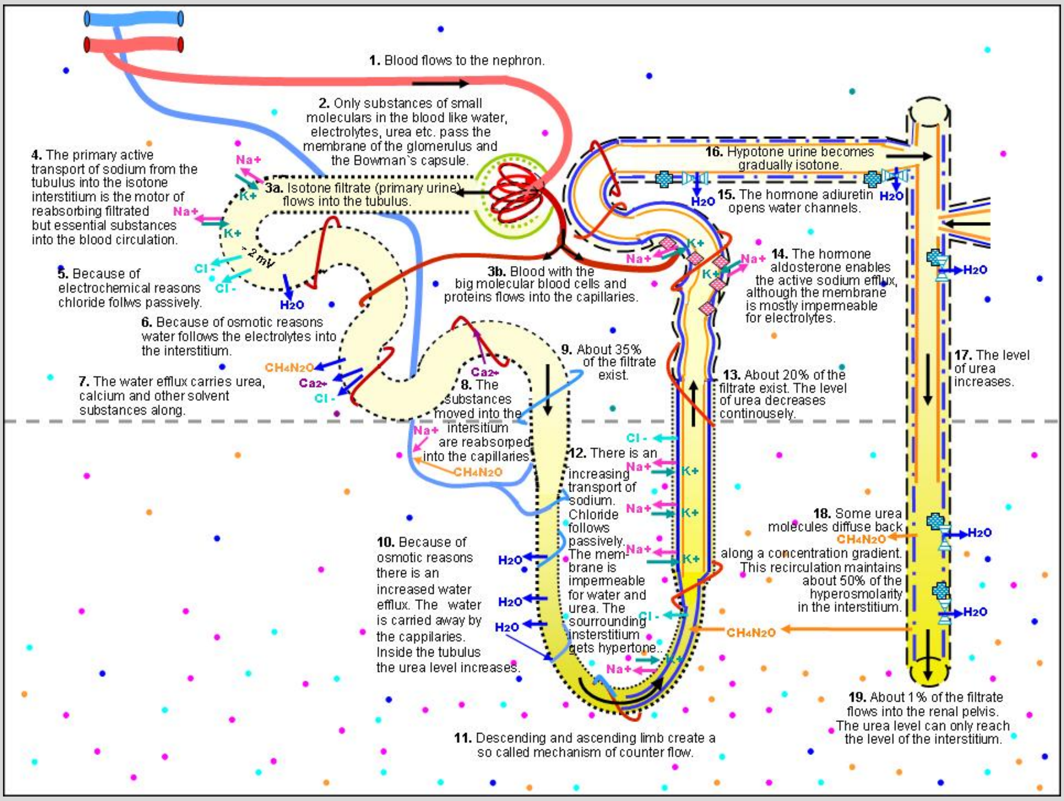


(a) Integrated format of the structure graphic

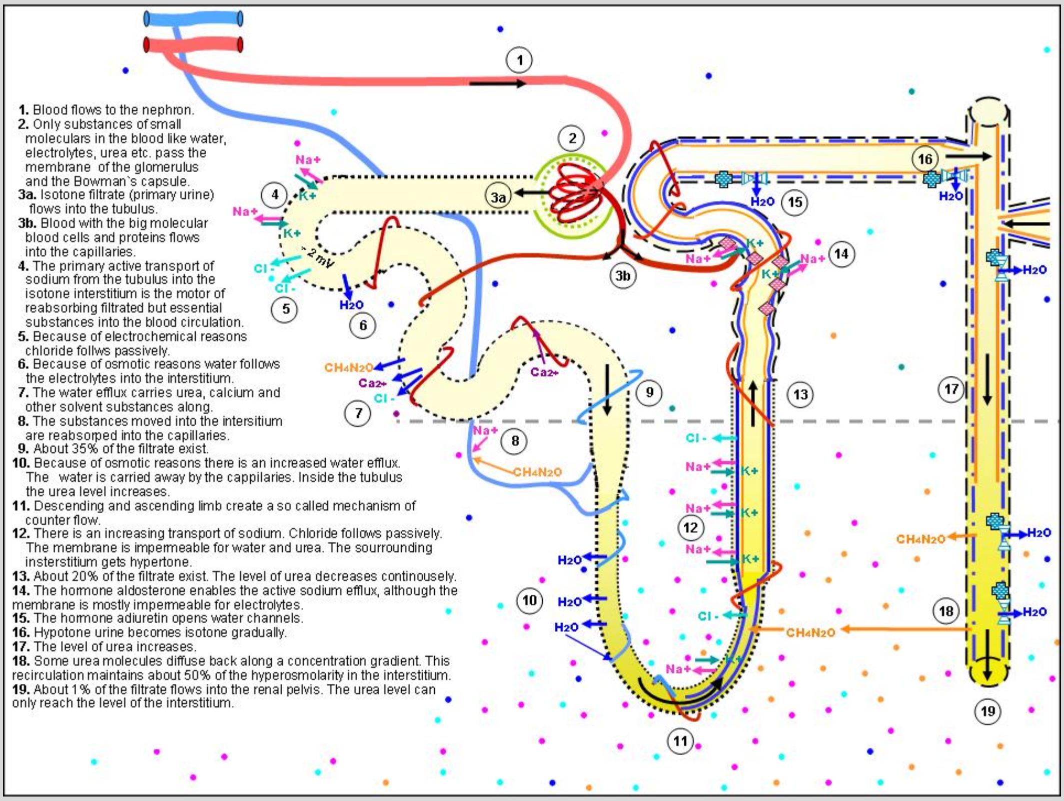


(b) Separated format of the structure graphic

Figure 13. Illustrations of the structural graphic of the learning material (English version)



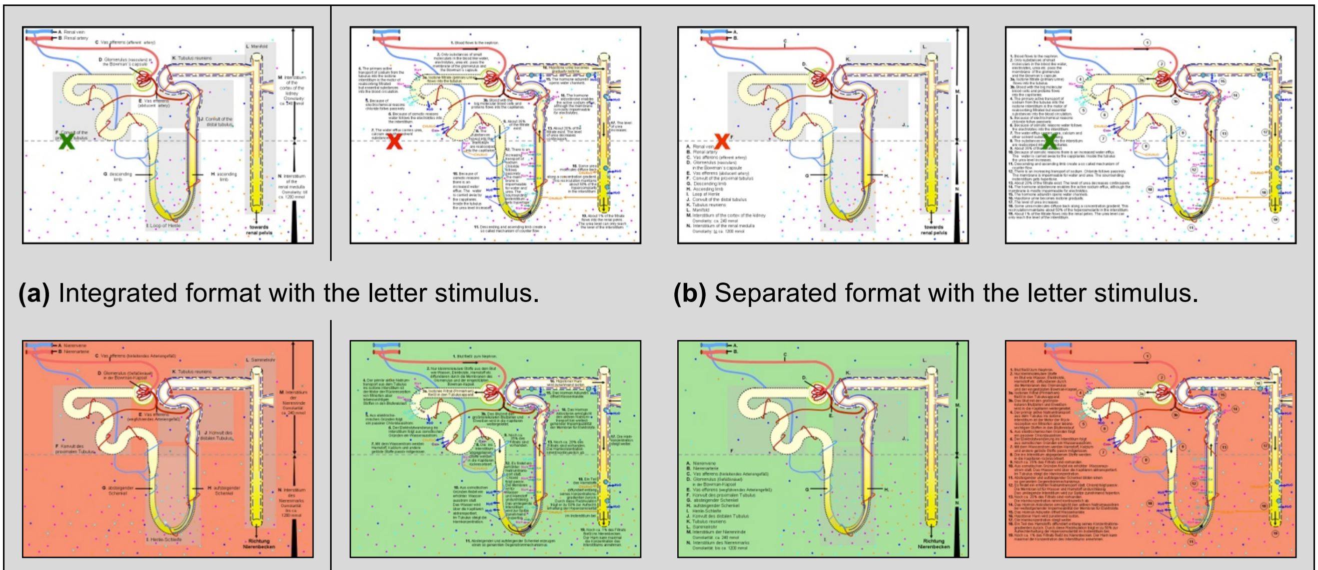
(a) Integrated format of the process graphic



(b) Separated format of the process graphic

Figure 14. Illustrations of the process graphic of the learning material (English version)

Secondary task: In creating a secondary task to measure participants' overall cognitive load during learning the following criteria were considered. First, performing the secondary task should not influence learning. Hence, the secondary task should be relative simple to avoid learning intrusion. Second, the perception of the stimulus should be equally manageable for both instructional format groups to not favor one over the other (see Chapter 3, structural interference). Third, the task should be similar to tasks already used in the cognitive load literature to produce results that can be compared with existing studies. Therefore, a color change task was chosen with two types of secondary task stimulus (see Figure 15). Learners had to press the space bar as fast as possible, if the stimulus was shortly presented (250 ms) in green color but not in red color. The task consists of realizing the stimulus, encoding the color, deciding whether to press the space bar or not, and to press the space bar or not. To perform the task best, learners had to find an optimal balance between speed and correctness.



(a) Integrated format with the letter stimulus. **(b)** Separated format with the letter stimulus.

(c) Integrated format with the background stimulus. **(d)** Separated format with the background stimulus.

Figure 15. Illustrations of the structure and process graphic with the letter and the background secondary task stimulus

(1) **Letter stimulus.** This stimulus was similar to the one used by Brünken et al. (2002). Here, the letter X (font: Arial black; size: 54) appeared either in green or red color on the left side of the learning materials (see Figure 15a and 15b). This place was chosen to minimize possible perceptual interference with the target letter's color and the colored graphic of the nephron. However, it could not be ruled out that this place caused higher perceptual interference either for the integrated or the separated

format, because of the different information presented on this place, and thus, the letter could be located either more closely or farther to learners' current visual attention depending on the instructional format. Therefore, an alternative secondary task was created.

(2) **Background stimulus.** The other secondary task stimulus was similar to the one used by DeLeeuw and Mayer (2008). The white background of the whole instructional materials changed its color to either green or red (see Figure 15c and 15d). All participants should be equally able to respond to the color change of the background, no matter on which information they looked at the specific moment of the color change. This secondary task stimulus should not interfere differently with learning on a perceptual level across the instructional formats.

To test whether the secondary task performance did not intrude learning, control conditions of both instructional format conditions without any secondary task were implemented.

5.2.2.2 Dependent Variables

Participants' learning outcomes, secondary task performance, and cognitive load ratings were the most important dependent variables. All measurement items were presented via MediaLab. To answer the test items participants had to click on labeled buttons.

Learning outcomes. To measure participants' learning outcomes, four computerized knowledge tests (terminology, labeling, complex facts, and transfer) were used. The same tests were used to measure learners' prior knowledge.

(1) The ***terminology*** test consisted of nine multiple-choice items conveying terms used to describe the structure of a nephron. The answers had to be selected out of four alternatives by clicking on the respective button (see Table 5).

Table 5

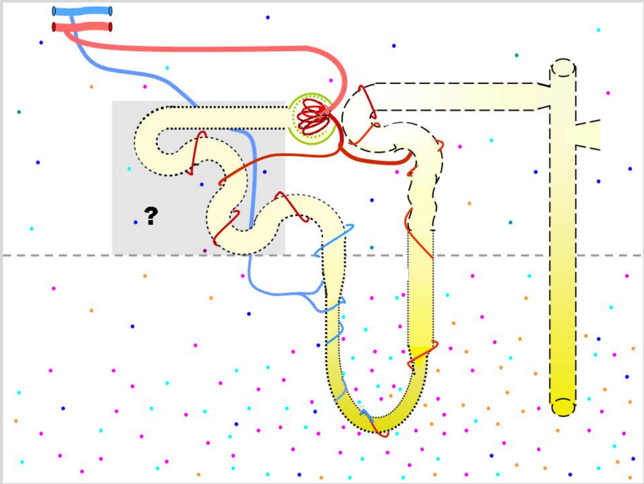
Item example of the terminology test

Which of the following terms names a structure of a nephron or its surrounding?			
a) Globulus	b) Glomulus	c) Glomerulus	d) Glemerulus

(2) The ***labeling*** test consisted of 12 multiple-choice items. Participants had to chose one out of twelve possible structure terms that matched the high-lighted part in a given graphic depcting a nephron by clicking on the respective term (see Table 6).

Table 6

Item example of the labeling test

Which structure is highlighted by the grey color?	
1.Renal medulla 2. Loop of Henle 3. Distal convult 4. Vas afferens 5. Manifold 6. Bowman's capsule 7. Ascending limb 8. Renal cortex 9. Tubulus reuniens 10. Proximal convult 11. Vas efferens 12. Glomerulus	

(3) The test on ***complex facts*** consisted of 22 sentences about the physiological processes in a nephron. Participants had to state whether these sentences were either correct or incorrect by clicking on either yes or no (see Table 7).

Table 7

Four item examples of the test on complex facts

Which of the following sentences are correct?
1) The Vas efferens transfers blood from the glomerulus into the capillaries. (Yes / No)
2) There is an increased sodium efflux in the ascending limb of loop of Henle. (Yes / No)
3) The membrane of the ascending limb is impermeable for water. (Yes / No)
4) The hormone adiuretin opens water channels in the descending limb and thus ensures that there is an increased water efflux. (Yes / No)

(4) The ***transfer*** test consisted of 20 sentences describing causes and effects in a nephron, which demanded participants to draw inferences. Again, participants had to state whether these sentences were either correct or incorrect (see Table 8).

Table 8

Four item examples of the transfer test

	Which of the following sentences are correct?
1)	If proteins are found in the urea test of a patient, a defect in the vas efferens can be assumed. (Yes / No)
2)	An artificial addition of adiuretin in the proximale convult can increase the concentration of urine by opening more water channels. (Yes / No)
3)	The filtering process of the kidney can be described as follows: only molecular components not needed by the body (e.g. urea) are filtered out of the blood by the glomerulus, and then, these molecules are transported through a tubulus system to the renal pelvis. (Yes / No)
4)	4. The longer the loop of Henle is in a nephron, the higher gets the concentration of urine. (Yes / No)

Secondary task performance. The above described secondary task stimuli appeared for 250 ms at random during intervals of ten seconds. The participants in the dual task conditions were asked to press the space bar as fast as possible, whenever a green stimulus appeared. They were supposed to ignore stimuli in red. Two types of secondary task performance were measured. First, participants' percentages of hits were calculated. Second, the reaction times to the hits (visual stimuli in green color) were measured. Participants' reactions were recorded via the probe task tool of MediaLab (Empirisoft) using Microsoft's DirectX®.

Cognitive load types. To measure the three types of cognitive load, three subjective rating scales with a labeled six-point Likert-type scale were used ranging from "not at all" (1 point) to "extremely" (6 points).

(1) ***Intrinsic*** cognitive load scale: *"How difficult was the learning content for you?"*

(2) ***Extraneous*** cognitive load scale: *"How difficult was it for you to learn with the material?"*

(3) ***Germane*** cognitive load scale: *"How much did you concentrate during learning?"*

5.2.2.3 Control variables

To control for and to explore possible cognitive and motivational influences, participants' cognitive learning prerequisites, domain and topic prior knowledge as well as their perceived task demands, and situational interest were measured.

Cognitive processing capacity. To measure participants' learning prerequisites the processing capacity sub-scale of the BIS-4 intelligence test was used in a paper-pencil version (Jäger, Süß, & Beauducel, 1997). The processing capacity consists of six sub-scales. Two of them measured verbal ability by asking persons to find correct word analogies and to distinguish between opinions and facts. Two scales measured figural ability by asking persons to find correct figural analogies and to draw the next step in a logical sequence of changing abstract figures. Two scales measured numerical ability by asking persons to write down the next number in a logical sequence of numbers and to estimate the correct result of complex formulas without exact calculation.

Domain knowledge in physiology. To measure participants' domain prior knowledge in physiology, they were asked whether 18 sentences about physiological processes were either correct or incorrect (see Table 9).

Table 9

Four item examples of the physiology test

Which of the following sentences are correct?
1) The sodium-potassium pump transports sodium out of the cell. (Yes / No)
2) During osmosis particles of a substance diffuse from areas with a higher concentration of the substance to areas with a lower concentration of the substance. (Yes / No)
3) A liquid is said to be isotonic compared to another liquid, if both have the same electrolyte level. (Yes / No)
4) The osmolarity indicates the number of active osmotic particles per liter dissolution. (Yes / No)

Perceived task demands. To measure participants' perceived task demands which might influence their actual investment of mental effort or their subjective ratings of cognitive load, participants were shown the process graphic in separated format (A) next to the process graphic in integrated format (B). The graphics were presented on the screen via MediaLab in such a resolution that the text could not be read. Participants were first asked "How hard do you think it is to learn with format A?" and then "How hard do you think it is to learn with format B?". A labeled six-point Likert-type scale was used ranging from “not hard” (1 point) to “extremely hard” (6 points).

Interest. Participants’ interest was measured by means of a labeled six-point Likert-type scale ranging from “not at all” (1 point) to “extremely” (6 points) and asking *“How interesting did you find the learning content?”*. Interest will be included in the mediation analyses to test for motivational effects.

5.2.3 Procedure

The study consisted of four phases: An initial phase of collecting general personal data (age, major, color blindness), measuring cognitive learning prerequisites (processing capacity) as well as prior knowledge in physiology, then a subsequent learning phase, a phase to rate cognitive load items, and a final post-test phase. Participants were either run in individual or pair-wise sessions. All sessions were run by one experimenter. At the beginning of the experiment the BIS-4-S was administered as paper-pencil test. Afterwards, the participants answered a computer-based knowledge test about physiological issues in general (domain knowledge) and the four knowledge tests about the nephron (terminology, labeling, complex facts, and transfer as topic knowledge). After pre-testing, participants’ baseline rates concerning reaction times for the color change task were measured in the dual task groups. Participants in the dual task groups had to react as fast as possible to a green stimulus and to ignore the red version of the stimuli for 3 minutes. Depending on the type of dual-task stimulus, the baseline stimuli consisted either of a short change (250 ms) of a black X in the middle of the screen to green or red or of a short change (250 ms) of a white screen to green or red, respectively. After the measurement of the baseline rates, participants were instructed to learn as intensive as possible, and, when they were in one of the dual task groups, at the same time to react again as fast as possible, whenever a stimulus (either X or background) in green color appeared in the learning materials. Participants were supposed to ignore stimuli in red color. All secondary task stimuli appeared for 250 ms at random during intervals of ten seconds. The green and red stimuli were equally distributed. The introduction of the learning phase lasted 3.5 minutes. Subsequent to the introductory followed the structure graphic (either in integrated or separated format) that was presented for 3 minutes and afterwards the process graphic that was presented for 10 minutes. After the learning phase, students had to rate the cognitive load type scales, then they had to answer the same four knowledge tests again (terminology, labeling, complex facts and transfer). Because knowledge retrieval is related to changes in memory awareness (Conway, Gardiner, Perfect, & Cohen, 1997), participants had to rate their confidence about the correctness of their answers on a five-point Likert-type scale ranging from "guessed" (0 point) to "very sure" (4 points) after each test item.

5.2.4 Data Analysis

Before the data were analyzed, several variables had to be computed from the raw data collected during the experiment. In the following it is described how these variables were computed. Moreover, an overview of the statistical analyses used to test the meditational role of the cognitive load types is given.

5.2.4.1 Learning outcomes

For each knowledge test item answered correctly participants were assigned one point, whereas zero points were given in case of a wrong answer. The answers to all test items were weighted with participants' confidence ratings ranging from 0 ("guessed") to 4 ("surely known") and concerning the correctness of their response, by multiplying both scores. Only correct answers (1 point) were weighted. If participants gave a wrong answer (0 points) it resulted in 0 points, no matter how sure they had stated to be, thereby ensuring that no negative knowledge outcome was generated. If participants guessed correctly and stated that they had guessed (0 points), their answer was multiplied by 0 thereby resulting in 0 points. The more surly participants stated that they knew a correct answer, the higher the knowledge score of this answer (1-4). By taking participants' knowledge consolidation into account, the nominal items were transformed into metrical ones. Based on the products, the percentage of the maximal score was determined for each participant on each knowledge test. Using the confidence information of participants' knowledge retrieval has several advantages. First, the problem of guessing probability (50% guessing probability in the test on complex facts and transfer) and rather easy recognition items (test on labeling) is bypassed. Second, the consolidation aspect of knowledge acquisition is taken into account (Conway et al., 1997), that is, answers that participants feel more confident about, are assumed to reflect knowledge that is already better retrievable. Third, nominal items are transformed into metrical ones. All three aspects should increase reliability.

5.2.4.2 Secondary task performance

Two types of parameters were generated from participants' reactions to the secondary probes. First, the percentage of hits (pressing the space bar when a green stimulus appears) participants made in performing the secondary task was computed during the phases of (a) the baseline (without learning), (b) reading the introduction, (c) learning the structure graphic, and (d) learning the process graphic. Second, the mean reaction time of these hits of each person and phase was computed.

5.2.4.3 Cognitive processing capacity

To measure participants' cognitive processing capacity, the correct answers for each sub-scale were added. All six sub-scale measures were integrated into one overall measure.

5.2.4.4 Multiple mediation analyses

Mediation analyses attempt to identify the intermediary processes that lead from the independent variable to the dependent variable. Hence, to test whether the effect of the instructional format on learning outcomes is transmitted through cognitive load, mediation analyses were used. As mediation analysis is not a standard analysis method used in the cognitive load or multimedia literature, the statistical rational of this analysis is outlined briefly. Mediation analyses mainly base on a causal steps strategy using multiple regression equations. The strategy comprises four steps in demonstrating and testing whether one or more variables (e.g., cognitive load types) mediate the relation between an independent variable (e.g., instructional format) and a dependent variable (e.g., learning outcomes; for a more detailed overview of the mediation concept see Baron & Kenny, 1986; Frazier, Tix & Barron, 2004; James & Brett, 1984; Preacher & Hayes, 2008).

Step 1 is to show that there is a significant relation between the independent variable or predictor (instructional format) and the dependent variable (learning outcomes) in the dependent variable model by using regression analysis. *Step 2* is to show that there is a significant relation between the independent variable (instructional format) and the mediator (cognitive load) in the mediator variable model. *Step 3* is to show that there is a significant relation between the mediator (cognitive load) and the dependent variable (learning outcomes) controlling for the predictor (instructional format) in the mediational dependent variable model. *Step 4* is to show that the magnitude of the relation between the independent and the dependent variable of the dependent variable model is significantly reduced when the relation is controlled for the mediator (like it is in the mediational dependent variable model). A variable is a complete mediator, if the relation between the predictor and the outcome becomes non-significant, when the mediator is included in the model, whereas a variable is a called a partial mediator, if the relation between the predictor and the outcome is still significant but becomes significantly smaller when the mediator is included in the model (step 4). In general, a variable is mediating an effect, only if the criterion for each of the four steps is met. Finally, the amount of mediation can be described in terms of the proportion of the total effect that is mediated (Shrout & Bolger, 2002).

Concerning step 4, there are two statistical ways to test whether the reduction in the magnitude of the relation is significant (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Preacher & Leonardelli, 2007). First, a significance test called Sobel test can be used. The Sobel test assumes a standard normal sampling distribution of the product term of the regression coefficients of the mediator regressed on the independent variable as well as of the dependent variable regressed on the mediator while controlling for the independent variable (Sobel, 1982). The Sobel test was developed for simple mediation. However, in this experiment three cognitive load types have to be tested by multiple mediation analyses. Because the Sobel test formula includes standard errors, it may specify the indirect effect of multiple mediations not precisely (MacKinnon, Warsi & Dwyer, 1995). Second, the method of bootstrapping can be used (Efron, 2003). Bootstrapping estimates the sampling distribution of the mediation effect non-parametrically by drawing many resamples (usually 1000 bootstrap resamples or more) randomly from the original empirical data sample and uses this information to subsequently generate confidence intervals (CIs; DiCiccio & Efron 1996; Preacher & Hayes, 2008). If the CI comprises the value of no effect, that is the difference between the measures compared (relation between independent and dependent variable with and without considering the multiple mediators) is zero, there is no mediation effect. However, the hypothesis that there is no mediation effect or that the reduction in the magnitude of the relation between independent and dependent variable in step 4 is not significant can be rejected, if the CI does not contain zero, thereby indicating significance. The advantages of using bootstrapping are that neither assumptions need to be made about the shape of the sampling distribution of the product term (as is done in the Sobel test) nor is a particular formula of the standard error required. This is especially important for rather small samples which normally violate the assumption of a normal distribution of the regression coefficients' product term and for models with multiple mediators (Preacher & Hayes, 2008). Therefore, the bootstrapping method and the use of CIs provided by Preacher and Hayes (2008) were applied in the mediation analyses because CIs also provide information on statistical significance.

5.3 Results

In the following the results of Experiment 1[4] will be presented. After checking for randomization and assignment, the learning outcomes were analyzed to probe the split-attention effect. Then, the secondary task performance and subjective load ratings were analyzed to test the explanations of the mechanism. Furthermore, detailed mediation analyses were run to test whether and which cognitive load types mediate possible instructional format effects. Additionally, several exploratory analyses concerning perceived task demands and interest serving as motivational variables were run.

For all variables tested in the following analyses, statistical distributional assumptions were tested. The Kolmogorov-Smirnov (K-S) test was used to test normal distribution assumed in all normal-theory tests (all analyses except bootstrapping analyses). Levene's test was used to test variance homogeneity across groups as assumed in ANOVAs. The Box-M-test was used to test covariance homogeneity as is assumed in RM-ANOVAs (two times of measures). Homoscedasticity as is assumed in regressional analyses was checked first, by testing the normal distribution of the residuals by means of the K-S test and second, by inspecting the scatter-plot of the residuals plotted against the predicted values. Whenever the distributional assumption was not fulfilled ($p < .05$), it is reported. Furthermore, the cause of the violation was checked and it was considered whether the violation influences the interpretability of the further test results or rather not.

For all univariate F-tests reported in this dissertation, Cohen's f is used as the effect size measure with $f = 0.10$ indicating a small effect, $f = 0.25$ indicating a medium, and $f = 0.40$ indicating a large effect (cf. Cohen, 1988). For all multivariate F-tests (including RM-ANCOVAs) reported, partial eta squared (η_P^2) is used as the effect size measure indicating the proportion of the effect plus error variance that is attributable to the effect.

[4] Parts of the results of Experiment 1 are published in Cierniak, Scheiter, and Gerjets (2009).

5.3.1 Randomization Checks and Exploratory Analyses

Before the statistical analyses of the hypotheses tests were run, it was tested whether cognitive variables like cognitive processing capacity and domain knowledge in physiology were equally distributed among the experimental groups to increase the internal validity in interpreting the effects of instructional format and secondary task on learning outcomes. Furthermore, the motivational variables perceived task demands of instructional formats and interest were explored. Whereas Table 10 summarizes the means and standard deviations of cognitive processing capacity, Table 11 summarizes the means and standard errors of the remaining control variables (that were controlled for the differences in cognitive processing capacity between the groups).

Table 10

Means and standard deviations of cognitive processing capacities as a function of instructional format and secondary task

		STS $_{Letter}$		STS $_{Background}$		STS $_{without}$	
		Integrated	Separated	Integrated	Separated	Integrated	Separated
		(*n* = 17)	(*n* = 16)	(*n* = 19)	(*n* = 17)	(*n* = 15)	(*n* = 14)
Cognitive Processing Capacity[a]	*M*	31.18	27.47	29.82	23.62	29.46	28.77
	SD	(7.74)	(7.36)	(6.57)	(8.74)	(7.35)	(8.79)

Note. [a] scale range : 0 to 54; STS = secondary task stimulus; *n* = sample size; *M* = mean; *SD* = standard deviation

Cognitive processing capacity. The IQ-subscales (cognitive processing capacity) data from five participants were incomplete because of their delayed appearance at the laboratory and therefore excluded from the analysis. Cronbachs' α of the overall scale was .68. This level is rather low and suggests that participants differ intraindividually in the different sub-scales concerning verbal, figural, and numerical tasks. A 2 (instructional format) x 3 (secondary task) ANOVA on processing capacity was run. Participants with integrated format performed better than participants with separated format ($F(1, 87) = 4.76$, $MSE = 60.45$, $p = .03$, $f = 0.23$). Participants with or without secondary task did not differ from each other ($F(2, 87) = 1.16$, $MSE = 60.45$, $p = .33$, $f = 0.15$). There was no interaction effect between instructional format and secondary task ($F < 1$). Because cognitive processing capacity (an IQ-subscale) was regarded as an important characteristic of the participants which might influence learning outcomes and because the groups differed significantly in their cognitive processing capacities, this variable was included as covariate (or predictor) into all following analyses, thereby controlling for the group differences statistically.

Domain knowledge in physiology. Cronbach's α of the prior knowledge test about physiology was .78. To test for group differences in domain knowledge in physiology, a 2 (instructional format) x 3 (secondary task) ANCOVA with cognitive learning prerequisites as covariate was run. There were no differences between participants with integrated format and separated format ($F < 1$). Furthermore, there were no differences between participants with secondary task and the participants of the control group without secondary task ($F < 1$). There was no interaction effect between instructional format and secondary task ($F < 1$). The groups did not differ in general domain knowledge about physiology and therefore there was no need to statistically control the influence of this variable. Cognitive processing capacity tended to influence the knowledge in physiology ($F(1, 86) = 3.35$, $MSE = 176.19$, $p = .07$, $\eta_P^2 = .04$).

Table 11

Means and standard errors of the remaining control variables as a function of instructional format and secondary task (controlled for cognitive processing capacity)

		STS_{Letter}		$STS_{Background}$		$ST_{without}$	
		Integrated (n = 17)	**Separated** (n = 16)	**Integrated** (n = 19)	**Separated** (n = 17)	**Integrated** (n = 15)	**Separated** (n = 14)
% Knowledge in Physiology	*M*	25.04	24.28	26.22	24.95	21.45	20.48
	SE	(3.26)	(3.43)	(3.23)	(3.33)	(3.55)	(3.68)
Perc. Task Demands[a]	*M*	2.67	2.74	2.10	2.35	2.42	2.30
	SE	(0.27)	(0.28)	(0.26)	(0.27)	(0.29)	(0.30)
Interest[a]	*M*	3.34	3.15	3.24	3.04	3.51	2.41
	SE	(0.28)	(0.29)	(0.27)	(0.29)	(0.30	(0.31)

Note. [a] scale range : 1 to 6; STS = secondary task stimulus; *n* = sample size; *M* = mean; *SE* = standard error; perc. = perceived

Perceived task demands. It was tested whether participants or groups differed with respect to perceived task demands, that is, how hard they think it is to learn with an integrated or separated format before learning. A paired t-test yielded that participants did not differ in how hard they assumed it is to learn with either separated or integrated format ($t(97) = -.46$, $p = .65$).

A correlation analysis yielded that participants who rated that it would be harder for them to learn with separated format rated that it would be easier for them to learn with integrated format and vice versa ($r(97) = -.28$, $p < .01$) indicating that participants preferred either an integrated or a separated format.

A 2 (instructional format) x 3 (secondary task) ANCOVA with cognitive processing capacity as covariate on the perceived task demands of the respective format participants learned with during the experiment did neither demonstrate a main effect of instructional format ($F < 1$), nor a main effect of secondary task ($F(2, 86) = 1.69$, $MSE = 1.99$, $p = .19$, $f = 0.20$), nor an interaction effect ($F < 1$), indicating that participants assigned to the conditions with separated format did not think that it was harder to learn with separated format than participants assigned to the conditions with integrated format thought about learning with integrated format. Because there were no differences between the conditions, further analyses did not controlled perceived task demands statistically.

Moreover, partial correlations between perceived task demands the and the ratings of the three cognitive load types were run and controlled for instructional format, secondary task and cognitive processing capacity to exclude any possible influence from these variables on the correlations. Extraneous load ratings showed a trend towards a positive correlation with perceived task demands ($r(88) = .18$, p = .09) indicating maybe a weak relation between extraneous cognitive load and learners' motivation to learn.

Interest. A 2 (instructional format) x 3 (secondary task) ANCOVA with cognitive processing capacity as covariate tested whether instructional format influenced interest as a possible motivational variable. Participants with integrated format found the learning content more interesting than participants with separated format ($F(1, 86) = 4.34$, $MSE = 1.26$, $p = .04$, $f = 0.22$) indicating maybe a weak effect of instructional format on participants' motivation. Whether participants had to react to a secondary task probe did neither influence the interest ratings ($F < 1$) nor interact with instructional format ($F(1, 86) = 1.59$, $MSE = 1.26$, $p = .21$, $f = 0.19$). Cognitive processing capacity did not influence the interest ratings ($F < 1$).

Moreover, partial correlations between the cognitive load type and interest were run and controlled for instructional format, secondary task and cognitive processing capacity to exclude any possible influence from these variables on the correlations.

The analyses revealed that higher intrinsic cognitive load ratings were related with lower interest ratings ($r(88) = -.49$, $p < .01$) indicating that the higher the topic difficulty was rated, the lower was the learners' interest in the topic.

5.3.2 Learning Outcomes

At first, Cronbach's α of the post-tests were calculated to get general information on the different tests. To analyze participants' knowledge gains and learning outcomes a 2 (time of test) x 2 (instructional format) x 3 (secondary task) RM-ANCOVA that controlled for cognitive processing capacity was run for each type of knowledge test. Time of test was the within subject variable, whereas instructional format and secondary task were between subject variables.

The Box-M-tests were significant in all analyses ($p <. 01$) and thereby indicated that the assumptions of covariance homogeneity was not fulfilled. These results were caused by the fact that participants' knowledge varied more highly in the post-tests than in the pre-tests. As this result does not seem to be caused by differences in variances across the instructional format conditions or secondary task conditions, the violation against this statistical precondition does neither seem to make non-parametrical tests necessary nor seem to influence the interpretability of the further results.

Because cognitive processing capacity interacted with the time of test (pre/-post test) in three out of the four RM-ANCOVA analyses, additional 2 (instructional format) x 3 (secondary task) ANCOVAs[5] that controlled for cognitive processing capacity were run for these pre- and post-tests to analyze the influence of cognitive processing capacity in more detail. Table 12 summarizes the adjusted means and standard errors of all knowledge tests as a function of time of test, instructional format, and secondary

---

[5] The results of the ANCOVAs showed the same patterns concerning learning outcomes and split-attention like the RM-ANCOVAS, and thus are not reported comprehensively. Moreover, when using the knowledge post-test scores without confidence weighting as learning outcome measures, the split-attention effects yielded on the post-tests remain for the knowledge tests on labels ($p < .01$, $f = 0.48$) and on complex facts ($p < .01$, $f = 0.29$) but not for the knowledge test on terms ($p = .37$, $f = 0.09$). ANCOVAs on the confidence weightings instead of the learning outcomes showed an instructional format effect on the confidence ratings for the test on terms ($p < .01$, $f = 0.477$), labels ($p < .01$, $f = 0.477$), and complex facts ($p < .01$, $f = 0.477$) but not on transfer ($p =.28$, $f = 0.477$). Thus, participants with integrated format stated higher security for their knowledge than participants with separated format thereby corroborating the positive effect of integrated formats on learning.

task stimulus controlled for cognitive processing capacity.

Labeling. Cronbach's α of the labeling post-test was .84 which is rather high and indicates that the better participants knew one part of the structure, the better they knew the whole structure. The 2 x 2 x 3 RM-ANCOVA yielded that participants in all conditions had a significant gain in knowledge on labeling the structure graphic as indicated by the main effect of time of test ($F(1, 86) = 19.82$, $MSE = 165.41$, $p < .01$, $\eta_P^2 = .19$). The interaction between time of test and instructional format ($F(1, 86) = 31.32$, $MSE = 165.41$, $p = .01$, $\eta_P^2 = .07$) was significant.

Table 12

Means and standard errors of % correct of the four knowledge tests as a function of instructional format, secondary task stimulus, and test time controlled for cognitive processing capacity

% correct			STS Letter		STS Background		STS without	
			Inte-grated	Separated	Inte-grated	Separated	Inte-grated	Separated
			(*n* = 17)	(*n* = 15)	(*n* = 17)	(*n* = 17)	(*n* = 14)	(*n* = 13)
Termi-nology	*Pre*	*M*	5.17	1.44	1.26	6.98	0.19	0.61
		SE	(2.31)	(2.43)	(2.29)	(2.36)	(2.52)	(2.61)
	Post	*M*	66.81	50.55	71.45	59.14	65.62	68.63
		SE	(4.07)	(4.28)	(4.03)	(4.16)	(4.43)	(4.60)
Label-ing	*Pre*	*M*	0.50	0.17	0.65	1.78	1.72	0.79
		SE	(0.65)	(0.69)	(4.51)	(0.67)	(0.71)	(0.74)
	Post	*M*	65.25	44.58	72.89	47.69	71.43	52.34
		SE	(4.55)	(4.79)	(4.51)	(4.65)	(4.96)	(5.14)
Comp-lex Facts	*Pre*	*M*	1.60	0.66	0.97	3.24	1.57	0.53
		SE	(0.91)	(0.96)	(0.90)	(0.93)	(0.99)	(1.03)
	Post	*M*	37.60	25.39	39.00	29.61	39.53	30.59
		SE	(2.67)	(2.81)	(2.64)	(2.73)	(2.91)	(3.01)
Trans-fer	*Pre*	*M*	2.73	0.52	1.43	3.93	2.12	0.57
		SE	(0.91)	(0.97)	(0.91)	(0.94)	(1.00)	(1.04)
	Post	*M*	23.38	24.24	25.27	20.47	23.97	26.87
		SE	(2.36)	(2.48)	(2.33)	(2.41)	(2.57)	(2.66)

Note. STS = secondary task stimulus; *n* = sample size; *M* = mean; *SE* = standard error.

Bonferroni-adjusted comparisons yielded that participants with integrated format outperformed participants with separated format on the post-test in the secondary task conditions and in the control conditions (all *p*s $< .01$) indicating a split-attention effect. Participants with integrated format did not differ from participants with separated format in the pre-test, except participants with separated format and the background secondary task stimulus tended to have higher prior knowledge than participants with integrated format ($p = .07$). The interaction between time of test and secondary task ($F < 1$) as well as the interaction between time of test and instructional format and secondary task ($F < 1$) were neither significant. These results indicate that the secondary task did not influence the learning outcomes on labeling. The interaction between time of test and cognitive processing capacity was significant ($F(1, 86) = 12.74$, $MSE = 165.41$, $p < .01$, $\eta_P^2 = .13$), indicating that the influence of processing capacity differed in the pre- and post- tests. 2 x 3 ANOCAs showed that cognitive processing capacity did not influence the pre-test outcomes ($F(1, 86) = 1.18$, $MSE = 7.03$, $p = .28$, $\eta_P^2 = .01$), but significantly supported the post-test outcomes ($F(1, 86) = 13.40$, $MSE = 343.12$, $p < .01$, $\eta_P^2 = .14$).

Complex Facts. Cronbach's α of the post-test about complex facts was .77 which is satisfactory and indicates that the better participants knew about one process, the better they knew about the other processes taking part in the kidney. The 2 x 2 x 3 RM-ANCOVA showed that participants in all conditions had a significant gain in knowledge on complex facts as indicated by the main effect of time of test ($F(1, 86) = 18.10$, $MSE = 61.80$, $p < .01$, $\eta_P^2 = .17$). The interaction between time of test and instructional format was significant ($F(1, 86) = 18.64$, $MSE = 61.80$, $p = .01$, $\eta_P^2 = .18$). Bonferroni-adjusted comparisons yielded that all participants with integrated format outperformed participants with separated format on the post-test (letter secondary task: $p < .01$; background secondary task: $p = .02$; and control conditions: $p = .04$), indicating a split-attention effect. Participants with integrated format did not differ from participants with separated format in the pre-test, except participants with separated format and the background secondary task stimulus tended to have higher prior knowledge than participants with integrated format in the control condition ($p = .09$). The interaction between time of test and secondary task ($F < 1$) as well as the interaction between time of test and instructional format and secondary task ($F < 1$) were neither significant. These insignificant results indicate that secondary task did not influence learning outcomes. The interaction between time of test and cognitive processing capacity was significant ($F(1, 86) = 8.99$, $MSE = 61.80$, $p < .01$, $\eta_P^2 = .10$), indicating that the influence of processing capacity differed in the pre- and post- tests. 2 x 3

ANOCAs showed that cognitive processing capacity did not influence the pre-test outcomes ($F < 1$), but significantly supported the post-test outcomes on complex facts ($F(1, 86) = 8.55$, $MSE = 117.74$, $p < .01$, $\eta_P^2 = .09$).

Transfer. Cronbach's α of the transfer post-test was .67 which is rather low and suggests that if learners succeeded in transfer on one item, they did not perform well in the transfer test in general. This makes sense in so far that there was not just one underlying principle of the physiology of the kidney to be learned but several. Hence, understanding one functioning principle of the kidney did not mean that another principle was also learned. The 2 x 2 x 3 RM-ANCOVA showed that participants in all conditions had a significant gain in knowledge on complex facts as indicated by the main effect of time of test ($F(1, 86) = 10.01$, $MSE = 47.29$, $p < .01$, $\eta_P^2 = .10$). The interaction between time of test and instructional format was not significant ($F < 1$), indicating that there was no split-attention effect on transfer tasks. Participants with integrated format did not differ from participants with separated format on the post-test. The interaction between time of test and secondary task was neither significant ($F(1, 86) = 1.09$, $MSE = 47.29$, $p < .34$, $\eta_P^2 = .02$), indicating that secondary task did not influence participants learning outcomes on the transfer post-test. The three-way interaction between time of test, instructional format, and secondary task was significant $F(1, 86) = 3.60$, $MSE = 47.29$, $p < .03$, $\eta_P^2 = .07$). Bonferroni-adjusted comparisons yielded that this effect was obviously caused by the fact that participants with separated format had high prior knowledge in the transfer test than participants with integrated format in the conditions with the background secondary task stimulus ($p = .04$). All other comparisons were not significant ($p > .10$). The interaction between time of test and cognitive processing capacity was again significant ($F(1, 86) = 6.56$, $MSE = 47.29$, $p = .01$, $\eta_P^2 = .07$). 2 x 3 ANOCAs showed that cognitive processing capacity did not influence the pre-test outcomes ($F < 1$), but significantly supported the post-test outcomes on transfer tasks ($F(1, 86) = 7.68$, $MSE = 91.94$, $p < .01$, $\eta_P^2 = .08$).

5.3.3 Cognitive Load Measures

Experiment 1 measured overall cognitive load by secondary task performance and the three cognitive load types by subjective ratings. In analyzing these measures different statistical analyses were applied. First of all and in accordance with former cognitive load studies, it was probed by means of repeated measurement RM-ANCOVAs (secondary task performance) and ANCOVAs (subjective ratings) whether there were mean differences across the groups caused on the secondary task performance as well as on the subjective ratings. Furthermore, correlations between the three subjective load type measures were analyzed to investigate how (dis)similar the measures are. Moreover and new in the context of cognitive load research, multiple mediation analyses were used to test whether, and if so, which type of cognitive load (and/or interest) mediated the split-attention effect demonstrated for knowledge about terms, labels, and complex facts.

5.3.3.1 Secondary task performance: Overall cognitive load

Three types of measures were calculated: (1) the percentage of hits during the baseline phase, the introduction phase, the structure graphic phase, and the process graphic phase, and (2) the mean reaction times for the hits during the baseline phase, the introduction phase, the structure graphic phase, and the process graphic phase of the four groups with a secondary task. The mean reaction times were probed for normal distribution by K-S tests. None of the K-S tests reached significance, thereby indicating normal distribution for all four types of mean reaction times per group. The means and standard deviations of the percentage of hits during all four phases are presented in Table 13, whereas the means and standard deviations of the mean reaction times during all four phases are presented in Table 14.

Table 13

Means and standard deviation of percentage of hits as a function of experimental phase, instructional format, and secondary task stimulus

		STS $_{Letter}$		STS $_{Background}$	
		Integrated	**Separated**	**Integrated**	**Separated**
		(*n* = 17)	(*n* = 16)	(*n* = 19)	(*n* = 17)
Baseline	*M*	94.12	97.92	98.25	97.06
	SD	(14.36)	(5.69)	(7.65)	(6.55)
Introduction	*M*	98.24	97.50	94.21	92.94
	SD	(3.93)	(4.47)	(16.44)	(19.61)
Structure Graphic	*M*	92.16	90.97	93.57	94.77
	SD	(6.53)	(12.32)	(12.46)	(9.72)
Process Graphic	*M*	93.92	96.04	97.54	96.47
	SD	(6.26)	(3.27)	(2.18)	(5.95)

Note. STS = secondary task stimulus; *n* = sample size; *M* = mean; *SD* = standard deviation.

Percentages of hits. A 2 (instructional format) x 2 (secondary task) x 4 (experimental phase) RM-ANOVA was run. Mauchly's test indicated that the assumption of sphericity was not fulfilled ($\chi^2(5) = 22.02$, $p < .01$) and that the range of hits differed largely across the groups. To consider these differences statistically the degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = .83$). There was a trend that the percentage of hits differed among the phases ($F(2.50, 195) = 2.26$, $MSE = 108.46$, $p = .09$, $\eta_P^2 = .03$). Inspecting the means and standard deviations, the phase with the structure graphic had the lowest hit rates with rather high standard deviations in all conditions. The interaction between phases and secondary task ($F(2.50, 195) = 1.96$, $MSE = 108.46$, $p = .12$, $\eta_P^2 = .03$), the interaction between phases and instructional format ($F < 1$), and the three-way interaction between phases, secondary task, and instructional format ($F < 1$) were not significant. However, Bonferroni-adjusted comparisons yielded that participants with letter secondary task and integrated format had significantly lower hits during learning the process graphic than participants with background secondary task and integrated format ($p = .02$), indicating that reacting to the letter secondary probe seemed to be more demanding in the integrated format condition where the letter was a bit closer to the periphery of the graphic than reacting to the background secondary probe which was visible all over the screen during the complex process graphic.

Table 14

Means and standard deviation of mean reaction times as a function of experimental phase, instructional format, and secondary task stimulus

		STS_{Letter}		$STS_{Background}$	
		Integrated	**Separated**	**Integrated**	**Separated**
		(*n* = 17)	(*n* = 16)	(*n* = 19)	(*n* = 17)
Baseline	*M*	411.77	423.17	402.79	422.06
	SD	(57.40)	(63.72)	(42.95)	(51.00)
Introduction	*M*	582.83	629.89	593.64	685.52
	SD	(66.92)	(100.93)	(66.26)	(256.19)
Structure Graphic	*M*	610.43	662.26	632.34	674.48
	SD	(72.20)	(109.20)	(106.93)	(124.06)
Process Graphic	*M*	587.33	604.05	604.91	591.80
	SD	(74.62)	(80.15)	(112.07)	(95.71)

Note. STS = secondary task stimulus; *n* = sample size; *M* = mean; *SD* = standard deviation.

Reaction times. A 2 (instructional format) x 2 (secondary task) x 4 (experimental phase) RM-ANOVA was run. Mauchly's test indicated that the assumption of sphericity was not fulfilled ($\chi^2(5) = 27.17$, $p < .01$). Therefore, the degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = .79$). There was a main effect of the within-factor phase ($F(2.37, 153.70) = 128.31$, $MSE = 7517.40$, $p < .01$, $\eta_P^2 = .66$.). Bonferroni-adjusted comparisons indicated that all participants were significantly faster during the baseline phase than during the three learning phases introduction, structure graphic, and process graphic (all *p*s < .01), independently from type of secondary task stimulus ($F < 1$). The interaction between phase and instructional format tended to be significant ($F(2.37, 153.70) = 2.72$, $MSE = 7517.40$, $p = .06$, $\eta_P^2 = .04$) thereby indicating that the influence of the instructional format differed with respect to the phase. Bonferroni-adjusted comparisons showed that participants with background stimulus and separated format tended to differ from participants with integrated format ($p = .06$) during the introduction phase without experimental manipulations indicating individual differences between the groups in reaction times. To control for these differences the following RM-ANCOVA was run. The means and standard errors of these reaction times controlled for individual reaction times during the baseline phase and introduction are presented in Table 15.

Table 15

Means and standard errors of reaction times (ms) as a function of instructional format, secondary task stimulus controlled for reaction times during the baseline phase and controlled for reaction times during the introduction

		STS $_{Letter}$		STS $_{Background}$	
		Integrated	**Separated**	**Integrated**	**Separated**
		(*n* = 17)	(*n* = 16)	(*n* = 19)	(*n* = 17)
Structure Graphic	*M*	628.33	656.78	647.55	644.75
	SE	(19.87)	(20.35)	(18.80)	(20.18)
Process Graphic	*M*	602.67	600.55	616.52	566.78
	SE	(18.22)	(18.66)	(17.24)	(18.50)

Note. ms = milliseconds; STS = secondary task stimulus; *n* = sample size; *M* = mean; *SE* = standard error.

A 2 (instructional format) x 2 (secondary task) x 2 (experimental phase) RM-ANCOVA was run that controlled for individual differences by including the mean reaction times of the baseline phase and the introduction phase as covariates. The box-M-test was not significant indicating that the covariances were equal across the groups. Only the interaction between instructional phase and instructional format was significant ($F(1, 63) = 4.48$, $MSE = 2697.57$, $p = .04$, $\eta_P^2 = .07$). Bonferroni-adjusted comparisons showed that learners with separated format (with both letter and background stimulus) responded significantly faster during learning the process than during the structure graphic (both *p*s < .01), whereas learners with integrated format and with background stimulus only tended to respond faster during learning the process graphic than during the structure graphic ($p = .07$) and learners with letter task did not respond differently during learning the two graphics ($p = .16$). These results indicate that the content of the learning material per se is an important characteristic. Moreover, Bonferroni-adjusted comparisons showed that participants with background secondary task and integrated format tended to react more slowly during the process graphic phase ($p = .06$) than participants with separated format. This result is in the opposite direction to the assumed extraneous cognitive load explanation.

5.3.3.2 Subjective ratings: Cognitive load types

In investigating the subjective cognitive load ratings, it was first tested whether and how the instructional formats influenced these ratings, and second, how the subjective ratings of the three load types were related with each other, and third, whether the subjective cognitive load ratings functioned as mediators between instructional format and learning outcomes. Again tests controlled for cognitive processing capacity because this variable is regarded as important cognitive factor which might distort cognitive load type results because conditions differ with regard to cognitive processing capacity. Adjusted means and standard errors of the cognitive load ratings are reported in Table 16.

Group differences. 2 (instructional format) x 3 (secondary task) ANCOVAs controlled for cognitive processing as covariate were run to test how instructional format influenced the subjective cognitive load ratings.

Intrinsic cognitive load. Participants with the separated format tended to find the learning content more difficult than participants with the integrated format ($F(1, 86) = 3.22$, $MSE = .91$, $p = .08$, $f = 0.19$). Whether participants had to react to a secondary task stimulus did neither influence the difficulty ratings of the learning content ($F < 1$) nor interact with instructional format ($F < 1$). Cognitive processing capacity did not influence the subjective ratings of content difficulty ($F(1, 86) = 1.67$, $MSE = .91$, $p = .20$, $\eta_P^2 = .02$).

Table 16

Means and standard errors of the subjective cognitive load ratings as a function of instructional format, secondary task, and controlled for cognitive processing capacity

		STS Letter		STS Background		STS without	
		Integrated	Separated	Integrated	Separated	Integrate d	Separated
		(*n* = 17)	(*n* = 15)	(*n* = 17)	(*n* = 17)	(*n* = 14)	(*n* = 13)
ICL[a]	*M*	3.78	4.15	3.92	4.26	3.84	4.22
	SE	(0.23)	(0.25)	(0.23)	(0.24)	(0.26)	(0.26)
ECL[a]	*M*	2.94	3.07	2.77	3.35	2.79	3.23
	SE	(0.21)	(0.23)	(0.23)	(0.22)	(0.23)	(0.24)
GCL[a]	*M*	4.31	3.95	4.57	3.90	4.91	4.53
	SE	(0.16)	(0.17)	(0.16)	(0.17)	(0.18)	(0.19)

Note. [a] scale range: 1 to 6; STS = secondary task stimulus; *n* = sample size; *M* = mean; *SE* = standard error, ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

Extraneous cognitive load. Participants with the separated format tended to find it more difficult to learn with separated materials than participants with integrated format ($F(1, 86) = 4.23$, $MSE = .76$, $p = .04$, $f = 0.22$). Whether participants had to react to a secondary task stimulus did neither influence the difficulty ratings of the learning content ($F < 1$) nor interact with instructional format ($F < 1$). Cognitive processing capacity did not influence how difficult the participants perceived the materials ($F < 1$).

Germane cognitive load. Participants with the integrated format reported to have concentrated more during learning than participants with the separated format ($F(1, 86) = 10.79$, $MSE = .45$, $p < .01$, $f = 0.32$). Whether participants had to react to a secondary task stimulus also influenced the ratings of the concentration level ($F(2, 86) = 6.46$, $MSE = .45$, $p < .01$, $f = 0.35$). Bonferroni adjusted post-hoc comparisons yielded that participants with the perceptual secondary task reported to have concentrated less than participants without a secondary task ($p < .01$). Similarly, participants with the background stimulus also reported to have concentrated less during learning than participants without secondary task ($p = .02$). However, instructional format did not interact with secondary task ($F < 1$). Cognitive processing capacity tended to influence the concentration ratings ($F(1, 86) = 2.92$, $MSE = .45$, $p = .09$, $\eta_P^2 = .03$).

Partial correlations between cognitive load types. Partial correlations between the three cognitive load type measures controlled for instructional format, secondary task and cognitive processing capacity were run to get a closer look on the relations between the three cognitive load types without the influence of the control variables. They revealed that higher intrinsic load ratings were related with higher extraneous load ratings ($r(88) = .45$, $p < .01$), and higher germane load ratings ($r(88) = .27$, $p = .01$). Extraneous load did not correlate with germane load ($r(88) = .07$, $p = .50$). The non-correlation between the extraneous and germane load measures might indicate two possible factors that might mediate the split-attention effect individually. Multiple mediation analyses were used to test this assumption.

Multiple mediation. To test whether participants' subjective ratings of content difficulty (intrinsic cognitive load), difficulty in learning with the instructional material (extraneous cognitive load), and level of concentration (germane cognitive load) were the mediating variables between instructional format and learning outcomes on the knowledge tests, mediation analysis were run. Because learners' *interest* was also influenced by the instructional format, it was included in the regression analyses to test whether it functions as a motivational mediator. Moreover, cognitive processing was again included in the model to control for its influence. Mediation analyses are based on multiple regression equations. To run the regression analyses needed for multiple mediation analyses the two nominal variables, instructional format and secondary task, were effect coded with unweighted means. Instructional format was coded in the following way: Participants with the separated format were coded -1 and participants with the integrated format were coded +1. For secondary task two coding variables (cv) were needed: Participants with the background stimulus were coded -1 and -1, participants with the letter stimulus were coded +1 and 0, and participants without secondary task were coded 0 and +1, respectively. When using unweighted effect coding the unstandardized B- and standardized Beta-weights represent the deviation of the outcome for each separate group from the mean of all groups with the means of each group contributing equally to the overall mean. This coding method is equivalent to the method used for ANOVA and ANCOVA (Cohen, Cohen, West, & Aiken, 2003). The cognitive load ratings, interest, and cognitive processing capacity were centered thereby making 0 to the mean and trying to keep multicollinearity as low as possible because the correlation of one predictor with other predictors maybe reduced. Subsequently the variables were included as continuous predictors in the respective models.

Three types of regression models were estimated to describe the multiple mediation of each knowledge test: Four simple dependent variable models (step 1), four mediator variable models (step 2), and four mediational dependent variable models (step 3). The models' results are reported in the following. Only the significance tests with regard to instructional format and the cognitive load types are reported. Subsequently, the bootstrapping analyses representing step 4 in mediation analysis are reported. Finally, the descriptive values concerning the amount of variance mediated by the respective variables are reported (final step of mediation analysis).

Step 1: Dependent variable models. The multiple regression models for the learning outcomes of the four knowledge tests included the following predictors: instructional format, secondary task coding variable (cv) 1 and cv2, interactions between instructional format and secondary task cv1 and cv2, and cognitive processing capacity. The regression models yielded a significant effect of instructional format for *terminology* ($t(86) = 2.40$, $p = .02$), for *labeling* ($t(86) = 5.46$, $p < .01$), for *complex facts* ($t(86) = 4.38$, $p < .01$) but not for *transfer* ($t(86) = 0.17$, $p = .87$). The parameter estimates (B-weights, standard errors, ß-weights, and R^2) of the dependent variable models are summarized in Table 17 and 18.

Table 17

Summary of the multiple regressions for the dependent variable models of the terminology and labeling knowledge test

	Terminology			Labeling		
	B	***SE_B***	***ß***	***B***	***SE_B***	***ß***
Instructional format	4.26	1.77	.24*	10.83	1.98	.47**
Secondary task cv1	- 5.02	2.43	- .24*	- 4.11	2.72	- .15
Secondary task cv2	3.42	2.53	.15	2.85	2.83	.10
Instructional format x Secondary task cv1	3.87	2.42	.18	- 0.49	2.71	- .02
Instructional format x Secondary task cv2	- 5.77	2.55	- .26*	- 1.28	2.85	- .05
Cognitive processing capacity	0.41	.23	$.18^+$	0.94	0.26	.32**
Model fit	$R^2 = .21$			$R^2 = .41$		

Note. $^+ p < .10$, * $p < .05$, ** $p < .01$; cv = coding variable; *B* = unstandardized coefficient, SE_B = standard error of B, *ß* = standardized coefficient.

Table 18

Summary of the multiple regressions for the dependent variable models of the complex facts and transfer test

	Complex Facts			Transfer		
	B	SE_B	*ß*	*B*	SE_B	*ß*
Instructional format	5.10	1.16	.41**	0.17	1.03	.02
Secondary task cv1	- 2.12	1.59	- .14	- 0.23	1.40	- .02
Secondary task cv2	1.44	1.66	.09	- 1.38	1.47	.11
Instructional format x Secondary task cv1	1.02	1.59	.07	- 0.60	1.40	- .05
Instructional format x Secondary task cv2	- 0.62	1.67	- .04	- 1.62**	1.67	- .13
Cognitive processing capacity	0.44	0.15	.28**	0.37	0.13	.29**
Model fit	R^2 = .31			R^2 = .15		

Note. $^+$ *p* < .10, * *p* < .05, ** *p* < .01; cv = coding variable; *B* = unstandardized coefficient, SE_B = standard error of B, *ß* = standardized coefficient.

Table 19

Summary of the multiple regression parameter estimates for the mediator variable models with intrinsic, extraneous, and germane cognitive load as the dependent variables

	ICL			ECL			GCL		
	B	SE_B	*ß*	*B*	SE_B	*ß*	*B*	SE_B	*ß*
Instructional format	- 0.18	0.10	- .20$^+$	- 0.19	0.09	- .22*	0.24	0.07	.31**
Secondary task cv1	- 0.07	0.14	- .06	- 0.02	0.13	- .02	- 0.23	0.10	- .26*
Secondary task cv2	0.01	0.15	.01	- 0.02	0.13	- .01	0.36	0.10	.39**
Instructional format x Secondary task cv1	- 0.01	0.14	- .01	0.13	0.13	.13	- 0.05	0.10	- .06
Instructional format x Secondary task cv2	- 0.01	0.15	- .01	- 0.03	0.13	- .03	- 0.05	0.10	- .05
Cognitive processing capacity	- 0.02	0.01	.14	0.01	0.01	- .01	- 0.02	0.01	- .17$^+$
Model fit	R^2 = - .05			R^2 = .06			R^2 = .28		

Note. $^+$ *p* < .10, * *p* < .05, ** *p* < .01. IV = independent variable. cv = coding variable.
ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.
B = unstandardized coefficient, SE_B = standard error of B, *ß* = standardized coefficient.

Step 2: Mediator variable models. The multiple regression models for intrinsic, extraneous, germane cognitive load and interest included the following predictors: instructional format, secondary task cv1 and cv2, interaction between instructional format and cv1, interaction between instructional format and cv2, and cognitive processing capacity. In accordance with the above reported ANCOVAs the regression models yielded a marginally significant effect of instructional format for *intrinsic cognitive load* ($t(86) = -1.54$, $p = .08$), a significant effect of instructional format for *extraneous cognitive load* ($t(86) = -2.06$, $p = .04$), for *germane cognitive load* ($t(86) = 3.28$, $p < .01$), and for *interest* ($t(86) = -2.08$, $p = .04$). The parameter estimates (B-weights, standard errors, ß-weights, and R^2) of the cognitive load mediator variable models are summarized in Table 19.

***Step 3: Mediational dependent variable models*.** The multiple regression models of the learning outcomes of the four knowledge tests included the following independent variables: instructional format, secondary task cv1 and cv2, interaction between instructional format and secondary task cv1, interaction between instructional format and secondary task cv2, cognitive processing capacity, as well as intrinsic, extraneous, as well as germane cognitive load, and interest. Interest was included in order to test and control for possible motivational effects on learning outcomes. The regression models yielded a marginally significant effect of instructional format for *terminology* ($t(86) = 1.68$, $p = .09$), significant effects for *labeling* ($t(86) = 3.97$, $p < .01$), for *complex facts* ($t(86) = 2.73$, $p < .01$), but again not for *transfer* ($t(86) = -0.90$, $p = .37$). Moreover, the analyses yielded that extraneous cognitive load was a marginally significant predictor of participants' learning outcomes on labeling ($t(86) = -1.72$, $p = .09$), and a significant predictor for learning outcomes on complex facts ($t(86) = -2.99$, $p < .01$) and transfer ($t(86) = -2.24$, $p = .02$). Germane cognitive load was a significant predictor for learning outcomes on complex facts ($t(86) = 2.18$, $p = .03$). Intrinsic cognitive load and interest were not related with the learning outcomes when extraneous and germane cognitive load were included in the models. The parameter estimates (B-weights, standard errors, ß-weights, and R^2) for the mediational dependent variable models are summarized in Table 20.

Table 20

Summary of the multiple regressions for the mediational dependent variable models of the test about terminology, labeling, complex facts and transfer

	Terminology			Labeling		
	B	SE_B	*ß*	*B*	SE_B	*ß*
Instructional format	3.34	1.99	.19+	8.43	2.12	.36**
Secondary task cv1	- 4.78	2.57	- .23+	- 3.47	2.74	- .13
Secondary task cv2	2.89	2.83	.13	1.45	3.02	.05
Instructional format x Secondary task cv1	4.08	2.48	.19	0.52	2.65	.02
Instructional format x Secondary task cv2	- 6.00	2.66	- .27	- 1.68	2.84	- .06
Cognitive processing capacity	0.40	0.24	.18	0.88	0.25	.30**
ICL	- 1.39	2.54	- .07	- 1.01	2.71	- .04
ECL	0.30	2.41	.01	- 4.41	2.57	- .17+
GCL	1.98	2.99	.08	4.27	3.19	.14
Interest	1.02	1.96	.06	1.46	2.09	.07
Model fit	R^2 = .23			R^2 = .48		
	Complex Facts			**Transfer**		
	B	SE_B	*ß*	*B*	SE_B	*ß*
Instructional format b	3.16	1.16	.25**	- 0.98	1.08	- .10
Secondary task cv1	- 1.46	1.50	- .10	0.20	1.40	.02
Secondary task cv2	0.23	1.65	- .12	1.15	1.54	.09
Instructional format x Secondary task cv1	1.96	1.45	.13	- 0.01	1.35	- .01
Instructional format x Secondary task cv2	- 0.98	1.55	- .06	- 2.02	1.45	- .16
Cognitive processing capacity	0.37	0.14	.23**	0.35	0.13	.28**
ICL	- 0.54	1.48	- .04	- 0.24	1.38	- .02
ECL	- 4.19	1.40	- .29**	- 2.93	1.31	- .25*
GCL	3.80	1.74	.23*	1.06	1.63	.08
Interest	1.33	1.14	.12	1.15	1.07	.13
Model fit	R^2 = .46			R^2 = .26		

Note. + p < .10, * p < .05, ** p < .01.
ICL = Intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.
cv = coding variable, *B* = unstandardized coefficient, SE_B = standard error of B, *ß* = standardized coefficient.

Step 4: Bootstrapping analyses and confidence intervals. The mediational dependent variable models demonstrated that instructional format still influenced the learning outcomes when controlling for cognitive load and interest. To test whether intrinsic cognitive load, extraneous cognitive load, and germane cognitive load mediated the split-attention effect on the learning outcomes of terminology, labeling and complex facts at least partially, bootstrapping analyses was used. The bootstrapping method was used to test the significance of the three mediators by using the SPSS Macro and the according commands provided by Preacher and Hayes (2008). The bootstrapping analyses estimated the indirect effect of instructional format on the learning outcomes on terminology, labeling, and complex facts through intrinsic, extraneous, and germane cognitive load and interest. 95 % bias corrected and accelerated bootstrap confidence intervals (CIs) were produced basing on 2000 bootstrap samples. For knowledge about *terminology* all CIs included 0 indicating that none of the cognitive load types functioned as a mediator. For knowledge about *labeling* the CIs of extraneous cognitive load (CI_{lower} = 0.06, CI_{upper} = 2.17) and germane cognitive load (CI_{lower} = 0.31, CI_{upper} = 2.15) did not include 0, indicating that both load types had a mediational function for the split-attention effect on labeling (see Figure 16).

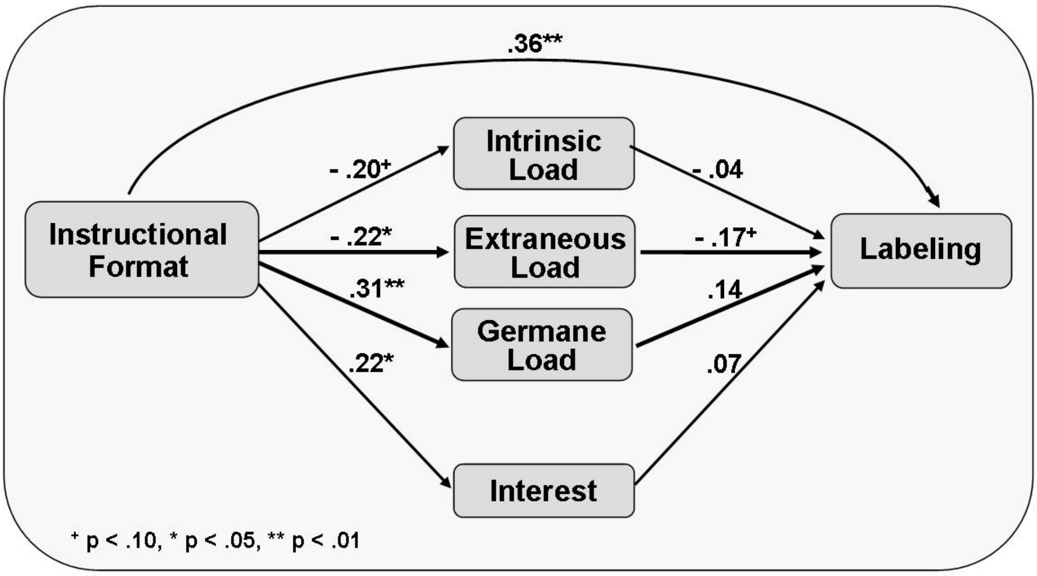


Figure 16. Illustration of the empirical multiple mediation model for labeling including ß-weights

For knowledge about *complex facts* the CIs of extraneous cognitive load (CI_{lower} = 0.01, CI_{upper} = 2.80) and germane cognitive load (CI_{lower} = 0.06, CI_{upper} = 3.05) did again not include 0, indicating that both load types had a mediational function for the split-attention effect on complex facts (see Figure 17). Figures 16 and 17 depict the multiple mediation models for labeling and complex facts by showing only the predictor instructional format.

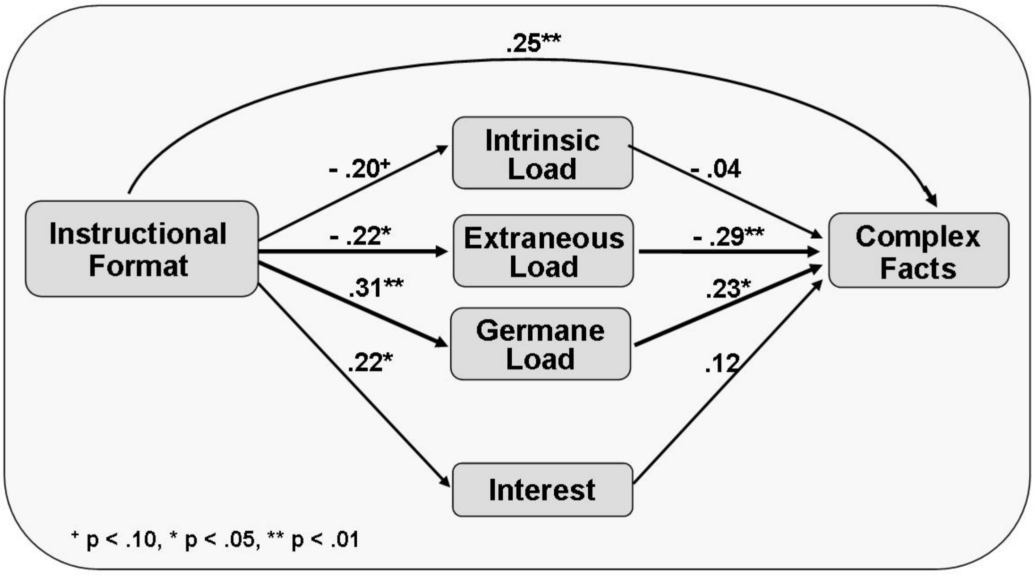


Figure 17. Illustration of the empirical multiple mediation model for complex facts including ß-weights

Final Step: Proportions of effects mediated. To summarize, the bootstrapping and CIs results indicated that extraneous and germane cognitive load were partial mediators for knowledge about labeling and complex facts. Extraneous cognitive load mediated 7.82 % of the total effect of instructional format on knowledge about labeling and 15.78% of the total effect on knowledge about complex facts. Germane cognitive load mediated 9.30% of the total effect of instructional format on knowledge about labeling and 17.50% of the total effect on knowledge about complex facts.

5.4 Summary and Discussion

The aim of Experiment 1 was to investigate the explanations of the split-attention effect in multimedia learning, because different cognitive processing mechanisms are assumed to underlie the effect as shown in the literature review in Chapter 4. It was analyzed why learners with integrated formats outperform learners with separated formats. More precisely, it was analyzed whether the split-attention effect is caused either only by a decrease in extraneous cognitive load reflected in a decrease of overall cognitive load in learners with integrated format (extraneous cognitive load explanation) or by a decrease in extraneous cognitive load and a concurrent increase in germane cognitive load (germane cognitive load explanation).

In accordance with the split-attention effect, the experiment demonstrated that participants learning with an integrated format about the kidney's physiology outperformed participants with a separated format on several knowledge tests. Secondary task performance was used as a direct and objective measure of overall cognitive load, whereas subjective rating scales were used to measure intrinsic, extraneous, and germane cognitive load individually. The results of the secondary task performance do not support the extraneous cognitive load mechanism assumed. The results of the subjective cognitive load type ratings support the germane cognitive load assumption. The hypotheses and results of this experiment are discussed in more detail in the following.

5.4.1 Split-Attention Effect: Learning Outcomes

According to the split-attention effect it was assumed that learners with integrated format should outperform learners with separated format. The results demonstrated a split-attention effect for tests measuring simple facts like knowing the correct terminology and knowing how to label the structural parts of the depicted nephron. Furthermore, a split-attention effect was shown for remembering complex facts about the physiological processes. However, the split-attention effect was not found on the transfer test demanding complex inferences. This finding is only partially in line with hypothesis *H 1.1* and not in line with many results of former studies or with the assumptions made by the CTML and CLT that both predict a particularly strong effect for transfer tasks basing on inferences but not necessarily for retention tasks asking for factual knowledge. However, our finding is in line with some other studies. For instance, in the domain of geography Purnell et al. (1991) demonstrated the split-attention effect also for factual knowledge but not for inference tasks. And Florax and Plötzner (2009) also replicated the split-attention effect only on retention but not on comprehension questions about the biochemical processes at synapses.

An explanation for the different findings concerning the knowledge tests might be that instructional design effects might progress from factual knowledge to comprehension depending on the complexity of the domain as already outlined. The more comprehensive and complex the factual knowledge is, the longer it may need to first develop factual knowledge representations and then a good comprehension of the domain. Similarly, instructional design effects might also appear first on factual knowledge tests, if there are many and rather complex facts, and are then shown on transfer tasks, if the factual knowledge is developed well enough. In contrast to the former studies in which learners might had already developed good factual knowledge representations, learners in the Purnell et al. (1991) study and in this study might still had difficulties with learning the facts and had not yet developed a deep understanding of all processes. The participants in this experiment reached only about 25% on the transfer test. This might be a hint that the test was rather difficult for the participants. Whether this explanation of the instructional design effect differences between factual knowledge and comprehension or transfer tests holds true, needs more research. Another explanation might be that the differences are caused by specific characteristics of the learning domains or of the tests. Whether one of the suggested explanations holds true, needs further research.

5.4.2 Explanations of the Split-Attention Effect: Cognitive Load Measures

Two different hypotheses patterns concerning overall cognitive load and the three cognitive load types were assumed according to the germane and the extraneous cognitive load assumptions. Before the results of the three cognitive load type scales will be discussed, the secondary task performance as a correlate of overall cognitive load is discussed.

5.4.2.1 Secondary task performance: Overall cognitive load

According to the germane cognitive load explanation overall cognitive load does not need to differ between learners with integrated and separated format, whereas the extraneous cognitive load explanation predicts that learners with separated format should suffer from higher overall cognitive load than learners with integrated format. Hence, according to the extraneous cognitive load explanation, participants with separated format should have lower percentage of hits and longer reaction times (decreased secondary task performance) than participants with integrated format. The results were not in line with this assumption but rather contradicted it. Participants with separated format and background secondary task had shorter reaction times than participants with integrated format and the same stimulus, even though this result was only demonstrated for the process graphic. Hence, hypothesis *H 1.2.1a* is to be rejected. However, the finding can be interpreted in favor of the germane cognitive load explanation. The result might suggest that participants with integrated format were probably loaded by germane cognitive load because they also had higher learning outcomes than participants with separated format. Hence, the results of the reaction times may be interpreted as support of hypothesis *H 1.2.1b*.

Another finding was that participants with integrated format and letter secondary task had lower percentages of hits during learning with the process graphic than participants with integrated format and background secondary task. This result shows that it does matter how the secondary task is designed. One can infer that reacting to the letter secondary task stimuli was more difficult than reacting to the background stimuli because the letter appeared at the left periphery of the nephron on the process graphic, and thus, the color was more difficult to perceive, whereas the background stimulus could be perceived easily no matter which information the participants processed. Although there was no interaction between secondary task stimulus and instructional format that would indicate a confounding of instructional format and structural interference, the results of the percentages of hits show that the type of secondary task stimulus and the learning materials can easily interact on a perceptual rather than a cognitive level. Hence, the results of the cited CLT studies using secondary task performance in Chapter 3 (see also next paragraph) should be interpreted very cautiously because they did not test the perceptual level.

The findings of the secondary task performance contrast with the result of Chandler and Sweller (1996) who demonstrated worse secondary task performance of learners with separated format compared to learners with integrated format. Both results are hard to compare, because the secondary task used by Chandler and Sweller (1996) was rather demanding and probably disturbed learning, and hence, is hard to interpret. Unfortunately, Chandler and Sweller (1996) did not test how performing the secondary task influenced learning behavior. In contrast to Chandler and Sweller (1996) but also in contrast to Brünken et al. (2002) as well as DeLeeuw and Mayer (2008), Experiment 1 explicitly tested, whether performing a secondary task influenced learning or subjective cognitive load ratings. In general, performing the secondary task either with the letter or the background stimulus during learning did not influence learning outcomes. Hence, the secondary tasks used fulfilled the condition that they should not disturb learning. Therefore, the secondary task performance can be used as a measure of overall cognitive load. Despite this overall success, there was a trend for an interaction effect on learning outcomes for the knowledge about the specific terminology of the structural parts in a nephron. Post hoc analyses yielded that students without secondary task who learned with separated format did not differ in their terminology outcomes from students without secondary task and learning with integrated format, whereas students with secondary task (especially with the letter stimulus) who learned with separated format did differ from students learning with integrated format. This indicates that even performing a rather easy secondary task, disturbed students' encoding of the terms under a separated format condition. Furthermore, participants performing a secondary task also reported to have concentrated less than participants without secondary task. This result shows that too demanding secondary tasks probably influence learning behavior. Because it was only a very slight disturbance in this experiment and because the reaction times were sensitive enough to differentiate between the baseline and the learning phases and because the reaction times were influenced by the instructional format they were nevertheless interpreted as overall cognitive load measure.

I suggest that future studies on the split-attention effect (but also on other instructional format effects) that use secondary task performance as overall cognitive load measure should first allow for testing the effect of the secondary task on learning outcomes and should second compare different secondary tasks to better distinguish between interference on a cognitive or a perceptual level. An option would be to use a secondary task stimulus with a modality different from the information to be learned. In testing the split-attention effect (spatial contiguity effect) an auditory stimulus would for instance be suitable. Comparisons of different secondary tasks are also necessary to investigate how different working memory components are involved in instructional design effects (cf. Baddeley, 2002). Although the CTML and especially CLT focus on working memory as the pivotal cognitive structure during learning, both frameworks still have to clarify how overall cognitive load and the load types conform to more elaborated working memory models (for a review of working memory models see Miyake and Shah, 1999).

5.4.2.2 Subjective ratings: Cognitive load types and interest

Because the two explanations proposed for the split-attention effect differ with respect to germane cognitive load but not with respect to intrinsic and extraneous cognitive load, different subjective rating scales were used as attempt to measure the three load types individually. Intrinsic cognitive load was addressed by asking about the difficulty of the learning content, extraneous cognitive load was addressed by asking about the difficulty to learn with the materials, and germane cognitive load was addressed by asking about the level of concentration during learning. Furthermore, it was explored whether instructional format influences the motivational aspect of situational interest.

Intrinsic cognitive load. According to both explanations learners should not differ with respect to intrinsic cognitive load. The finding that learners with the separated format tended to find the learning content more difficulty than learners with the separated format is not in line with both explanations. Hence, hypothesis *H 1.2.2a* was not supported. However, whether participants' ratings reflected intrinsic cognitive load as suggested by CLT is questionable. The finding that intrinsic and extraneous cognitive load ratings were positively correlated, suggests that participants had difficulties in distinguishing between these two load types. There are two reasons why intrinsic and extraneous cognitive load ratings correlated. First, the formulations of the scales may have been too similar. Therefore, some learners may have thought that they were asked to rate more or less the same thing twice and did not distinguish between content and material. Second, it is very possible that learners encountering a new domain can hardly differentiate whether the content or the design makes learning difficult. This might be a limitation of introspection and experience, because it requires knowledge on how difficult the same content would be perceived, if presented in a different way keeping prior knowledge constant. However, this is an experience a person can never make. The slight positive correlation between intrinsic and germane cognitive load is also not predicted by CLT. Notably, Schnotz and Kürschner (2007) argue that intrinsic and germane cognitive load should be related. Despite the doubts that the participants were able to rate intrinsic cognitive load without being influenced by the difficulty of the design, it is interesting to note that the difficulty ratings of the content (intrinsic cognitive load) did not predict learning outcomes, when extraneous and germane cognitive load were simultaneously included in the multiple regression models. This finding suggests that extraneous and germane cognitive load ratings are better or stronger predictors of learning outcomes.

Extraneous cognitive load. According to both explanations extraneous cognitive load should mediate the split-attention effect. In line with this hypothesis was the finding that participants' ratings of how difficult they found the materials to learn with indeed mediated the instructional format effects on knowledge tests about labeling and complex facts. Learners with separated format had higher ratings of material difficulty and higher difficulty ratings were related to worse learning outcomes. Hence, the difficulty ratings of the materials met the relational criteria of extraneous cognitive load between instructional design, load type, and learning outcomes as defined by CLT (Gerjets et al., 2009). Because the validity of the extraneous cognitive load scale was successful, the subjective ratings can be interpreted as evidence that extraneous cognitive load mediates the split-attention effect. Therefore, both hypothesis *H 1.2.2b* and *H 1.2.2c* were supported by the results. Interestingly, correlational analyses showed that perceived task demands and interest ratings were related to the subjective ratings of extraneous cognitive load, thereby suggesting that the extraneous cognitive load ratings might be also influenced by motivational aspects of the learners. This finding shows that subjective ratings are sensitive to individual learner characteristics. This is an issue that needs further investigation.

Germane cognitive load. According to the germane cognitive load explanation germane cognitive load should mediate the split-attention effect, whereas according to the extraneous cognitive load explanation germane cognitive load is not responsible for the split-attention effect. Mediation analyses revealed that the rated level of concentration mediated the split-attention effect. Learners with integrated format reported higher levels of concentration and the higher the concentration was, the better were the learning outcomes. Hence, the subjective ratings of concentration level during learning met the relational criteria of germane cognitive load between instructional design, load type, and learning outcomes. Nevertheless, the ratings were also influenced by secondary task performance and learners' processing capacity. Despite the fact that the subjective ratings were sensitive to environmental demands and learner characteristics, the finding supports the germane cognitive load mechanism suggesting that germane cognitive load plays an important role in learning with integrated text-graphic formats. Therefore, hypotheses *H 1.2.2d* and *H 1.2.2.e* are interpreted as corroborated, whereas hypothesis *H 1.2.2g* is to be rejected..

However, the results do not completely meet the explanations proposed. The finding that extraneous and germane cognitive load were not related was not predicted. Rather, it was assumed that a reduction in extraneous cognitive load is accompanied by an increase in germane cognitive load. This should result in a negative relation, that could not be obtained. Therefore, hypothesis *H 1.2.2f* is to be rejected. The finding that a reduction in extraneous cognitive load does not automatically lead to an increased germane cognitive load can be interpreted in different ways. First, the findings that the subjective ratings were influenced by other factors than the instructional format alone show that the subjective ratings consist of a rather large error variance which might lead to statistical misspecifications, and thus, reducing the possibility to find shared variance. Alternatively, one might argue that instructional designs like a separated format generally cause higher extraneous cognitive load (as demonstrated by the group differences) but nevertheless leaves enough room for learners to individually decide how much germane cognitive load they want to invest. In such a case, individual differences in situational arousal or learning strategies may moderate learners' investment of germane cognitive load (Gerjets & Scheiter, 2003).

Moreover, the germane cognitive load explanation assumes that the specific cognitive load type pattern mediates the complete effect of instructional format on learning outcomes. The degree of mediation, however, of both extraneous and germane cognitive load explained about 20 % in the case of the labeling test and about 35 % in the case of complex facts only. Although being able to explain 20 – 35% variance can be considered a success, when dealing with complex processes like learning, there are still 80 - 65% of the variance left to be explained. One reason of the rather small mediation degrees might be that there are more factors mediating the learning outcomes besides instructional format and the related cognitive load types. Learners' cognitive prerequisites like processing capacity influenced for example three of four learning outcomes. Future research should investigate systematically further possible issues related to ability and to motivation (cf. Paas et al., 2005). Another reason might be related to measurement error. First, the analyses yielded that instructional format influenced the subjective ratings but that there was still a lot of unexplained variance. Second, the measurement of the load types was very parsimonious. Only one item was used for each cognitive load type. Hence, it was not possible to control for measurement error. However, a low reliability of a mediator's measure causes biased effects. In this case, the effect of the mediator on the outcome is underestimated, whereas the effect of the predictor on the outcome controlled for the mediator is overestimated (Kenny, 2006). For detailed analyses of the amounts concerning the effects between instructional format, cognitive load, and learning outcomes, future studies may use measurement error adjusted latent variable analyses that require multiple measures of a construct and/or measures that collect cognitive load measures multiple times during learning.

Interest. Although researchers advocating both explanations have considered spatial contiguity to only influence cognitive load, it was demonstrated that learners' interest was also influenced by the instructional format even though it was no mediator. Furthermore, participants' interest was related with their intrinsic cognitive load ratings. These findings suggest that instructional format might also be a relevant factor in keeping learners' persistence at school or university, especially when students have to study much longer materials than those used in the experiment (see Alexander & Jetton, 1996). Future studies may investigate in greater detail how instructional format influence different aspects of motivation that determines the direction, intensity, and persistence of human behavior (Heckhausen & Heckhausen, 2008). Moreover, to be better able to use and interpret subjective ratings, it is also important to understand how motivational issues like interest but also social desirability influence subjective cognitive load ratings.

5.4.3 Conclusions

Experiment 1 showed that not only extraneous but also germane cognitive load is an important mediator in learning with text-graphic instructions differing in spatial contiguity. This finding shows that a simple change in the instructional format has a profound effect of how learners perceive and presumably process the materials. Furthermore, it showed that CLT's strong focus on inhibiting processes is not enough to explain learning. Rather, both constraints and affordances (Greeno, 1994, 1998) inherent in an instructional format need to be considered more carefully. Therefore, one can conclude that there is no simple one to one correspondence between a specific instructional design characteristic and a single specific cognitive load type as often assumed in literature based on CLT or the CLTM. This conclusion can be derived from the Experiment's results based on the cognitive load measures developed to distinguish between the three cognitive load types. These results further suggest that differentiating measures seem to be an important tool to find out why specific instructional design characteristics are effective under specific conditions (e.g., learner characteristics; cf. Schnotz & Kürschner, 2007).

5.4.4 Limitations and Further Research Questions

Although Experiment 1 helped to clarify how the split-attention effect is mediated through cognitive load, its results seem to be too limited to conclude that all is now known about the influence of spatial contiguity between text and picture on multimedia learning. The following sections show which questions are still to be answered before the effect of spatial contiguity between text and graphic is understood. These questions concern learners' level of prior knowledge and learners' processing activities like reading behavior of text and graphic.

Level of prior knowledge. According to the findings of Experiment 1 it is tempting to generalize the results and claim that integrated formats generally support germane and reduce extraneous processing, whereas separated formats generally inhibit germane and increase extraneous processing. However, one important caveat for such a general conclusion is that only learners with almost no prior knowledge were tested. Thus, the results of Experiment 1 should not be generalized overly without considering individual differences in prior knowledge. Generalizing the results across all levels of prior knowledge would be especially questionable, because first, prior knowledge is known to be a pivotal learner characteristic in general (Dochy, Segers, & Buehl, 1999; Shapiro, 2004). Second, prior knowledge is an important factor in the models of the

CTML as well as CLT and Mayer and Sweller assume that it is a critical variable in learning. Mayer (2003) assumes that prior knowledge influences the mental integration processes of the verbal and pictorial model. Furthermore, Mayer and Gallini (1990) also assume that high-knowledge learners do not rely so much on pictorial information to build an integrated model, whereas low-knowledge learners do and therefore need to learn with integrated formats. Sweller et al. (1998) assume that prior knowledge influence intrinsic cognitive load because it is associated with element interactivity and schemas available. Furthermore, Sweller (2005) assumes that prior knowledge influences not only intrinsic cognitive load but also redundancy of information that is associated with extraneous cognitive load (Sweller, 2005). Whereas instructional design characteristics get first relevant when the topic to be learned is high in element interactivity (Chandler & Sweller, 1994), Kalyuga (2007) claims that extraneous cognitive load caused by redundancy moderates the effectiveness of different instructional designs as is demonstrated in the so called expertise reversal effect. Consequently, the questions arise: How does learners' prior knowledge influence their cognitive load during learning with integrated and separated text-graphic formats? And if their cognitive load is influenced, how do these changes of learners' cognitive load influence their learning outcomes? According to the focus on promoting processes suggested by the CTML, learners' prior knowledge might influence germane cognitive load and thereby their learning outcomes, whereas according to CLT's focus on inhibiting processes caused by poor instructions, learners' prior knowledge might influence intrinsic and extraneous cognitive load and thereby change their learning outcomes. However, it is still unclear how prior knowledge influences promoting or inhibiting processes during learning with integrated or separated formats. This lack of knowledge leads to the second limitation of Study 1 because it was not investigated how the participants actually processed the integrated and separated formats. Hence, so far it is still unknown which processing activities are related with which type of cognitive load.

Behavioral activities, learning processes, and cognitive load. A second limitation of Experiment 1 is that although overall cognitive load was measured by secondary task behavior and the three cognitive load types were measured separately by three subjective ratings scales, it is still unclear how learners handle or process integrated and separated formats. Do learners with separated format indeed switch more often between text and picture because of their need for mental integration as suggested by Erhel and Jamet (2006)? Or is it rather the other way round and learners with integrated format switch more often between text and picture as suggested by the results of learners' integrative saccades provided in the study of Holsanova et al.

(2009)? Moreover, one can ask whether prior knowledge influences how learners switch between text and picture. Because Experiment 1 did not record any learners' behavioral activities like viewing behavior, it does not provide any information about how learners actually process integrated and separated formats, neither does it provide any information about how such processing behavior like switching between text and illustrations represents cognitive load. To find out more about the relations of learners' activities and cognitive load and learning outcomes, it would be necessary to combine viewing behavior measures with other measures of cognitive load (e.g., subjective ratings) as well as with learning outcomes. It can be assumed that the more research is devoted to the measurement issue, the more will be learned about cognitive load. This will enhance our understanding of how integrated and separated formats are processed not only on the behavioral level but also on the cognitive level.

In order to increase our understanding of what happens in learners' minds during learning with integrated and separated text-graphic formats, Study 2 considered prior knowledge and learners' processing activities. However, before the question of how learners' prior knowledge influences the behavioral and related cognitive processing during learning with integrated and separated formats is approached experimentally, an overview on the research on prior knowledge and different instructional formats will be presented. This research has currently been subsumed under the term expertise reversal effect (Kalyuga et al., 2003).

6 The Expertise Reversal Effect: Cognitive Load Explanations and their Empirical Evidence

Fair is foul, and foul is fair.
(From Shakespeare's Macbeth; Act 1, Scene 1)

As shown in the work presented above, spatial contiguity between text and picture is an important characteristic of instructional formats which should be considered for learners with low prior knowledge due to its effects on extraneous and germane cognitive load and their effects on learning outcomes. However, how important is spatial contiguity between text and picture in multimedia instructions for learners with high prior knowledge? And if it is important, how does it influence the cognitive load of these learners? Without knowing an answer based on empirical results to the first of the two questions it can be assumed that spatial contiguity is also an important design characteristic for learners with high prior knowledge because of two reasons. First, prior knowledge is an important factor in the CTML as well as in CLT. Prior knowledge represents learners' schemas and therefore should influence how complex a topic is for differently knowledgeable learners. Second, there is this general phenomenon that instructional methods or formats that are highly effective for less knowledgeable learners can lose their effectiveness or even have detrimental effects when used with more knowledgeable learners. This phenomenon has been known for many years and was subsumed under the more general phenomena of Aptitude Treatment Interaction (ATI) effects (Cronbach & Snow, 1977). Among several aptitudes, prior knowledge showed the most consistent ATI effects (Bracht, 1980). Early ATI research (Cronbach & Snow, 1977) showed that low-aptitude learners benefit more from instructions with more guidance (e.g., integrated formats), whereas high-aptitude learners benefit more from instructions with less guidance (e.g., separated formats). This early conclusion is in line with more recent research results (Kalyuga, 2007). Since 2003, CLT researchers like Kalyuga et al. (2003) have called the general phenomenon that prior knowledge moderates instructional effectiveness *expertise reversal effect*. Because of this effect, instructional design researchers recommend to adapt instructions to learners' level of prior knowledge (Kalyuga, 2005; 2007; Snow & Lohman, 1984). Concerning the answer to the latter of the two questions above, however, no easy answer can be provided. As was the case with explaining the split-attention effect, there are competing explanations in the literature of how prior knowledge moderates the effect of spatial contiguity between text and picture. Generally, it is assumed that prior knowledge influences how learners learn from integrated or separated formats by influencing learners' cognitive load. Before a more detailed review on these explanations will be provided, the empirical evidence of an expertise reversal concerning spatial contiguity in multimedia learning is summarized in the following.

Mayer and Gallini (1990) were the first who discovered that prior knowledge plays also an important role in multimedia learning and spatial contiguity of text and illustrations. Differently knowledgeable students were asked to learn how mechanical devices work (Experiment 1: breaks; Experiment 2: pumps; Experiment 3: generators) with either a text only format, text with illustrations containing only the names of the parts (separated format[6]), text with illustrations containing only information of the processes (semi-separated format), or illustrations with integrated information about parts and processes (fully integrated format). The results of Experiment 1 demonstrated that low prior knowledge students with the fully integrated format recalled more conceptual knowledge and solved more problems than students with the other formats, whereas there were no differences in conceptual recall or problem solving performance among students with high prior knowledge. A similar pattern of results was demonstrated in Experiment 2 and 3. Although this study confounded spatial contiguity and multiple representations (text only vs. three levels of spatial contiguity), the effect seems to be caused at least partially by spatial contiguity of text and illustration, because later on Mayer et al. (1995, Experiment 2) replicated the effect. They demonstrated with text and static graphics that low-knowledge learners studying with integrated format performed better in problem solving tasks than low-knowledge students with separated format, whereas high-knowledge students with integrated format did not differ from high-knowledge students with separated format. A similar result was found by CLT researchers. Kalyuga et al. (1998) conducted a longitudinal study and demonstrated that novice learners benefited from integrated formats but suffered from separated formats and graphics only, whereas more advanced learners benefited from graphics only and suffered from integrated text-graphic formats. Although this study confounded the increasing complexity of learning task and learners' increasing prior knowledge and lacked the strict comparison with a

[6] Because only students with the illustrations including information of parts and processes did differ from students with the text only format, whereas students with lower levels of spatial contiguity did not differ from students with text only, I interpret the results in favor of a spatial contiguity effect (cf. Mayer, 2001).

separated format (cf. Experiment 1 and 3), the result pattern fits the expertise reversal effect. Hence, prior knowledge seems to moderate (neutralize or reverse) the split-attention effect generally found for learners with low prior knowledge.

In a rather recent review on the expertise reversal effect on a more general level, Kalyuga (2007) cited 26 research papers from several instructional design fields published between 1990 and 2007, which demonstrated the effect in 42 (partially combined) experiments. The mean effect size difference (d_{diff} = sum of ($d_{\text{effect size of high knowledge learners}} - d_{\text{effect size of low knowledge learners}}$); cf. Mayer, 2001) of the 50 effect size differences reported was 1.24 (SD = .60) with a minimum of .45 and a maximum of 2.99, indicating a rather large effect. Kalyuga (2007) focused predominantly on the interaction effects with regard to learning outcomes but missed to review and analyze the explanations provided for these effects as well as their empirical evidence thoroughly. Instead of a comprehensive review on the explanations, Kalyuga (2007) claimed that all expertise reversal effects (except of three studies investigating the imagery technique) were caused by extraneous cognitive load. According to the extraneous cognitive load explanation, already presented in Chapter 2.3, high-knowledge learners suffer from redundant information provided in well-guided formats (e.g., integrated formats, cohesive texts). Well-guided formats are said to force high-knowledge learners to “waste limited resources on co-referring internal and external representations of the same information” (p. 515; Kalyuga, 2007). Such extraneous processing leaves too few working memory resources for learning relevant processing. However, CTML as well as CIM researchers explain the same effect with germane cognitive load. These researchers assume that high-knowledge learners can handle or even benefit from unguided formats (e.g., separated formats, incohesive texts), because they can compensate for information gaps by actively applying their prior knowledge (e.g., Mayer & Gallini, 1990; McNamara & Kintsch, 1996).

Because the expertise reversal effect is a more general phenomenon and not restricted to one specific instructional design characteristic like for example spatial contiguity but subsumes specific characteristics under the more general characteristics like the guidance or clarity of structure of instructional formats, studies on the expertise reversal effect differ with respect to the specific instructional design used to investigate this effect. Because CLT with its triarchic model of cognitive load translates specific cognitive processes into the general three cognitive load types, it does not matter from a cognitive load perspective that the experiments reviewed were conducted in different domains (e.g., text comprehension, problem solving or language learning). What counts instead, is the argumentation on which type of cognitive load is caused by the instructional design depending on the level of prior knowledge. Consequently, the explanations provided in the literature can be theoretically transferred to integrated (higher guidance) and separated (lower guidance) formats, although the studies reviewed did not necessarily investigate spatial contiguity but a different specific instructional design characteristic. Hence, the evaluation of the evidence concerning the cognitive load explanations of the expertise reversal effect is based on a more general level comprising different specific instructional design characteristics that can be classified as more or less guided/structured. The following review of 51 studies consisting of 58 partially combined experiments focuses on the cognitive load measures that were used to provide empirical experiments in favor of the respective explanation. Thirty-four studies[7] listed in Kalyuga's (2007) review table of the expertise reversal effect were chosen as basis. Seventeen additional studies were included that comprised on the one hand, studies published after 2007 in a recent special issue on the expertise reversal effect (Kalyuga & Renkl, 2010), and on the other hand, the first studies of McNamara and colleagues. The reviewed experiments were analyzed and categorized with regard to the cognitive load explanation postulated. This procedure resulted in three groups arguing either for (1) compensatory processes increasing germane cognitive load, for (2) redundancy processes increasing extraneous cognitive load, or unexpectedly for (3) schema-influenced processes reducing intrinsic cognitive load. One study was exploratory without a specific assumption (Shin, Shallert, & Savenye, 1994), and thus, it was not categorized into any of the groups. Table 21 summarizes the reviewed studies, their theoretical framework, type of interaction[8], type of cognitive load measure, and whether the cognitive load measure could be interpreted in support of the respective explanation or not.

[7] The work of Reisslein (2005) was not included, because only experiments published in English peer-reviewed journals are considered in this dissertation thesis.

[8] The phenomenon that instructional design effects disappear with more knowledgeable learners (no differences in learning outcomes between the instructional design groups of more knowledgeable learners but differences between the groups of less knowledgeable learners) is statistically called ordinal or hybrid interaction, whereas the phenomenon that instructions with positive (negative) effects on less knowledgeable learners have negative (positive) effects on more knowledgeable learners (group differences between the instructional design groups of less and more knowledgeable learners in the opposite direction) is statistically called disordinal interaction (Leigh & Kinnear, 1980). Both phenomena are subsumed under the expertise reversal effect, although only a disordinal interaction demonstrates a complete reversal (Kalyuga et al., 2003).

Table 21

Summary of the reviewed papers addressing the expertise reversal effect

Study	Instructional formats	Theoretical framework	Type of Interaction	CL measure	CL measure supports explanation
GCL Explanation					
Amadieu, Tricot, & Mariné (2009) [a]	Hypertext with a hierarchical structure vs. with a network structure	CLT	Ordinal	Mental effort ratings	No
				Reading sequences	Yes
Amadieu, Van Gog et al. (2009) [a]	Hypertext with a hierarchical structure vs. with a network structure	CLT, CIM	Ordinal	Mental effort ratings	No
				Fixation durations	Possibly
Ayres (2006b) [a] *Experiment 2*	Interacting elements strategy vs. isolated elements strategy	CLT	Ordinal	Difficulty ratings	Possibly
Bodemer & Faust (2006) [a] *Experiment 1*	referencing vs. interactive integration	CLT, CTML	Ordinal	Drag-and-drop behavior	Possibly
Calisir & Gurel (2003)	Linear text vs. hypertext	—	Ordinal	—	—
Cooper, Tindall-Ford, Chandler, & Sweller (2001) *Experiment 4*	Studying vs. imagining	CLT	Disordinal	—	—
Ginns, Chandler, & Sweller (2003) *Experiments 1 and 2*	Studying vs. imagining	CLT	Disordinal	Studying times	No
Lambiotte & Danserau (1992)	Knowledge maps vs. lists	CMAE	Disordinal	—	—
Leahy & Sweller (2005) *Experiment 1*	Studying vs. imagining	CLT	Ordinal	—	—

Table 21 *(continued)*	Instructional formats	Theoretical framework	Type of Interaction	CL measure	CL measure supports explanation
Leahy & Sweller (2005) *Experiment 2*	Studying vs. imagining	CLT	Disordinal	—	—
Mayer & Gallini (1990) *Experiment 1*	Text with differently integrated illustrations vs. text only	CTML	Ordinal	—	—
Mayer & Gallini (1990) *Experiment 2*	Text with differently integrated illustrations vs. text only	CTML	Ordinal	—	—
Mayer & Gallini (1990) *Experiment 3*	Text with differently integrated illustrations vs. text only	CTML	Ordinal	—	—
Mayer et al. (1995) *Experiment 2*	Illustration with integrated vs. illustration with separated text	CTML	Ordinal	—	—
McNamara (2001) [a]	Sequence of in/cohesive text vs. cohesive-incohesive text	CIM	Disordinal	—	—
McNamara; Kintsch, Songer, & Kintsch (1996) [a, b] *Experiment 2*	Cohesive vs. incohesive text	CIM	Disordinal	Studying times	No
McNamara & Kintsch (1996) [a] *Experiment 1*	Cohesive vs. incohesive text	CIM	Ordinal	Studying times	Yes
McNamara & Kintsch (1996) [a] *Experiment 2*	Cohesive vs. incohesive text	CIM	Ordinal	Studying times	Yes
Ollerenshaw, Aidman, & Kid (1997)	Text with animated simulation vs. text only	CTML, CLT	Ordinal	—	—
Potelle & Rouet (2003)	Structured hierarchical maps vs. semantic network maps	CIM	Ordinal	—	—
Schnotz & Rasch (2005) *Experiment 1*	Animated vs. static pictures	CLT	Ordinal	Picture inspection times	Yes
Schnotz & Rasch (2005) *Experiment 2*	Manipulation vs. simulation pictures	CLT	Ordinal	—	—

Table 21 *(continued)*	Instructional formats	Theoretical framework	Type of Interaction	CL measure	CL measure supports explanation
Seufert (2003)	Directive vs. non-directive help vs. no help	CTML, CIM	Ordinal	—	—
Seufert, Jänen, & Brünken (2007) [a] *Experiment 3*	Inter-representational hyperlinks vs. no help	CLT	Ordinal	Mental effort ratings	No
Shapiro (1999)	Text with vs. text without structuring overview	CIM	Ordinal	—	—
ECL Assumption					
Blayney, Kalyuga, & Sweller (2010) [a]	Isolated vs. interacting elements format	CLT	Ordinal	—	—
Homer & Plass (2010) [a]	Symbolic vs. iconic presentations	CLT	(Dis)Ordinal	—	—
Kalyuga (2008)	Animated vs. static diagrams	CLT	Disordinal	—	—
Kalyuga et al. (1998) *Experiments 1 and 3*	Diagram with integrated text vs. diagram only	CLT	Disordinal	Difficulty ratings	Yes
				Studying times	Yes
Kalyuga, Chandler, & Sweller (2000)	Animated diagram with narrated text vs. diagram only	CLT	Disordinal	Difficulty ratings	Yes
Kalyuga et al. (2001) *Experiment 1*	Worked examples vs. problem solving	CLT	Ordinal	Difficulty ratings	No
Kalyuga et al. (2001) *Experiment 2*	Worked examples vs. problem solving	CLT	Ordinal	Difficulty ratings	No
Kalyuga & Sweller (2004) *Experiment 3*	Worked examples vs. problem solving	CLT	Disordinal	—	—
Lee, Plass, & Homer (2006)	Symbolic vs. iconic-symbolic presentations	CLT, CTML	Disordinal	—	—

Table 21 *(continued)*	Instructional formats	Theoretical framework	Type of Interaction	CL measure	CL measure supports explanation
Nückles, Hübner, Dümer, & Renkl (2010) [a] *Experiment 1*	(Meta)cognitive prompts vs. no prompts in writing	CLT	Ordinal	Mental effort ratings	No
Nückles, Hübner, Dümer, & Renkl (2010) [a] *Experiment 2*	Permanent (meta)cognitive prompts vs. faded (meta)cognitive prompts	CLT	Ordinal	Mental effort ratings	No
Oksa, Kalyuga, & Chandler (2010) [a] *Experiments 1 and 2*	Modern explanatory notes in historical literature vs. no notes	CLT	Disordinal	Difficulty ratings	Yes
				Retrospective reports	Possibly
Pawley, Ayres, Cooper, & Sweller (2005) *Experiment 1*	Checking strategy vs. non-checking	CLT	Ordinal	—	—
Pawley, Ayres, Cooper, & Sweller (2005) *Experiment 2*	Checking strategy vs. non-checking	CLT	Disordinal	Difficulty ratings	Yes
Reisslein, Atkinson, Seeling, & Reisslein (2006)	Example-problem vs. problem-example sequence	CLT	Ordinal	Studying times	No
Salden, Aleven, Schwonke, & Renkl (2010) [a] *Experiment 1*	Problem solving vs. fixed fading vs. adaptive fading in software tutors	CLT	Ordinal	Studying times	No
Salden, Aleven, Schwonke, & Renkl (2010) [a] *Experiment 2*	Problem solving vs. fixed fading vs. adaptive fading in software tutors	CLT	Ordinal	Studying times	No
Van Gog, Paas, & Van Merriënboer (2008) [a]	Sequences of process- and product-oriented worked example	CLT	Ordinal	Mental effort ratings	No
				Studying times	Yes

Table 21 *(continued)*	Instructional formats	Theoretical framework	Type of Interaction	CL measure	CL measure supports explanation
Yeung et al. (1997) *Experiments 2 and 3*	Vocabulary definitions in integrated vs. separated format	CLT	Disordinal	—	—
Yeung et al. (1997) *Experiments 4 and 5*	Vocabulary definitions in integrated vs. separated format	CLT	Disordinal	Difficulty ratings	Yes
ICL Explanation					
Clarke et al. (2005)	Sequential vs. concurrent presentation	CLT	Ordinal	Difficulty ratings	Yes
Kalyuga, Chandler, & Sweller (2001) *Experiment 2*	Worked examples vs. exploratory learning	CLT	Ordinal	Difficulty ratings	Yes
				Studying times	Yes
Pollock et al. (2002) *Experiments 1 and 2*	Isolated-interacting elements vs. interacting-interacting elements sequence	CLT	Ordinal	Difficulty ratings	Yes
				Studying times	Possibly
Pollock et al. (2002) *Experiments 3 and 4*	Isolated-interacting elements vs. interacting-interacting elements sequence	CLT	Ordinal	Difficulty ratings	No
Tuovinen & Sweller (1999)	Worked examples vs. exploratory learning	CLT	Ordinal	Mental effort ratings	Possibly
No explanation					
Shin, Shallert, & Savenye (1994)	Hypertext with limited-access vs. with free-access	—	Ordinal	Difficulty ratings	*n.d.*
				Studying times	*n.d.*

Note. [a] experiment not included in Kalyuga's (2007) table on summarized results of the expertise reversal effect. [b] experiment cited in Kalyuga (2007) but not included in his summary table. [c] experiment not cited in Kalyuga (2007). CL = Cognitive load, GCL = germane cognitive load, ECL = extraneous cognitive load, ICL = intrinsic cognitive load. CLT = Cognitive Load Theory, CTML = Cognitive Theory of Multimedia Learning, CIM = Construction-Integration Model, CMAE = Cognitive Model of Assimilation Encoding (Mayer, 1979), *n.d.* = not definable.

6.1 Literature Review of Expertise Reversal Studies

The next sections present a qualitative review of central studies on the expertise reversal effect and their cognitive load explanations. Some further studies not included in Table 21 are cited, if they contribute to the argumentation of the respective explanation but are otherwise out of the scope of this review. This review starts with the overview of studies arguing for the germane cognitive load explanation because this explanation was the established one that has been challenged by Kaluga et al. (1998).

6.1.1 Germane Cognitive Load Explanation: Compensatory Processing

Scientists from different research fields assume that cognitive processes causing germane cognitive load are responsible for the expertise reversal effect. Their assumptions are provided in the context of the CTML as well as in the context of Kintsch's *Construction-Integration Model* (CIM) developed in text comprehension research. The CIM is based on a connectionist approach that is beyond the scope of this thesis (for a detailed overview see Kintsch, 1988; 1998; McNamara, 2009). Moreover, some CLT researchers also argue with referring to germane cognitive load. The selected studies are presented according to the type of cognitive load measurement used or behavioral processing data collected to support the explanation.

Learning outcomes. Learning outcomes are not considered to be measures of cognitive load in this dissertation thesis (but cf. Brünken et al., 2003). Nevertheless, the following studies are subsumed under this label because they did not use any cognitive load measur but relied on interpreting students' learning outcomes only. In explaining the expertise reversal effect found for multimedia instructions that differ in spatial contiguity, Mayer and colleagues (Mayer & Gallini, 1990; Mayer et al., 1995) argued that high-knowledge learners do not suffer from separated formats, because by using their prior knowledge high-knowledge learners can compensate for a lack in instructional guidance compared to low-knowledge learners. Mayer (2001) states that this assumption is "based on the idea that high-knowledge compensates for poor instructions" (p.167). He assumes that while reading the text information high-knowledge learners are able to apply imagery strategies, and thus, do not depend on pictorial information (cf. Alexander & Judy, 1988). Such active learning processes involved in imagery strategies correspond to generative processing or germane cognitive load and help high-knowledge learners to focus on the relevant information, so that explicit illustrations are not needed. Therefore, according to Mayer one can assume that high-knowledge learners do not switch very frequently between text and corresponding pictorial information to build a coherent mental representation. Rather, high-knowledge learners are expected to focus on the textual information and to actively use their domain knowledge. Empirical evidence with regard to any type of cognitive load measure or to process measure, like for example learners' viewing behavior during studying, however, has not been reported by Mayer so far. Hence, Mayer's assumptions are based on theoretical considerations and learning outcomes of different knowledge tests only.

Another study that did neither measure any cognitive load type, and therefore does also not provide any empirical evidence on the cognitive load mechanism underlying the expertise reversal effect, is the one by Seufert (2003). This study is described more thoroughly, because according to Kalyuga (2007) it is "a relevant study" (p. 521) that shows that high-knowledge learners do not need instructional guidance but benefit from unguided instructions according to his extraneous cognitive load explanation. Notably, this study was presented in the framework of both the CTML and the CIM but not within the framework of the CLT. Seufert examined whether and which type of help supported the generation of referential connections between corresponding verbal and pictorial information in multimedia learning. The instructional help was presented either in a more directive or non-directive way. Learners were distinguished not only in more and less knowledgeable but in low-, medium-, and high-knowledge learners. Seufert demonstrated that low-knowledge learners did not benefit from help in an instruction about biochemical processes or might even suffer from non-directive help, whereas medium-knowledge learners benefited from help, especially from directive help. High-knowledge learners had higher learning outcomes than less knowledgeable learners but were not affected by help. Seufert argued that low-knowledge learners lacked the necessary prior knowledge to profit from the help, whereas medium-knowledge learners' prior knowledge was activated by the help and made them focus on the relevant information. Concerning high-knowledge learners Seufert argued that they remain inactive and did not use the help. She suggested that this might be due to an illusion of knowing. She further argued that high-knowledge learners should be prompted in order to use the help more actively (such processes might be interpreted as germane cognitive load). With respect to the low-knowledge learners who did not benefit from help she discussed the possibility that this might be due to cognitive overload. If one ignores the small number of participants, this study seems to demonstrate the basic assumption of the expertise reversal effect, namely that high-knowledge learners can benefit from more unguided instructions or do at least not suffer from them. However, the study does not help to clarify which cognitive load mechanism underlies the expertise reversal effect, because it provides not even a single cognitive load measure. Nevertheless, it is a good example showing how Kalyuga's review (2007) more or less ignored not only the authors' own theoretical argumentation but also the lacking empirical evidence in favor of or against the argumentations, thereby biasing readers towards the extraneous cognitive load explanation.

Studying times. Although many studies on the expertise reversal effect did not measure any cognitive load measure and just relied on learning outcomes, some studies analyzed studying times. McNamara and colleagues investigated the effect of text cohesion on learners' comprehension outcomes in the context of the CIM. In a series of studies, they demonstrated that low-knowledge learners benefited more from high-cohesive texts (e.g., many anaphoric referents, sentence connectives, background information, as well as meaningful headings and paragraphs), whereas high-knowledge learners benefited more from low-cohesive texts that lack structuring information. This result pattern was first demonstrated by using differently cohesive versions of a biology text about heart disease (McNamara et al., 1996) and differently cohesive versions of a history text about the Vietnam War (McNamara & Kintsch, 1996). McNamara and colleagues argued that low-cohesive texts force high-knowledge learners to engage in compensatory processing to infer unstated relations in these texts (germane processing), whereas high-cohesive texts seduce high-knowledge learners to more passive processing instead of activating relevant prior knowledge on their own. Their assumption was supported by the finding that high-knowledge learners with low cohesive texts had longer reading times than equally knowledgeable learners with low cohesive texts (McNamara & Kintsch, 1996). Although studying times are a critical cognitive load measure, longer reading/studying times in relation with higher learning outcomes can be interpreted as germane cognitive load (see Chapter 3.2). However, the studying times of learners did not differ in the study of McNamara et al. (1996).

Besides measuring studying times, McNamara and colleagues claimed to have provided more evidence in favor of the germane cognitive load explanation by investigating the moderating effect of reading skills. In a more recent study on text cohesion, prior knowledge and reading skills, O'Reilly and McNamara (2007) showed that learners with high prior knowledge and low reading skills did not benefit from high-cohesive texts (as expected according to the expertise reversal effect), whereas skilled learners with high knowledge and high reading skills did successfully learn with high-cohesive texts. The authors argued that the latter learners who had high reading skills were not seduced to inactive processing despite high text cohesion. According to this argument, good reading skills help high-knowledge learners to engage in germane processing and prevent them from a passive reading behavior. Reading skills in high-knowledge readers thus work against cognitive inactivity. Similar results were shown by Ozuru, Dempsey, and McNamara (2009). Thus, high-knowledge learners do not benefit or suffer from high-cohesive texts only, when they are seduced to not engage in germane processing. However, if high-knowledge learners are skilled readers, they also know how to apply active processing strategies with well-guided texts, and hence, can also benefit from high-cohesive texts. With taking reading skill as second moderator (and prior knowledge as a first moderator) into account, O'Reilly and McNamara (2007) more or less reversed the expertise reversal effect with skilled readers. This was demonstrated by a three-way-interaction of text cohesion, prior knowledge, and reading skill. Although these results and argumentations are very interesting, no further empirical evidence of any cognitive load measure was provided.

Schnotz and Rasch (2005) also provided an argumentation based on germane cognitive load and processing time in a study on static pictures (lower guidance) and animations (higher guidance). However, these authors assumed that animations inhibit low-knowledge learners to invest germane cognitive load. In their experiment 1, half of the participants were asked to learn time and date differences on earth with a hypertext and so called manipulation and simulation pictures that were animated in a specific way, when learners clicked specific buttons. The other half of the learners were asked to study the topic with the hypertext and static versions of the manipulation and simulation pictures. Concerning circumnavigation questions, learners with high learning prerequisites (combination of higher intelligence test scores and higher prior knowledge) did benefit equally well from the animated and static pictures, whereas learners with low learning prerequisites benefited more from static pictures than animated pictures. Because learners with low learning prerequisites inspected the static pictures longer than animated pictures, Schnotz and Rasch assumed that the animations seduced the learners with low learning prerequisites to not use their mental capacity. Because longer picture inspection times were related with higher learning outcomes in low-knowledge learners, picture inspection times were interpreted as indicators of germane cognitive load (see Chapter 3.2). Although the interpretation of inspection times refers to germane cognitive load, it should be noted that not high-knowledge learners but low-knowledge learners were assumed to be inhibited to invest germane cognitive load when studying well-guided instructions (in that case animations). This argumentation is at odds with the general germane cognitive load explanation. However, one can assume that high-knowledge learners did not suffer from animations because they were skilled enough to invest germane cognitive load in processing them.

Subjective ratings. In the context of CLT and the expertise reversal effect, Ayres (2006b) argued in accordance with the germane cognitive load explanation. In this study, cognitive load was measured by means of subjective difficulty ratings. Ayres investigated whether the solving of only one calculation per mathematical bracket problem (isolated element strategy with lower intrinsic cognitive load) or whether the solving of all four calculations in such a problem (interacting element strategy with higher intrinsic cognitive load) is more effective. The study showed that less knowledgeable students benefited from an isolated elements strategy in solving mathematical bracket problems compared to an interacting element strategy, whereas more knowledgeable learners benefited from an interacting elements strategy but suffered from an isolated elements strategy. Learners were asked to rate the perceived difficulty of each calculation. Because the isolated elements strategy was rated the least difficult one by all students, Ayres suggested that more knowledgeable learners studying with the isolated-elements strategy did not engage in sufficient germane processing, because they perceived the strategy as too simple. Ayres also argued that more knowledgeable learners are seduced to not engage in germane processing and stay passive, when they are not challenged by the instructional design. The perceived difficulty ratings of that study seem to be a measure of intrinsic cognitive load (cf. Chapter 3.2), because Ayres did not manipulate the instructional design between groups. However, according to Kalyuga (2007) one might disagree because he argues that techniques and procedures reducing intrinsic cognitive load for novices may become redundant for experts and thereby increase extraneous cognitive load. According to this extraneous cognitive load argumentation, the perceived difficulty measures might actually be a measure of extraneous cognitive load. However, if one takes Kalyuga's argument serious and argues that the perceived difficulty ratings represented extraneous cognitive load (see Chapter 3.2), then the result clearly contradicts the extraneous cognitive load explanation. This explanation postulates that high-knowledge learners with the isolated element strategy should be loaded by higher extraneous cognitive load compared to high-knowledge learners with the interacting element strategy. However, this was not the case. Hence, Ayres (2006b) stated that his results are better explained with a lack of germane cognitive load in high-knowledge learners than with an increase in extraneous cognitive load as suggested by Kalyuga et al. (2003). As perceived difficulty is not considered to be a direct measure of germane cognitive load, however, the finding can only be taken as possible indirect evidence in favor of the germane cognitive load explanation.

Different from Ayres, Seufert et al. (2007) collected mental effort ratings. They argued that high-knowledge learners benefited from inter-representational hyperlinks between text and illustrations in a multimedia learning environment about biochemical functions of vitamin C, because these learners should be loaded only minimally by intrinsic cognitive load, and therefore should have enough capacity left to engage in integrative processing elicited by inter-representational hyperlinks (germane processing). Low-knowledge learners should be loaded by too high intrinsic cognitive load, and therefore cannot engage in integrative or germane cognitive load processing, even when provided with inter-representational hyperlinks. Thus, whether low-knowledge students learn with inter-representational hyperlinks or with a separated format should make no difference. These assumptions were supported by the comprehension test result. Mental effort ratings showed that the difference of the load ratings between the two instructional groups was bigger for low-knowledge learners than for high-knowledge learners. This result is not in line with the hypothesized germane cognitive load explanation. However, if one argues that both the extraneous as well as the germane cognitive load mechanism worked simultaneously (as shown in Experiment 1 of this dissertation), one would not expect differences between the two groups of high-knowledge learners in mental effort ratings. Low-knowledge learners without hyperlinks might have had higher extraneous cognitive load than those with hyperlinks, whereas high-knowledge learners without hyperlinks might have invested germane cognitive load according to the germane cognitive load explanation and those with hyperlinks might have suffered from extraneous cognitive load according to the extraneous cognitive load explanation (i.e. processing of redundant information). Such interpretations of overall cognitive load measures like mental effort, however, stay highly hypothetical. In case that both explanations work, these measures do not provide substantiated empirical evidence for or against one of the cognitive load explanations.

Two recent studies about hypertext learning also used mental effort ratings as cognitive load measure. Amadieu, Tricot et al. (2009) as well as Amadieu, Van Gog et al. (2009) showed that low-knowledge learners benefited from a hierarchical hypertext concept-map consisting of organizational links compared to a network hypertext map consisting of relational links, whereas high-knowledge learners benefited equally well from both types of hypertext structures. According to the germane cognitive load explanation high-knowledge learners with the network hypertext should have had higher mental effort ratings than high-knowledge learners with the hierarchical hypertext. However, the authors did not find any differences between these groups. As mental effort ratings may be the result of all three load types representing total cognitive load, it might be possible that the ratings did not differ between the groups, hypothesizing again that the germane as well as the extraneous cognitive load explanation worked simultaneously. The groups might just differ in their cognitive load type patterns but not in their overall cognitive load. Whereas high-knowledge learners with hierarchical hypertext were maybe loaded by extraneous cognitive load (i.e., processing of redundant information), high-knowledge learners with the network hypertext were maybe engaged in germane cognitive load. Hence, if both mechanisms would work, overall cognitive load measures are not very useful to differentiate between these explanations. However, because high-knowledge learners benefited equally well from both hypertext types (ordinal interaction), it might also be possible that these learners did just not differ in cognitive load types at all.

Behavioral activities. More fine-grained information was provided by few studies that gathered some types of behavioral activity data. These studies are presented in the following. In contrast to Mayer's assumption, Bodemer and Faust (2006) doubted that learners invest germane cognitive load on their own just because text is integrated into illustrations Thus, they recommend dragging and dropping as a method to actively enhance germane cognitive load and to decrease unnecessary visual search processes. This active integration approach was tested against different alternative instructional designs. The authors demonstrated that only high-knowledge learners studying the principles of heat pumps benefited from the active integration of textual information into illustrations by dragging and dropping that should elicit processing that can be interpreted as germane cognitive load. Low-knowledge learners, however, suffered from the active integration method. The authors showed by means of the log file protocols of learners' dragging and dropping behavior that some of the low-knowledge learners were just not able to successfully integrate all relevant textual information into the graphic. This result suggests that only high-knowledge learners were able to actively apply the relevant prior knowledge that was needed to compensate the missing relations between text and illustration. Hence, similar to Seufert's (2003) explanation, Bodemer and Faust (2006) argued that some learning activities need a specific amount of prior knowledge that can be used for germane processing.

Moreover, in a series of studies investigating the imagery strategy, the active processes of imagery are assumed to be responsible for the so called imagery effect (e.g., Leahy & Sweller, 2005). In several studies comparing the two learning strategies studying and imagining, it was demonstrated that high-knowledge learners profited from imagining problem solution steps compared to studying only, whereas low-knowledge learners profited from studying the learning materials only compared to imagining the problem solution steps to be learned (Cooper et al., 2001; Ginns et al. 2003; Leahy & Sweller, 2005). Sweller and colleagues assumed that high-knowledge learners have already acquired the necessary schemata by studying and that imagery supports the automation of these schemata. For more advanced learners, schema automation is more important than focusing on schema construction. Think-aloud protocols from an imagery study (not addressing the expertise reversal effect) showed that advanced learners in an imagery condition engaged in an imagination process that differed from the normal studying process (Leahy & Sweller, 2004). Learners in the imagery condition used working memory to rehearse problem solving procedures (schema automation), whereas learners in the studying condition made no attempts to imagine the solution procedures but rather tried to understand the solution (schema construction). Although the imagery effect is meanwhile explained by germane cognitive load (see Kalyuga, 2007; Sweller, 2009a), the first studies on the imagery effect were less clear in their explanation and referred partially to extraneous processing in high-knowledge learners in the studying conditions as expressed in sentences like "studying the material again is now a redundant activity that interferes with additional learning." (Leahy & Sweller, 2004, p. 274). This argumentation is in line with the extraneous cognitive load explanation of expertise reversal effects referring to redundancy.

The most recent studies that collected several behavioral activity data during learning were the above mentioned hypertext concept-map studies by Amadieu and colleagues (2009). Amadieu, Tricot et al. (2009) showed that high-knowledge learners followed more coherent reading sequences than low-knowledge learners when using a network hypertext concept-map. Thus, these authors suggested that high-knowledge learners are able to compensate the lack of organizational cues in network hypertexts, because "they are able to process non-linear information building active reading sequences based on semantic coherence of the contents" (p.387). On the other hand, low-knowledge learners have not enough prior knowledge to compensate for information gaps and do not find meaningful reading sequences on their own, and therefore suffer from network hypertexts. However, the result of the reading sequences was not that clear in the study of Amadieu, Van Gog et al. (2009). However, Amadieu, Van Gog et al. (2009) also measured learners' fixation duration by means of the eye tracking methodology. In line with the learning outcomes on concept knowledge, they found differences between the low-knowledge learners but not for the high-knowledge learners. This result might suggest that low-knowledge readers with the hierarchically organized concept-map hypertext engaged in more germane processing as reflected in longer fixation times than those with the network structure, whereas high-knowledge learners were able to deal with both structure types equally well. Whether high-knowledge learners' fixation times did not differ because those learners with the hierarchical concept-map had somewhat higher extraneous cognitive load, whereas those with the network concept-map had somewhat higher germane cognitive load might be possible but remains unclear.

6.1.2 Extraneous Cognitive Load Explanation: Redundant Processing

As already mentioned, Kalyuga (2005; 2007; Kalyuga et al., 1998; 2003) has challenged the established assumption that germane processing is responsible for the expertise reversal effect. Kalyuga assumes that high-knowledge learners do not benefit or even suffer from high-structured materials (e.g., integrated formats) because they are overloaded by extraneous processing due to redundant information.

Learning outcomes. The following three experiments relied on learning outcomes only and did not use any cognitive load measure. In a series of experiments about where to put explanatory notes in foreign language texts, Yeung et al. (1997) showed that explanatory notes in reading passages presented in an integrated format enhanced 5th-graders comprehension but not their vocabulary knowledge compared to a separated format (Experiment 2; split-attention effect). However, adult readers with an integrated format showed lower comprehension scores but higher vocabulary knowledge scores compared to the separated format (Experiment 3). According to Yeung et al. (1997) this result pattern indicated that for comprehension the presence of vocabulary meanings integrated in the text increased extraneous load because of redundancy for high-knowledge learners, and thus, reduced comprehension performance of these learners. For vocabulary learning, however, the integrated format tended to reduce cognitive load, and thus, resulted in higher vocabulary performance compared with the separated format. Although the argumentation was not tested by obtaining any cognitive load measures, these combined experiments were taken as evidence in favor of the extraneous cognitive load explanation and gave rise to research on the expertise reversal effect in CLT research.

Studying times. Several studies measured studying times to provide some evidence in favor of the extraneous cognitive load explanation. For example, Reisslein et al. (2006) tested the sequence of worked-examples – problem solving versus problems solving - worked-examples versus fading in the domain of parallel electrical circuit analysis. The authors demonstrated that high-knowledge learners outperformed low-knowledge learners in the problem solving – worked example sequence condition but they did not differ from low-knowledge learners in the other two conditions. According to the extraneous cognitive load explanation, high-knowledge learners in the problem solving – worked example condition should experience lower extraneous cognitive load and therefore also study shorter than those in the worked example – problem solving condition. Studying times however, did not differ between the groups. Hence, they did not support the ECL explanation.

Similar results were demonstrated by Salden et al. (2010). They investigated the effectiveness of different fading conditions in a computer based cognitive tutor about angles geometry. Because learners acquire knowledge with practice it was argued that worked examples should be gradually faded out to keep redundancy and extraneous cognitive load low. Furthermore, it was argued that the best method of fading out would be a mechanism adaptive to the learners understanding in comparison to a fixed adaptation mechanism or problem solving tasks only. Although learners of the adaptive fading condition outperformed the learners of both other conditions, the studying times did not differ between the groups. This result was found in a lab as well as in a classroom experiment. Studying times, however, were in line with the extraneous cognitive load explanation in a study by a Van Gog et al. (2008) on the sequence of process- and product-oriented worked examples and in two combined experiments by Kalyuga et al. (1998) on integrated formats and illustrations only (a more detailed description of the experiments see below).

Subjective ratings. The expertise reversal effect in the two above mentioned experiments (Experiments 2 and 3) by Yeung et al (1997) were also found in two further combined experiments (Experiments 4 and 5) with 8th-graders in Hong Kong with low vs. high prior knowledge in English as their second foreign language. In these experiments, however, difficulty ratings were obtained. Low-knowledge students' perceived difficulty of comprehension and of vocabulary did not show a significant interaction effect with instructional format. However, high-knowledge students' perceived difficulty of comprehension was higher in the integrated format condition than in the separated one but perceived difficulty ratings of the vocabulary was lower in the integrated format condition than in the separated one. If the perceived difficulty ratings measured extraneous cognitive load only as assumed, the ratings of the high-knowledge learners are in line with the extraneous cognitive load explanation.

Further evidence supporting the extraneous cognitive load explanation was provided by research on multimedia instructions. In two experiments about the instruction of electric circuits, Kalyuga et al. (1998) showed that advanced learners performed better with a diagram only format than with an integrated text-diagram format, whereas low-knowledge learners performed better with an integrated format than with a separated or diagram only format. Kalyuga et al. claimed that high-knowledgeable learners had already acquired the necessary schemata to understand the diagram in isolation. Thus, the verbal information integrated into the diagram did not only become unnecessary but rather redundant for these learners by placing an excessive extraneous cognitive load on their working memory. This assumption was supported by studying times as well as by subjective ratings of difficulty. High-knowledge learners with the integrated format studied longer and rated the difficulty of the integrated format higher than high-knowledge learners with the diagram only format.

In several studies comparing worked examples with problem solving, however, the difficulty measures provided a rather unclear picture. For example, in the domain of writing programmable logic controller programs for relay circuits, Kalyuga, Chandler, Tuovinen, and Sweller (2001) demonstrated that advanced learners with problem solving tasks did not differ from advanced learners with worked examples, whereas novices benefited from worked examples in two experiments. Kalyuga et al. (2001) assumed that when knowledge increases, the need for worked examples to demonstrate the problem solution decreases, because guidance can shift from an external source to internal schemata. At some point worked examples become redundant for advanced learners. The difficulty ratings of these experiments did not fully support the extraneous cognitive load explanation, because the groups with high-knowledge learners did not differ. With regard to non-significant learning outcome results, the question arises, however, whether it is necessary at all that the difficulty ratings have to differ between these groups.

Other authors used mental effort as cognitive load measure. Although Van Gog et al. (2009) showed a disordinal expertise reversal effect on reversed sequences of different types of worked examples, mental effort measures were only collected during the knowledge tests, and therefore, cannot be used as cognitive load indicators during learning (see Chapter 3). Moreover, mental effort measures did not support the extraneous cognitive load assumption in a study on (meta)cognitive prompting in learning how to write short articles (Nückles, et al. 2009). In contrast, mental effort ratings of experienced writers with prompts decreased even more strongly over time compared to the ratings of experienced writers without prompts, although mental effort should have increased according to the extraneous cognitive load explanation. Interestingly, it was also found that the enjoyment of experienced writers in the prompts condition decreased in a similar pattern like their mental effort ratings. Hence, the finding would better fit with the germane cognitive load explanation stating that advanced learners with well-guided instructions do not invest germane cognitive load, and thus, have lower overall cognitive load as indicated by the mental effort measures.

Behavioral activities. To find out more about the processes during studying some authors asked learners to retrospectively report about their learning process.

Oksa et al. (2010) integrated modern English interpretations into Shakespearean play extracts presented in original Elizabethan English. The retrospective reports indicated that the experts not only read the modern interpretations and reflected on them but also that they felt interrupted by the interpretations or even that they did not share the interpretations. Only one expert found them interesting and helpful. Whereas Oksa et al. (2010) interpreted these reports as indication that experts crosschecked redundant information that caused extraneous cognitive load, it is not clarified so far, whether the task of text interpretation is comparable to tasks like understanding biological or physical systems with a well-defined functioning. The finding that several experts did not share the interpretations provided (conflicting prior knowledge) might suggest that other mechanisms might apply in ambiguous domains like literature than when dealing with the objective functioning of biological or mechanical systems.

6.1.3 Intrinsic Cognitive Load Explanation: Schematized Processing

Unexpectedly, there are several studies in the expertise reversal effect literature whose explanations of the effect do not correspond with the two main explanations introduced above. In these studies, intrinsic processing is regarded as the main mechanism underlying the expertise reversal effect, even though lower intrinsic cognitive load can be related to lower extraneous cognitive load to cause the expertise reversal effect. Hence, germane cognitive load is not assumed to be directly responsible for the expertise reversal effect in these studies, whereas extraneous cognitive load can play a role. In these studies, the overall argumentation was that high-knowledge learners have lower intrinsic cognitive load, and therefore, are not overloaded by extraneous cognitive load due to unguided instructions so that these instructions can be handled. All studies in this category were discussed in the context of the CLT.

Studying times. As mentioned in Chapter 3, studying times can be used to indicate intrinsic cognitive load. A reduced intrinsic cognitive load in high-knowledge learners was made responsible for the expertise reversal effect in four experiments about how much information about the interrelatedness of interacting elements should be presented (Pollock et al., 2002). Pollock et al. (2002) compared instructions containing all information about the elements and their interrelations among each other (interacting elements instruction) with instructions containing the elements without their interrelations among each other (the isolated elements instruction). Without enough prior knowledge the interacting elements instruction should cause too much intrinsic cognitive load. In two combined studies about information on electrical safety tests and in two combined studies about the complex electrical circuit of an industrial oven, it was demonstrated that low-knowledge learners benefited when they studied first with an isolated elements instruction and afterwards with an interacting elements instruction compared with low-knowledge learners studying twice with an interacting elements instruction. In contrast, there was no difference between high-knowledge students. The authors argued that isolated elements can be easily held and processed in working memory, and thus, can be easily learned although with a reduced understanding. However, once the elements are learned, the interrelations among the elements can be learned with a reduced intrinsic cognitive load because of sufficiently established schemata, thus, facilitating learning. Hence, if advanced learners have acquired sufficient schemata to process all the elements and the required interrelations simultaneously in working memory, there is no more need for an isolated elements approach and no differences between the two instructions for high-knowledge learners are expected. This assumption was partially corroborated by studying times. Low-knowledge learners with an isolated-interacting elements sequence studied marginally shorter than low-knowledge learners with an interacting-interacting elements sequence in Experiment 1, whereas there were no differences between high-knowledge learners in the second phase (interacting elements) of Experiment 2. Because the results of all four groups of both experiments were not analyzed in one single analysis, it is somewhat difficult to assess an assumed interaction. According to the means reported, however, high-knowledge learners obviously studied shorter than low-knowledge learners indicating that high-knowledge learners benefited from their prior knowledge. Similar results to Experiment 1 were found in the study on worked examples and exploratory learning in the study by Kalyuga et al. (2001).

Subjective ratings. Except for one study all other studies of this category reported difficulty ratings. The just aforementioned study by Pollock et al. (2002) showed that low-knowledge learners with the isolated-interacting sequence in Experiment 1 rated the difficulty of the materials lower than the low-knowledge learners with the interacting only sequence, whereas there were no differences between the high-knowledge learners in Experiment 2. These two results are in parallel with the learning outcomes and support the intrinsic cognitive load explanation, if the difficulty ratings measured intrinsic cognitive load. However, the differences in perceived difficulty of low-knowledge learners were not replicated in Experiment 3 thereby not supporting the explanation that low-knowledge learners in the interacting condition were loaded by higher intrinsic cognitive load than the learners in the isolated condition. The reduced intrinsic cognitive load in high-knowledge learners was also said to explain why high-knowledge learners were able to study well with exploratory instructions, whereas low-knowledge learners suffered from them compared to worked example instructions (Kalyuga et al., 2001; Tuovinen & Sweller, 1999). Without sufficient schemata, learners experience high extraneous cognitive load in unguided instructions because it may be difficult for these students to generate suitable aspects of the area to explore. However, with sufficient schemata learners can guide their exploration and make decisions on their own. This processing is related with lower extraneous cognitive load and lower intrinsic cognitive load. Kalyuga et al. (2001) showed that inexperienced learners rated the difficulty of worked examples lower than the difficulty of the exploratory instruction, whereas this difference disappeared when these learners were more advanced. In an experiment about the concurrent or sequential use of spreadsheets to assist learning in mathematics similar results were found (Clarke et al., 2005). Whether the difficulty ratings measured intrinsic or extraneous cognitive load or both types is unclear. Nevertheless, advanced learners were less loaded by the exploratory instruction than novice learners.

The assumption of a reduced intrinsic cognitive load because of sufficiently developed schemata in high-knowledge learners was also made by Tuovinen and Sweller (1999). In contrast to perceived difficulty they asked learners to rate their mental effort. The intrinsic cognitive load explanation was indirectly supported by the result that low-knowledge learners rated the mental effort in the exploratory instruction higher than in the worked examples instruction, whereas high-knowledge learners had rather low mental effort ratings in both instructional conditions (Tuovinen & Sweller, 1999). If mental effort ratings are a measure of overall cognitive load, the finding shows at least that high-knowledge learners are generally less loaded than low-knowledge learners and therefore support the intrinsic cognitive load explanation, even though one does not know whether the difference is really based on intrinsic cognitive load only or on reduced intrinsic and extraneous cognitive load.

6.2 Quantitative Summary

In contrast to Kalyuga's (2007) review whose main goals were on the one hand to demonstrate the quantitative evidence corroborating the expertise reversal effect (overview of effect size differences) and on the other hand to push his extraneous cognitive load explanation, this review aimed at a more thorough look on the explanations suggested by the authors of the analyzed papers and the empirical evidence in favor of their explanations. The analysis yielded several findings.

First, three explanations were found in the reviewed papers. Although only the extraneous cognitive load and the germane cognitive load explanation was expected according to the most prominent argumentations of CLT, CTML, or CIM studies, there were also some papers that argued with an intrinsic cognitive load explanation. In 25 studies (49%) researchers explained the expertise reversal effect by referring to an increased germane cognitive load in high-knowledge learners with unguided instructions (germane cognitive load explanation, but cf. Schnotz & Rasch, 2005). In 20 studies (39%) researchers argued that high-knowledge learners with well-guided instructions suffer from extraneous cognitive load. In five studies (10%) researchers argued that the reduction of intrinsic cognitive load because of (schematized) prior knowledge is enough to make more knowledgeable learners benefit from such instructions. One study (2%) on hypertext did not argue with relation to any theoretical framework (Shin et al., 1994), and thus, is not included in the further quantitative analyses. Figure 18 depicts an overview of the quantitative summary.

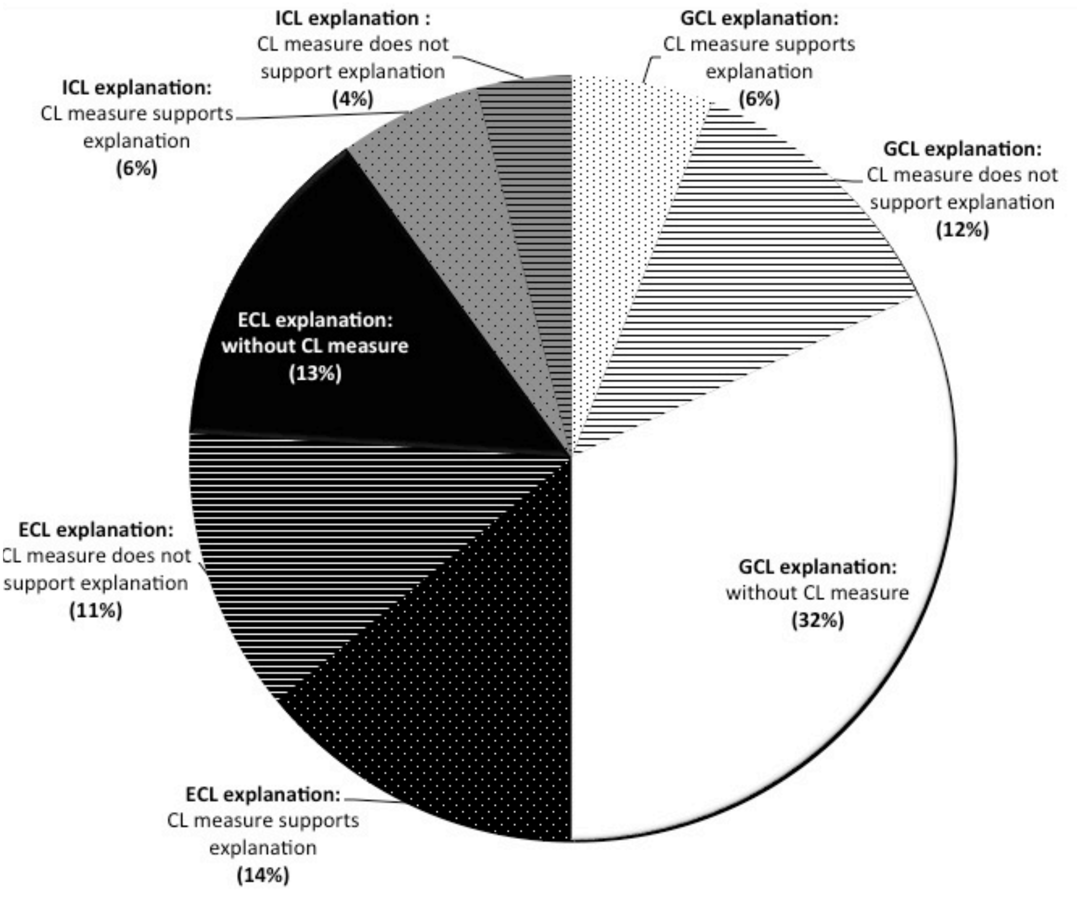


Figure 18. Quantitative summary of 50 studies concerning three cognitive load explanations of the expertise reversal effect

Second, 27 studies (54%) out of the remaining 50 studies measured cognitive load by at least one measure. Four types of cognitive load measures were used. Cognitive load was measured in 13 studies (48%) by means of *studying times*, in 12 studies (44%) by means of *perceived difficulty ratings*, and in seven studies (26%) by *mental effort rating*s. Other cognitive load measures like secondary task performance were not used. Moreover, only four studies (15%) used *behavioral activity measures* additionally to investigate more deeply how learners processed the materials.

Third, although cognitive load is thought to mediate the expertise reversal effect, studies arguing for a *germane cognitive load explanation* did often not measure cognitive load. Only nine (36%) of these 25 studies measured cognitive load at all and only six of these studies (24%) used a cognitive load measure that might be interpreted as germane cognitive load – namely studying times (but cf. Chapter 3.2). Three of the 25 studies (12%) can be said to corroborate the germane cognitive load explanation. These three studies make up 50 % of the empirical evidence provided to evaluate the germane cognitive load explanation. In comparison, thirteen studies (65%) of 20 arguing for the *extraneous cognitive load explanation* measured cognitive load. Ten of these studies (50%) used a cognitive load measure that might be interpreted as extraneous cognitive load – namely *studying times* again (see above) and *perceived difficulty ratings*. Whereas two studies (10%) using studying times supported the extraneous cognitive load explanation, three studies (15%) did not support this explanation. And whereas five studies (25%) using perceived difficulty ratings supported the extraneous cognitive load explanation, only two (10%) did not support it. In sum, 55% (seven studies) of the empirical evidence provided to test the extraneous cognitive load explanation supports it. Five studies argued for *the intrinsic cognitive load explanation* and all of them measured cognitive load. Except for one study, the others (80%) used cognitive load measures that are interpretable as intrinsic cognitive load and/or extraneous cognitive load – namely *studying times* and *perceived difficulty ratings*. About 57% of the empirical evidence supported the intrinsic cognitive load explanation. To sum up, each cognitive load explanation was supported by about (at least) 50% of the empirical evidence provided. However, one should keep in mind that 50% mean only 3 studies in the case of the germane cognitive load explanation. Interestingly, none of the 50 studies reviewed, even when they were conducted within the framework of CLT, measured or tried to measure all three cognitive load types separately. Thus, none of these studies provided clear evidence that only the germane or the extraneous or the intrinsic cognitive load explanation holds true. Hence, the quantitative evidence concerning the explanation of the expertise reversal effect is so far inconclusive and shows that research does not provide a clear picture of the mechanism underlying the expertise reversal effect.

Finally, each explanation assumes that prior knowledge moderates the effect of the instruction on cognitive load, and that cognitive load mediates this moderation effect on learning outcomes. Despite this assumption (no matter which type of cognitive processing is assumed to mediate the moderation effect) none of the studies tested this assumption by means of statistical mediated moderation analyses. Studies which used an adequate cognitive load type measure (e.g., difficulty ratings in studies favoring the extraneous cognitive load explanation) could have used this statistical analysis. Other studies which used rather inappropriate measures (to distinguish between cognitive load types) like mental effort as overall cognitive load, however, could not use it, because the results would not have been interpretable.

6.3 Conclusion and Research Outlook

This review shows that the cognitive load mechanism underlying the expertise reversal effect has not been fully understood so far. This result is similar to the finding of the review concerning the split-attention effect (see Chapter 4). As was the case with the split-attention effect, there also exist competing cognitive load explanations of the expertise reversal effect in literature. Whereas some researchers argue for a germane cognitive load explanation, other researchers argue for an extraneous cognitive load explanation, whereas a few also argue for an intrinsic cognitive load explanation. Despite these competing explanations, however, the empirical evidence of the underlying mechanisms is very limited. Although the expertise reversal effect is thought to be a rather general phenomenon that applies for many specific instructional design characteristics that can be distinguished in providing a more and less guided instruction, and thus, 51 studies were reviewed with regard to their explanations of the expertise reversal effect, it is not yet possible to explain how the level of spatial contiguity between text and picture influences learners' pattern of cognitive load types. Although a few studies on spatial contiguity with more and less knowledgeable learners provided first evidence for an expertise reversal effect, it cannot be concluded so far whether more knowledgeable learners with an integrated text-picture format cannot benefit or even suffer from it due to increased extraneous or due to inhibited germane cognitive processing. Neither can be decided so far, whether more knowledgeable learners with a separated text-picture format benefit from it due to reduced extraneous and intrinsic or due to increased germane cognitive processing.

Although the competing explanations seem to be contradictory at first sight, the possibility exists according to the triarchic model of cognitive load that both mechanisms may apply at the same time (cf. Experiment 1 of this dissertation thesis). Concerning an expertise reversal effect on spatial contiguity between text and picture, it might be possible that high-knowledge learners with an integrated format might not invest germane cognitive load and *additionally* suffer from extraneous cognitive load. On the other hand, high-knowledge learners with separated formats might be freed from extraneous cognitive load *and* invest germane cognitive load. The finding that some studies of the review did not demonstrate any differences in mental effort ratings (measure of overall cognitive load) between high-knowledge learners might be explained by this possible pattern of cognitive load types. Moreover, a few studies concentrating on germane cognitive load found that studying times of high-knowledge learners with unguided instructions sometimes increased. However, studying times are rather difficult to interpret. Moreover, there were also studies which showed the opposite direction. Other studies showed that difficulty ratings of high-high knowledge learners with unguided instructions decrease. This is a finding in favor of the extraneous cognitive load explanation but not necessarily one against the germane cognitive load explanation. Because no study has ever measured both germane as well as extraneous cognitive load separately, there is no empirical evidence whether both explanations hold true or only one of them.

The review also showed that in addition to both aforementioned main explanations, a third explanation suggests that the expertise reversal effect might just be explained by the reduced amount of intrinsic cognitive load in high-knowledge learners that frees working memory resources and thereby enables high-knowledge learners to handle unguided instructions that overload low-knowledge learners. Interestingly, the reduced amount of intrinsic cognitive load in high-knowledge learners was almost not mentioned by researchers favoring either the extraneous or the germane cognitive load explanation. Seufert et al. (2007) are a scarce exception in this respect. In contrast to Seufert et al. (2007) who favored the germane cognitive load explanation, however, the intrinsic cognitive load explanation does not assume that freed working memory resources are used for germane cognitive load. Rather, learners have more resources to not suffer from overload by previously too high intrinsic cognitive load or by previously too high extraneous cognitive load with unguided instruction. Although according to CLT intrinsic cognitive load should be lower in high-knowledge learners in general, this cognitive load type was neither measured in experiments favoring the other two cognitive load explanations. Hence, in testing the mechanism underlying the expertise reversal effect concerning spatial contiguity the measurement of intrinsic cognitive load is also an important issue.

Moreover, another finding of the review was that only very few studies tried to measure how learners actually processed the materials. Oksa et al. (2010) collected verbal reports afterwards, whereas Amadieu and colleagues collected behavioral processing data already during learning by logfiles and learners' viewing behavior (Amadieu, Tricot et al., 2009; Amadieu, Van Gog et al., 2009; see also Shin et al., 1994). However, behavioral processing data like viewing behavior that is measured online during learning seem to provide more insights into more and less knowledgeable learners' cognitive processing of integrated and separated formats. As stated in Chapter 3 on the cognitive load measurement, behavioral activities can be a possibility to measure cognitive load, if there are assumptions concerning the cognitive processing and its associated cognitive load type. CTML and CLT researchers have different assumptions about how learners process integrated and separated formats. Moreover, if the competing cognitive load explanations can be related to competing assumptions about how differently knowledgeable learners read integrated and separated formats, viewing behavior seems to be a suitable measure to help to disambiguate the competing explanations.

Summarizing, to test whether the expertise reversal effect is mediated by germane or/and extraneous or intrinsic cognitive load, an experiment is needed that should fulfill two necessary criteria. First, the three cognitive load types should be measured individually to test the cognitive load type pattern consisting of intrinsic, extraneous, and germane cognitive load. If an overall cognitive load measure is used and instructional conditions do not differ, no definite conclusions can be drawn concerning the cognitive load type explanation – a problem of many existing research results. As shown in Experiment 1 of this thesis, a differentiating measurement of the three load types seems possible by means of subjective rating scales. Thus, it seems rational to use these rating scales to also investigate the expertise reversal effect. Second, the statistical approach should use mediated moderation analyses to test whether the influence of prior knowledge on the instructional format effect is mediated by germane, extraneous, or intrinsic cognitive load. Figure 19 depicts the path model that has to be tested to answer the question which cognitive load type(s) mediate(s) the expertise reversal effect.

Moreover, as more and more instructional researchers (see Van Gog & Scheiter, 2009) expect a great potential in the eye tracking methodology, it seems worthwhile to apply this method in investigating the expertise reversal effect by testing whether and how high-knowledge learners differ from low-knowledge learners during learning with integrated (well-guided) or separated (less guided) formats. Tracking and analyzing learners' viewing behavior helps investigating different processing assumptions

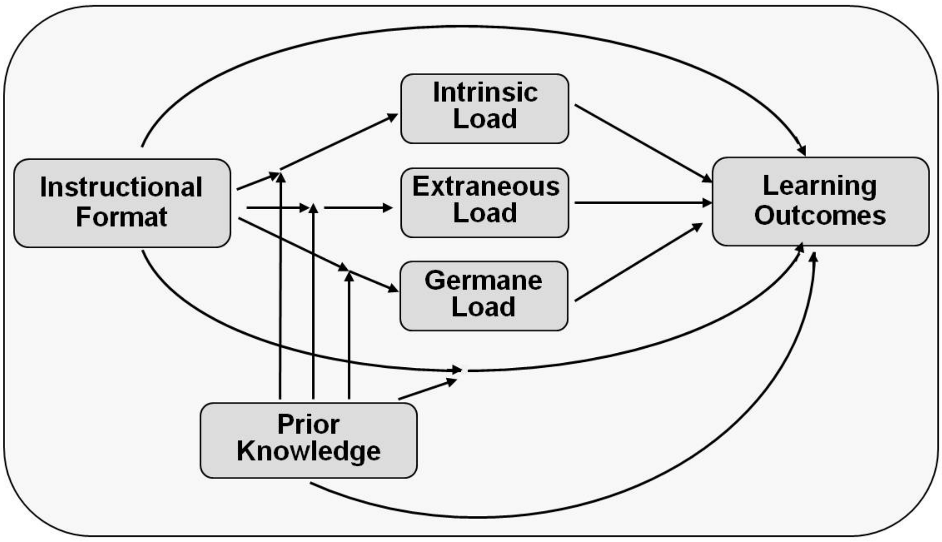


Figure 19. Path model to test the cognitive load explanations of the expertise reversal effect

made by CTML researchers and CLT researchers. With respect to the germane cognitive load explanation, Mayer and Gallini (1990) argued that all learners (no matter which level of prior knowledge) would concentrate on the text during learning with separated formats but that only high-knowledge learners would be able to handle such processing by using prior knowledge during applying imagery strategies (germane cognitive load explanation), whereas with respect to the extraneous cognitive load explanation, Kalyuga et al. (1998) argued that high-knowledge learners would concentrate on the illustration, if they would not be distracted by integrated text (extraneous cognitive load explanation). Concerning low-knowledge learners Erhel and Jamet (2006) assumed that these learners make many switches between text and picture. This assumption is in contrast to the aforementioned assumption by Mayer and Gallini (1990). These competing assumptions on learners' viewing behavior have not yet been investigated so far, but they can be investigated by means of the eye tracking methodology. Eye tracking seems even suited to assess processes that are not necessarily accessible by means of introspection (Van Gog et al., 2009). By using the method of triangulation, eye tracking data can be related with further data like subjective ratings and learning outcomes to investigate the meaning of different measures of viewing behavior (e.g., fixation duration, dwell time, switches) with regard to the three cognitive load types (cf. Ozcelik et al., 2009). Information about these relations should help to disambiguate whether the germane and/or the extraneous cognitive load explanation is more suited to explain an expertise reversal effect with regard to spatial contiguity.

7 Experiment 2: What Explains the Expertise Reversal Effect?

To investigate whether an expertise reversal effect can be found for the split-attention effect shown for Experiment 1, and if so, how cognitive load mediates such a reversed effect, Experiment 2 was conducted. This experiment was the first one that tested three alternative explanations: (1) the germane cognitive load explanation suggested by researchers of different instructional fields (e.g., multimedia: Mayer, 2003; text comprehension: McNamara & Kintsch, 1996), (2) the extraneous cognitive load explanation suggested first by researchers favoring the CLT, and (3) the intrinsic cognitive load explanation that is also based on the CLT but not favored by most CLT researchers. To test these explanations it was necessary to measure the three types of cognitive load separately. The three cognitive load scales developed in Experiment 1 were also used to measure the three cognitive load types in Experiment 2. These measures were used to investigate the assumptions of all explanations that cognitive load mediates the hypothesized interaction between spatial contiguity and prior knowledge with regard to learning outcomes. With regard to different levels of prior knowledge non-medical students were chosen to represent learners with low prior knowledge and medical students were chosen to represent learners with higher prior knowledge. Furthermore, this experiment applied statistical mediated moderation analyses to test the different explanations with regard to the mediating role of the cognitive load types, thereby applying the appropriate statistical method needed for testing the assumed mediated moderation effect. Moreover, the eye tracking methodology was used to measure participants' viewing behavior during learning. It was tested whether these behavioral process data would support the different assumptions of the germane and extraneous cognitive load explanation with regard to learners' viewing behavior. To improve the cognitive interpretation of the eye tracking data, it was further explored how eye tracking measures were related to cognitive load type ratings and learning outcomes. This chapter will first present the hypotheses based on the three alternative explanations, followed by a description of the design and experimental materials used in Experiment 2. Finally, the results of the experiment will be presented and discussed.

7.1 Hypotheses

The hypotheses tested by the following experiment refer to learning outcomes, cognitive load types, and learners' viewing behavior and will be outlined successively.

7.1.1 Hypotheses for Learning Outcomes

With regard to learning outcomes all three cognitive load explanations predict an expertise reversal effect, that is, an interaction effect with regard to learning outcomes, especially when test items ask for complex information or demand inferences but not when test items ask for simple vocabularies. More specifically, it is assumed that low-knowledge learners suffer from a separated text-graphic format and benefit from an integrated text-graphic format *(**H 2.1a**)*, whereas high-knowledge learners should learn equally well with both formats *(*ordinal interaction: ***H 2.1b****)* or might even benefit from a separated format *(*disordinal interaction: ***H 2.1c****)*.

7.1.2 Hypotheses for Cognitive Load Types

The three cognitive load explanations differ with respect to their assumptions with respect to the three cognitive load types. Table 22 summarizes the hypotheses of both the germane and the extraneous cognitive load assumptions about the expertise reversal effect and its cognitive load mechanism. Moreover, the intrinsic cognitive load explanation is stated in the according hypotheses. In the following the different hypotheses for all three cognitive load types will be outlined.

Table 22

Overview of the difference-hypotheses concerning the expertise reversal effect

	Germane CL Explanation	**Extraneous CL Explanation**
Learning Outcomes	***ICL-Expl.** $IF_{LP} > SF_{LP}$; $IF_{HP} = / < SF_{HP}$	***ICL-Expl.** $IF_{LP} > SF_{LP}$; $IF_{HP} = / < SF_{HP}$
ICL	***ICL-Expl.** $IF_{LP}, SF_{LP} > IF_{HP}, SF_{HP}$	***ICL-Expl.** $IF_{LP}, SF_{LP} > IF_{HP}, SF_{H}$
ECL	***ICL-Expl.** $IF_{LP} < SF_{LP}$; $IF_{HP} < SF_{HP}$	$IF_{LP} < SF_{LP}$; $IF_{HP} = / > SF_{HP}$
GCL	$IF_{LP} > SF_{LP}$; $IF_{HP} = / < SF_{HP}$	***ICL-Expl.** $IF_{LP} = SF_{LP}$; $IF_{HP} = SF_{HP}$

Note. *ICL-Expl. = intrinsic cognitive load explanation,
IF = integrated format, SF = separated format,
LP = low prior knowledge, HP = high prior knowledge,
ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

Intrinsic cognitive load hypotheses. According to CLT high-knowledge learners should experience lower intrinsic cognitive load than low-knowledge learners, because high-knowledge learners already possess schemata that facilitate the processing of novel information. Accordingly, *all explanations* predict a main effect of prior knowledge on intrinsic cognitive load. ***(H 2.2.1a)***. Furthermore, there should be no differences between the instructional format groups with regard to this measure because experimental conditions only differ with regard to spatial contiguity but not with regard to the complexity of the content ***(H 2.2.1b)***.

Extraneous cognitive load hypotheses. The *germane cognitive load explanation* predicts that integrated formats are perceived as easier than separated formats no matter whether learners have low or high prior knowledge, because the perceived instructional demand characteristics of an instructional format should not change due to prior knowledge ***(H 2.2.2a)***. The same assumption is made by the *intrinsic cognitive load explanation*. In contrast, the *extraneous cognitive load explanation* assumes that high-knowledge learners with integrated format suffer from higher extraneous cognitive load due to redundancy as compared to high-knowledge learners with separated format, whereas low-knowledge learners suffer from higher extraneous cognitive load when they learn with separated format than with integrated format due to searching for corresponding information. Depending on the type of expertise reversal effect (ordinal or disordinal interaction), the interaction effect with regard to extraneous cognitive load can be either ordinal or disordinal. In the ordinal case, low-knowledge learners with separated format should always yield higher ratings than low-knowledge learners with integrated format but high-knowledge learners with different formats do not necessarily differ (***H 2.2.2b***). In the disordinal case, high-knowledge learners with integrated format should yield higher extraneous cognitive load than high-knowledge learners with separated format (***H 2.2.2c***). Furthermore, extraneous cognitive load should mediate the interaction between instructional format and prior knowledge with regard to learning outcomes ***(H 2.2.2d)***. This assumption on mediation is not shared by the germane or intrinsic cognitive load explanation.

Germane cognitive load hypotheses. The *germane cognitive load explanation* predicts that high-knowledge learners with integrated format should have lower germane cognitive load than high-knowledge learners with separated format, whereas low-knowledge learners with integrated format should have higher germane cognitive load than low-knowledge learners with separated format. This pattern corresponds to a disordinal interaction effect ***(H 2.2.3a)***. However, according to O'Reilly and McNamara's (2007) suggestion that high-knowledge learners with good strategies can also process well guided instructions more actively, high-knowledge learners with integrated format and good processing strategies need not necessarily differ from high-knowledge learners with separated format. This pattern corresponds to an ordinal interaction effect ***(H 2.2.3b)***. Furthermore, it is assumed that germane cognitive load mediates the interaction between instructional format and prior knowledge with regard to learning outcomes ***(H 2.2.3c)***. In contrast, the *extraneous and the intrinsic cognitive load explanations* predict no differences between high- or low-knowledge learners or between learners with integrated and separated format ***(H 2.2.3d)***.

7.1.3 Hypotheses for Viewing Behavior

Hypotheses with regard to learners' viewing behavior during learning with integrated and separated formats can only be derived from the germane (e.g., Mayer & Gallini, 1990) and the extraneous cognitive load explanation (e.g., Erhel & Jamet, 2006). According to the literature on eye tracking studies (see Chapter 3), the assumptions of learners' perceptual processing of multimedia material can be described with respect to (1) the average fixation duration (Ozkelic et al., 2009), (2) the dwell time of learners' visual attention that is allocated on textual and pictorial information (Folker et al., 2005), and (3) switching between textual and pictorial information (Holsanova et al, 2009). Table 23 summarizes the hypotheses on learners' viewing behavior of both the germane and the extraneous cognitive load explanation.

Table 23

Overview of the hypotheses concerning viewing behavior and the expertise reversal effect

	Germane CL Assumption	Extraneous CL Assumption
Fixation durations	IF_{LP}, SF_{LP} > IF_{HP}, SF_{HP}	IF_{LP}, SF_{LP} > IF_{HP}, SF_{HP}
Dwell time on text	IF_{LP} < SF_{LP}; IF_{HP} < SF_{HP}	IF_{LP} > SF_{LP}; IF_{HP} > SF_{HP}
Dwell time on graphic	IF_{LP} > SF_{LP}; IF_{HP} > SF_{HP}	IF_{LP} < SF_{LP}; SF_{LP}, IF_{HP} < SF_{HP}
Switches within text	IF_{LP} < SF_{LP}; SF_{LP}, IF_{HP} < SF_{HP}	IF_{LP} > SF_{LP}; SF_{LP}, IF_{HP} > SF_{HP}
Switches within graphic	IF_{LP} > SF_{LP}; SF_{LP}, IF_{HP} > SF_{HP}	IF_{LP} < SF_{LP}; SF_{LP}, IF_{HP} < SF_{HP}
Corresponding switches between text and graphic	IF_{LP} > SF_{LP}; SF_{LP}, IF_{HP} > SF_{HP}	IF_{LP} > SF_{LP}; IF_{HP} > SF_{HP} < SF_{LP}
Non-corresp. switches between text and graphic	IF_{LP} < SF_{LP}; IF_{HP} =/< SF_{HP} < SF_{LP}	IF_{LP} < SF_{LP}; IF_{HP} < SF_{HP} < SF_{LP}

Note. IF = integrated format, SF = separated format, LP = low prior knowledge, HP = high prior knowledge, ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

Average fixation durations. The average fixation duration is often interpreted as the temporal amount of processing engagement. Therefore, it is generally assumed that high-knowledge learners have shorter average fixation durations than low-knowledge learners, because of the reduced intrinsic cognitive load they should experience ***(H 2.3.1a)***. This prediction is in line with the *germane, extraneous, and intrinsic cognitive load explanation*. Whether average fixation duration might also represent germane or extraneous cognitive load is an issue for explanatory analyses. Former eye tracking research on learning suggests that longer fixation durations may indicate higher germane cognitive load. According to this assumption, an interaction effect can be derived for fixation durations from the *germane cognitive load explanation*. Low-knowledge learners with the integrated format should have higher fixation durations than low-knowledge learners with a separated format, whereas high-knowledge learners' fixation durations should not differ or show the reverse result pattern ***(H 2.3.1b)***.

Dwell time on text and graphic. How long learners process one information representation (text or graphic) is described by the summed fixation times of the respective areas of interest (AOIs). AOIs either contain text or graphical information. Dwell times are measured by summing up the duration times of all fixations within the respective AOIs.

Dwell time on text. According to the *germane cognitive load explanation* (Mayer & Galini, 1990), learners with separated format concentrate on textual information and do not actively process and integrate graphical information. Hence, the germane cognitive load explanation predicts that learners with a separated format should process textual information longer than learners with an integrated format. ***(H 2.3.2a)***.

Dwell time on graphic. According to the *extraneous cognitive load assumption* it is suggested that learners with separated format process the graphic longer than learners with integrated format. Concerning low-knowledge learners with separated format, it is assumed that they have to search for elements in the graphic corresponding to the verbal information and thus process the graphic longer than learners with integrated format. In this case, long dwell times on the graphic might represent an aspect of visual search. Concerning high-knowledge learners with separated format, it is assumed that they might process the graphic longer than all other learners, because they prefer to and can ignore textual information which is redundant. In that case, long dwell times on the graphic might represent an aspect of the ability/strategy to learn primarily with the graphic and thus not being overloaded by redundant text information. Summarily, an interaction effect is assumed ***(H 2.3.2b)***.

Switching behavior. Learners' switching behavior during learning with multimedia materials can be categorized into switches within one type of representation, that is, switches between different textual information units or switches between different graphical information units, and switches between the two types of representations. The switches between two representation types (text and graphic) can be further divided into switches between non-corresponding textual and graphical units and switches between corresponding textual and graphical units (see Chapter 3).

Switches within one representation. Because the *germane cognitive load explanation* suggests that high-knowledge learners learning with separated format apply elaborated reading strategies and process the text more deeply, it is assumed that high-knowledge learners show a stronger switching behavior within the text than learners with less prior knowledge or when learning with integrated format. This processing behavior should result in an interaction effect with regard to switches between different textual information units ***(H 2.3.3a)***. In contrast, according to the extraneous *cognitive load explanation* it is assumed that high-knowledge learners with the separated format switch the least between different textual units but the most between different graphical units, because they prefer to ignore the text. Hence, an interaction effect with regard to switches within the graphic and within the text is assumed ***(H 2.3.3b)***.

Switches between representations. According to the *germane cognitive load explanation* learners with the separated format do not actively integrate text and graphic, and thus, they should switch less often between text and graphic than learners with integrated format. This should hold true especially for learners with high prior knowledge, because they can apply specific imagery strategies and thus do not rely on information in the graphic. Hence, they do not need to switch to understand the text. However, if high-knowledge learners with separated format switch between text and graphic, they switch more often between *corresponding* information units than low-knowledge learners with separated format because their prior knowledge helps them to decide where to switch on the graphic ***(H 2.3.3c)***. The *extraneous cognitive load explanation* predicts that low-knowledge learners with separated format switch more often between text and graphic than low-knowledge learners with integrated format and than high-knowledge learners with separated format because of their need for mental integration. This should hold true especially for switches between *non-corresponding* information units which might represent an aspect of visual search ***(H 2.3.3d)***.

Exploratory analyses. To explore whether and which of the eye tracking measures represent which cognitive load types, the correlations between objective eye tracking measures and subjective cognitive load ratings, as well as between the eye tracking measures and learning outcomes were analyzed.

7.1.4 Hypotheses for Control Variables

Several control variables were obtained to increase the internal

Cognitive variables. Participants' cognitive ***learning prerequisites*** were measured to ensure the equality among groups. The ***domain*** and ***topic prior knowledge*** were registered for manipulation check concerning high prior knowledge (non-medical students were assumed to have lower prior knowledge, whereas medical students were assumed to have higher prior knowledge). Furthermore, these variables were used to ensure that prior knowledge was equal among low-knowledge participants and among high-knowledge participants, respectively.

Motivational variables. In addition to the cognitive variables, participants' ***perceived task demands*** and ***interest*** were explored to investigate whether motivational aspects not considered so far by the cognitive load explanations play a role for expertise reversal effects. In order to ensure the internal validity of the experiment it was tested whether the perceived task demands were equal among groups. Moreover, it was explored whether learners with separated format reported the same level of interest than learners with integrated format after learning. Following Tobias (1994) who showed that interest in a topic increases the more knowledge learners have, it was tested whether prior knowledge influenced interest. It was also explored whether one of the cognitive load types measured was related to these motivational aspects to ensure internal validity.

Studying times. Because the version of the learning environment was self-paced for high-knowledge learners, high-knowledge participants' studying times were measured to explore how long they studied and to control for learning times if necessary. Because it was assumed that learners with high prior knowledge might not need the pre-defined time set for low-knowledge learners to study the materials, they had the options to either study maximally as long as low-knowledge learners or to study shorter. This was seen as necessary because otherwise the data of high-knowledge learners' viewing behavior might not have been interpretable, if they had to process information longer than they needed. In contrast, the learning time for low-knowledge learners was predefined (see procedure in Experiment 1 and 2), in order to make sure that they learn but do not exceed a realistic time for experiments. Because the learning content was rather difficult, especially engaged or interested learners would have otherwise studied much longer than others.

7.2 Methods

The following sections describe the participants, materials, procedure, apparatus, and data analysis of Experiment 2.

7.2.1 Participants and Design

Sixty university students, 39 females and 21 males, with an average age of 22.79 (*SD* = 2.73) years participated in the study for either payment or course credit. All participants were native German speakers with normal or corrected-to-normal vision. Three participants had to be excluded because of several problems. One participant was too small for the EyeLink's chin rest to sit comfortably during learning. After the learning phase she was complaining that the discomfort made it almost impossible for her to learn. Thus, her learning outcomes and subjective ratings were probably intrigued by this problem were and therefore excluded from the analyses. The second participant was a medical student with almost no specific prior knowledge (less than 20 %). Because prior knowledge was not used as a continuous but as a dichotomous variable (low vs. high) in this experiment, the data of this medical student were not representative for the high prior knowledge group. During testing the third excluded participant the eye tracking system failed to start the program because of technical problems that could not be fixed during that session. From the remaining 57 participants 29 studied subjects like psychology, politics, or history. These students served as low prior knowledge learners and were randomly assigned to either the separated or the integrated format condition[9]. Twenty-eight participants were medical students. The medical students served as high prior knowledge learners and were randomly assigned to either the separated or the integrated format condition. This resulted in a 2 x 2 design with prior knowledge (low vs. high) and instructional format (separated vs. integrated) as independent variables.

According to Kalyuga's (2007) review a full reversal (a disordinal interaction) was usually not obtained in strictly controlled longitudinal studies in which the same novice learners were gradually trained to eventually become experts in specific task domains,

[9] These two experimental conditions were identical to the two conditions without secondary task of the first experiment.

but mostly in cross-sectional studies. Moreover, according to Cohen's (1988) power table 55 cases are needed to detect a moderate to large interaction effect (f^2 = .15) for a power of 1 – β = .80, α = .05, and assumed without measurement error in the predictors seems to be justified because non-medical and medical students were assigned to the low-knowledge and high-knowledge groups, respectively (measurement errors in the predictors usually reduce power). Thus, by comparing non-medical students with medical students of higher semesters in a cross-sectional design the remaining 57 participants of this study should be enough to detect an expertise reversal effect, if there is a moderate to large effect in the assumed population.

7.2.2 Materials

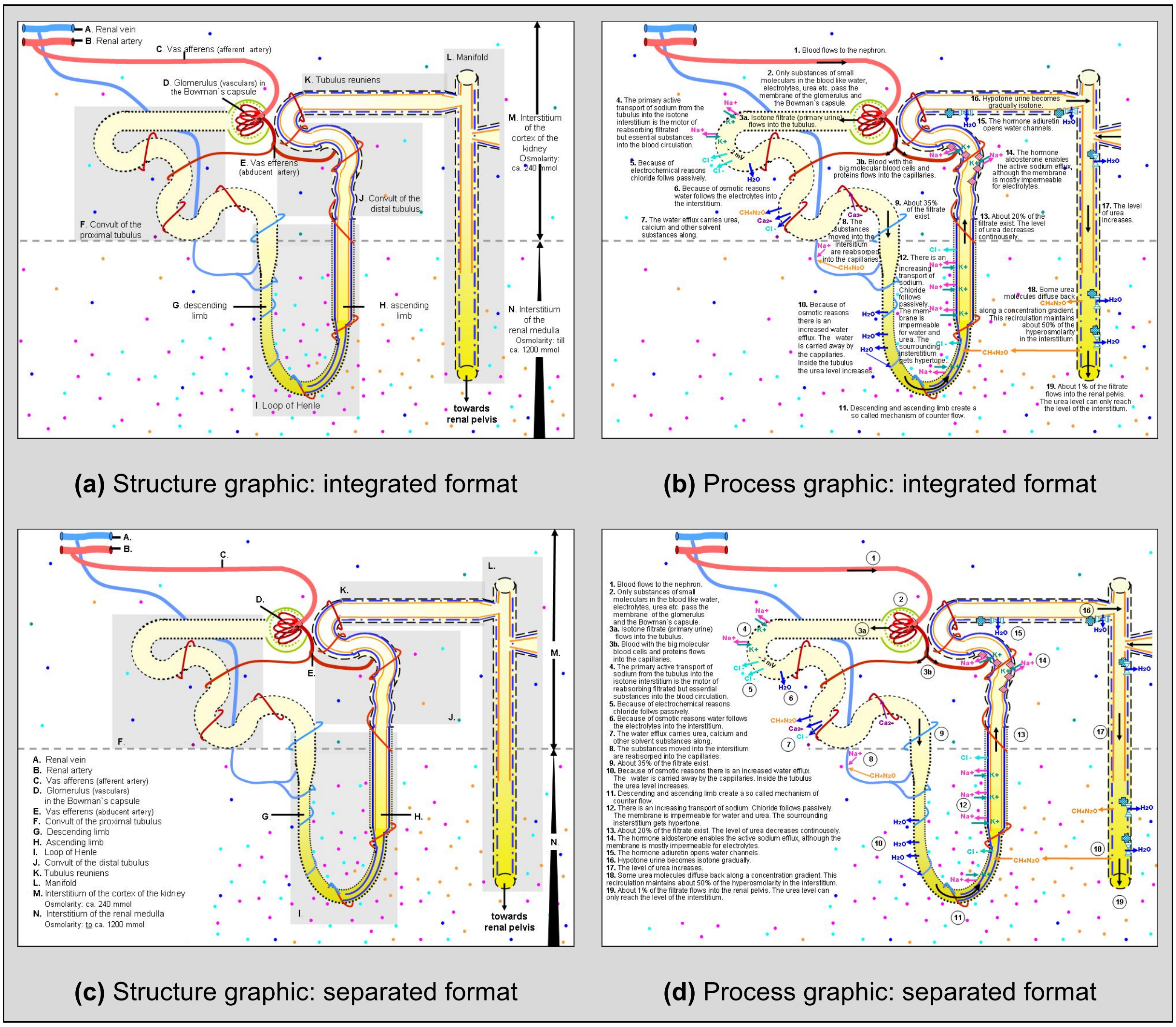
The learning materials and the materials to measure participants' prior knowledge as well as learning outcomes were the same as those used in Experiment 1. Furthermore, because the cognitive load items were successfully used during Experiment 1, they were again used in Experiment 2. Therefore, the materials are only described briefly. For a detailed description of the materials see the methods section in Chapter 5.

7.2.2.1 Independent variables

Two variables were manipulated in Experiment 2. First, participants with different prior knowledge were chosen by asking medical and non-medical students to take part in the experiment. Second, the instructional format of the learning materials differed with respect to spatial contiguity.

Prior knowledge. To validate the assignment of medical students as learners with high prior knowledge compared to non-medical students as learners with low prior knowledge, participants' domain prior knowledge in physiology and their specific topic knowledge were measured. The physiology test consisted of 18 sentences about physiological issues (e.g., "The so-called sodium-potassium pump is an enzyme which converts ADH in ADP and phosphate by energy consumption."). Participants had to state whether these sentences were either right or wrong. To test for topic knowledge the four knowledge tests used to measure learning outcomes (see below) were administered in a pre-post test design.

Instructional format. The learning materials consisted of a computerized learning environment about the physiological functioning of the nephron, the functional unit of the kidney and were already used in Experiment 1 (see Figure 20). The environment consisted of a short introduction into the topic and two complex instructional graphics with accompanying text. The introduction was the same for all participants and was about general functions of the kidney. Subsequent to the introduction, the first instructional graphic was presented consisting of a colored graphic of a nephron with verbal information about its *structure* parts (46 words; font: Arial; size: 11). Afterwards, a second instructional graphic was presented consisting of the visualization of the physiological *processes* in the nephron accompanied by verbal explanations (249 words; font: Arial; size: 9). Without knowledge about the structure of a nephron their verbal information was unintelligible in isolation, because the text about the physiological processes lacked specific spatial information about the structural places, where the processes take place. Both instructional graphics and their accompanying text were presented either in separated or integrated format, and thus, differed only with respect to the spatial contiguity between verbal and corresponding graphical information.



(a) Structure graphic: integrated format
(b) Process graphic: integrated format
(c) Structure graphic: separated format
(d) Process graphic: separated format

Figure 20. Illustrations of the learning materials on the functioning of a nephron used in both format conditions

7.2.2.2 Dependent variables

Participants' learning outcomes and cognitive load ratings were the main dependent variables. The materials were identical to those used in Experiment 1 (without secondary task).

Learning outcomes. To measure participants' learning outcomes four computerized knowledge tests (terminology, labeling, complex facts, and transfer) were used. The identical tests were used to validate learners' topic prior knowledge. For further examples see Experiment 1.

(1) The ***terminology*** test consisted of nine multiple-choice items about the structural terms of a nephron, where the answer had to be selected out of four alternatives.

(2) The ***labeling*** test consisted of 12 multiple-choice items. Participants had to chose one out of twelve possible structure terms that matched the high-lighted part in a given graphic that depicted a nephron.

(3) The test about ***complex facts*** consisted of 22 sentences about the physiological processes in a nephron (e.g., "The urea concentration increases in the descending limb of loop of Henle."). Participants had to state whether these sentences were either correct or incorrect.

(4) The ***transfer*** test consisted of 20 sentences about causes and effects in a nephron (e.g. "If proteins are found in the urea test of a patient, a defect in the vas efferens can be assumed."). Again, participants had to state whether these sentences were correct or incorrect.

Cognitive load types. To measure the three types of cognitive load, three subjective rating scales with a labeled six-point Likert-type scale were used ranging from "not at all" (1 point) to "extremely" (6 points).

(1) ***Intrinsic*** cognitive load scale: *"How difficult was the learning content for you?"*

(2) ***Extraneous*** cognitive load scale: *"How difficult was it for you to learn with the material?"*

(3) ***Germane*** cognitive load scale: *"How much did you concentrate during learning?"*

7.2.2.3 Control variables

To control for and explore possible cognitive and motivational influences participants' cognitive processing capacity, domain knowledge in physiology, perceived task demands, and interest were measured.

Cognitive processing capacity. To measure participants' learning prerequisites the processing capacity sub-scale of the BIS-4 intelligence test was used (Jäger et al., 1997).

Domain knowledge in physiology. To measure participants' domain prior knowledge in physiology, they were asked to state whether 18 sentences about physiological processes were either correct or incorrect.

Perceived task demands. To measure participants' perceived task demands which might influence learners' actual investment of mental effort or the subjective ratings of cognitive load, participants were shown the process graphic in the separated format (A) next to the process graphic in the integrated format (B) and were first asked "How hard do you think it is to learn with format A?" and then "How hard do you think it is to learn with format B?". A labeled six-point Likert-type scale was used ranging from "not hard" (1 point) to "extremely hard" (6 points).

Interest. To measure participants' interest, a labeled six-point Likert-type scale was used ranging from "not at all" (1 point) to "extremely" (6 points). The scale asked *"How interesting did you find the learning content?"*.

7.2.3 Procedure

The study consisted of four phases: An initial pre-test phase, a learning phase, a subsequent phase to rate cognitive load scales and a final post-test phase. Participants were run in individual sessions. One session lasted about 2 hours. In the first phase of the experiment participants were asked to answer first the processing capacity scales of the BIS intelligence test, second a general knowledge test about physiology, third four knowledge tests about the nephron (terminology, labeling, complex facts, and transfer), and fourth they were asked to rate how hard they think the topic is to learn with either an integrated or a separated format (perceived task demands). After pre-testing, participants changed seats and their eyes were calibrated with the eye tracker system. After calibration, participants started the computer based learning environment by pressing the keyboard's space bar and were instructed to learn as well as possible. Whereas the presentation time of the learning environment was fully system paced in the low prior knowledge conditions (structure graphic: 180 s; process graphic: 600 s), the presentation time in the high prior knowledge conditions were participant-paced. Participants in these conditions were allowed to go on by pressing the space bar before the system paced time ended, whenever they thought they had learned the content. This difference in learning times was allowed to happen because it was assumed that if high-knowledge learners are forced to process the instruction longer than needed their viewing behavior could not be used as valid measures of learning activities or correlates of cognitive load. In contrast, the topic of the instruction was so complex that some learners with low prior knowledge might have studied much longer than others. After the learning phase, students changed again seats and had to rate the cognitive load scales, then they had to answer the four knowledge tests again (terminology, labeling, complex facts, and transfer). Because the items of the test on complex facts and on transfer had a guessing probability of 50% and knowledge retrieval is related to changes in memory awareness (Conway et al., 1997), participants had to rate their confidence about the correctness of their answers on a five-point Likert-type scale ranging from “guessed” (0 point) to “very sure” (4 points) after each test item.

7.2.4 Apparatus

During the learning phase, participants sat in a distance of about 60 cm from a 21 inch computer monitor with a flicker rate of 100Hz and resolution of 1152 x 864 pixels in a darkened room. While subjects studied the learning materials, their eye movements were recorded every ms from the right eye by a video-based EyeLink 1000 Hz tracker (SR Research) with integrated head support device and gaze accuracy of 0.25° to 0.5°. The calibration was done with a 9 point grid. Each calibration was validated. The calibration was optimized until the EyeLink’s validation measure was < .5 indicating good calibration quality.

7.2.5 Data Analysis

Before the data could be analyzed, raw scores had to be transformed into several variables. These transformations concerned the items of the knowledge tests as well as the eye tracking data and are described in more detail below. Moreover, the statistical analyses used to test the mediated moderation assumption are outlined.

7.2.5.1 Learning outcomes

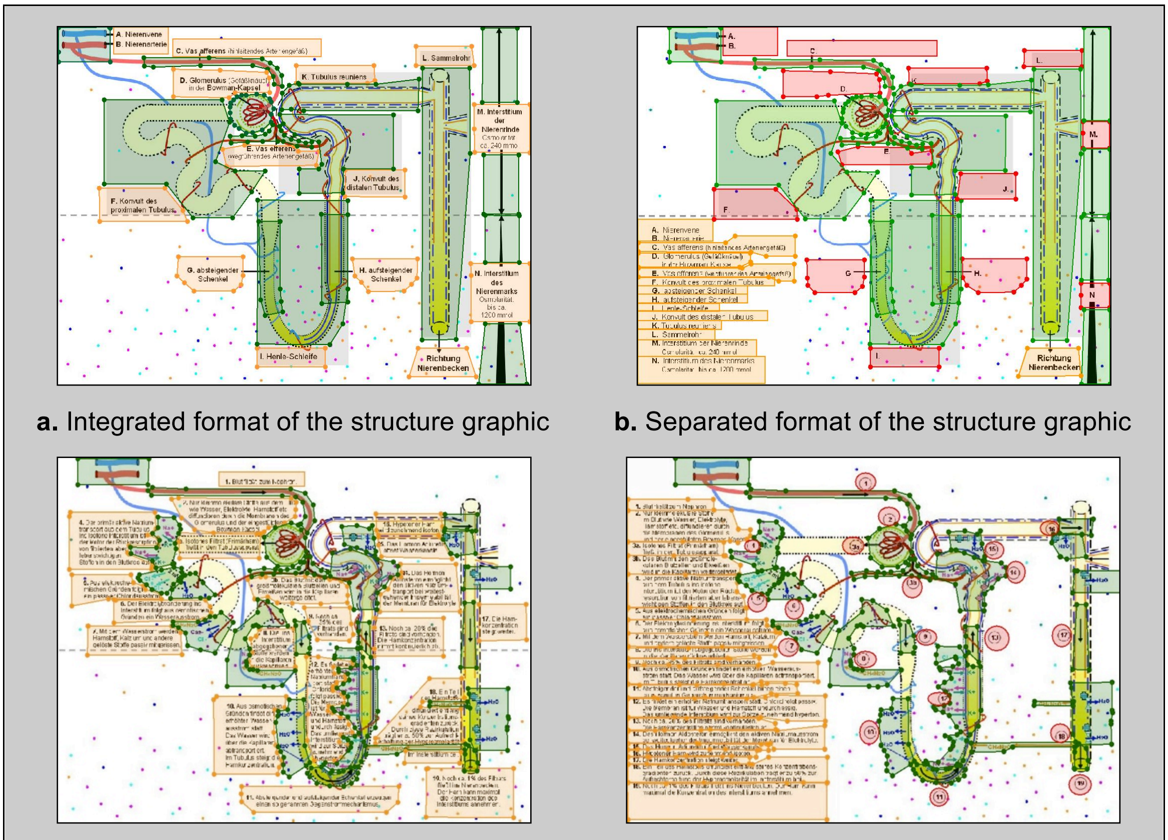
The learning outcome measures were computed according to the rational of Experiment 1. For each correctly answered test question participants were assigned 1 point, whereas 0 points were assigned in case of a wrong answer. The answers to all test questions were weighted with participants' confidence ratings ranging from 0 ("guessed") to 4 ("surely known") concerning the response correctness by multiplying both scores. Only correct answers (1 point) were weighted. If participants gave a wrong answer (0 points) it resulted in 0 points, no matter how sure they had stated to be, thereby ensuring that no negative knowledge outcome was generated. If participants guessed correctly and stated that they had guessed (0 points), their answer was multiplied by 0 thereby resulting in 0 points. The more surly participants stated that they knew a correct answer, the higher the knowledge score of this answer (1-4). By taking participants' knowledge consolidation into account, the nominal items were transformed into metrical ones. Based on the products, the percentage of the maximal score was determined for each participant on each knowledge test. Using the confidence information of participants' knowledge retrieval has several advantages. First, the problem of guessing probability (50% guessing probability in the tests on complex facts and transfer) and rather easy recognition items (test on labeling) is bypassed. Second, the consolidation aspect of knowledge acquisition is taken into account (Conway et al., 1997), that is, answers that participants feel more confident about, are assumed to reflect knowledge that is already better consolidated and retrievable. Third, nominal items are transformed into metrical ones. All three aspects should increase reliability.

7.2.5.2 Eye tracking data

The EyeLink 1000 system used is based on a saccade detection algorithm to determine fixations and saccades. The setting of the saccade sensitivity was set on medium level corresponding to a velocity threshold of 30°/s and an acceleration threshold of $8000°/s^2$. All eye tracking data were extracted by means of the EyeLink DataViewer software. When necessary (e.g., transition matrices), the data were further processed in a spreadsheet program. All eye tracking variables consist of the combined gaze data of the structure as well as the process graphic.

Quality check. First of all, the quality of the eye tracking data was checked by inspecting the noise level reflected in very short fixations (fixation duration < 25 ms), also called jitter, when the saccades between two very short fixations are very fast. A huge amount of such short fixations, caused for example by multiple corneal reflexes, reflects bad quality of fixation counts (Holmqvist et al., 2011.). To check whether the EyeLink algorithm produced jitter-free fixation counts, the merging function of the DataViewer Software was used. Fixations shorter than 25 ms were merged with the nearest fixation. Then the fixation counts of the unmerged data and the fixation counts of the merged fixation data were used to compute the fixation loss due to very short fixations. The overall mean of the noise level was smaller than 1% with a minimum of 0% and a maximum of 4% indicating high quality data with regard to jitter.

Switching behavior. To analyze the gaze recordings in such a manner that switches or transitions between text and graphic units could be analyzed, AOIs were created. For each text unit as well as for each pictorial unit one AOI was created. This resulted in 14 text-AOIs and 14 graphic-AOIs representing the different structures on the structure graphic and in 20 text-AOIs and corresponding 20 graphic-AOIs representing the different physiological processes described on the process graphic (see Figure 21). Because the position and layout of the text differed between the instructional formats, the text-AOIs differed in size according to the layout of the instructional formats. The summed size of all text-AOIs on the structure graphic was 83.95 cm^2 for the integrated and 63.75 cm^2 for the separated format. The summed size of all text-AOIs on the process graphic was 159.55 cm^2 for the integrated and 139.75 cm^2 for the separated format. The size of the graphic-AOIs of both instructional formats had the same size because the graphic was the same in both instructional formats. The summed size of the graphic-AOIs on the structure graphic was 197.75 cm^2 and on the process graphic 87.95 cm^2. Additionally, label-AOIs were created for the capitals on the structure graphic and the numbers on the process graphic of the separated format. Whereas the label-AOIs on the structure graphic corresponded to the text-AOIs of the integrated format, the label-AOIs on the process graphic were circles (size about 1 cm^2) around the numbers next to the graphical units. To use the information of the AOIs, transition matrices were exported from the DataViewer software. Transition matrices are a tabular representation of transitions to and from each defined AOI (Ponsoda, Scott & Findlay, 1995; Roetting, 2001). Using the information from these matrices four types of switches were determined.



a. Integrated format of the structure graphic **b.** Separated format of the structure graphic

c. Integrated format of the process graphic **d.** Separated format of the process graphic

Figure 21. Structure and process graphics with text-AOIs (orange), label-AOIs (red), and graphic-AOIs (green)

Switches within one representation: (1) transitions between different text-AOIs, and (2) transitions between different graphic-AOIs.

Switches between representations: (3) transitions between text- and any other graphic-AOIs (*non-corresponding* switches), and (4) transitions between *corresponding* text- and graphic-AOIs (several graphic-AOIs referring to the same text-AOI were treated as one graphical unit).

Concerning the number of transitions, there are two possibilities in counting the transitions. Either all transitions between two AOIs are counted (simple transition frequencies) or the mere existence of one or more transitions between two AOIs is counted as one and the non-existence of transitions as zero. Thus, instead of counting each switch between two AOIs in a transition matrix, it is only checked whether there is at least one transition between two AOIs. Even if there were more switches between two AOIs, only 1 transition is counted between these AOIs. This measure is called spatial transition density according to Goldberg and Kotval (1999).

There are two reasons why the measure of spatial transition density seems to be more suited in this dissertation than the number of mere frequencies. First, many AOIs (especially the text AOIs in the separated format but also the corresponding text and graphic AOIs in the integrated format) are very close to each other in the materials used. Hence, it is not possible to get high precision in the data concerning participants' frequencies of switches. Second, the mere frequency of learners' switches does not easily provide information about how systematically the participants processed the instructions, because frequencies do not tell whether a high number of switches is caused by many switches between only a few AOIs or by less switches between many AOIs. To get more information about how strategically learners processed instructions with regard to information integration, the measure of spatial transition density seems to be more suited because spatial transition density allows investigating whether participants' visual attention is distributed equally across the AOIs. Therefore, the measure of spatial transition density was computed for the four types of transitions or switches according to Goldberg and Kotval (1999). The counts of mere existence of one or more transition between two AOIs were summed up and divided through the total number of the matrix cells representing all possible transitions between two AOIs. For example, the transition matrix of the structure graphic consists of 14 text-AOIs, in which the existence of at least one transition from each text-AOI to all other text-AOIs is counted. The total number of possible transitions between all different text-AOIs is 182 (formula: 14*14 – 14 (the diagonal representing saccades within the same AOI)). If an eye tracked person would have switched between 10 different AOI pairs of this matrix, the person's spatial transition density would be 10/182 = 0.05 or 5.00%. To account for high-knowledge students' differences in learning times, the density measure was divided by the learning time in minutes resulting in the measure of transition density % / min (cf. Folker et al., 2005). The higher this number is, the more transitions between two different AOIs pairs were conducted per minute.

7.2.5.3 Mediated moderation analyses

To test whether the interaction effect of instructional format and prior knowledge on learning outcomes is conveyed through cognitive load, mediated moderational analyses were used. Mediated moderational analyses attempt to identify whether the effect of an independent variable on a dependent variable depends on a moderator variable and whether this moderation is conveyed through a mediator variable. As mediated moderation is not a standard analysis method used in the cognitive load literature, the statistical rational of this analysis is outlined briefly analogous to the simple mediation analyses described in Chapter 5.2 (for a detailed overview of conditional indirect effects see Muller, Judd & Yzerbyt, 2005; Preacher, Rucker & Hayes, 2007).

To demonstrate mediated moderation, four steps basing on three regression models have to be conducted. The steps are described by referring to the variables used in this study and to model 2 in the systematic about conditional indirect effects of Preacher et al. (2007). *Step 1* is to show an interaction (or so called moderation) effect between instructional format and prior knowledge on learning outcomes (simple dependent variable model). That is, the magnitude of the influence of instructional format on learning outcomes depends on prior knowledge. *Step 2* is to show a moderation effect of instructional format and prior knowledge on cognitive load (mediator variable model). That is, the magnitude of the influence of instructional format on cognitive load depends on prior knowledge. *Step 3* is to show that the assumed mediator (cognitive load) influences the learning outcomes after controlling for instructional format, prior knowledge, and the moderation of instructional format and prior knowledge (mediational dependent variable model). Moreover, the residual moderation effect of instructional format and prior knowledge should be reduced compared to the one in the simple dependent variable model. *Step 4* is to show that this residual moderation effect is significantly reduced in magnitude compared to the moderation effect of instructional format and prior knowledge in the simple dependent variable model (see step 1; Muller et al., 2005).

To test the significance of a mediated moderation effect (step 4) the method of bootstrapping can be used (Preacher et al., 2007). If a moderator variable is dichotomous, the mediated moderation effect for the two categories of the moderator can be specified. When using bootstrapping, the sampling distribution of the mediated moderation effect is estimated non-parametrically, and thus, no specific assumptions about the sampling distribution have to be considered. By using the information from the bootstrap sampling distribution, confidence intervals (CI) for the mediated moderation effect for the specific values of the moderator can be generated. If the CI embraces the value of no effect, that is the difference between the measures compared (relation between the interaction of the independent variables and the dependent variable with and without considering the mediator) is zero, there is no mediated moderation effect. However, the hypothesis that there is no mediated moderation effect or that the reduction in the magnitude of the relation between the interaction of instructional format and prior knowledge on learning outcomes in step 4 is not significant can be rejected, if the CI does not contain zero, thereby indicating significance.

7.3 Results

In the following the results of Experiment 2 are presented. After checking for randomization and assignment to prior-knowledge groups, the learning outcomes were analyzed to probe the expertise reversal effect. Then, the subjective cognitive load ratings were analyzed to test the assumptions on how cognitive load mediates the possible expertise reversal effect. Finally, the measures of viewing behavior were analyzed.

For all variables tested in the following analyses statistical distributional assumptions were tested. The Kolmogorov-Smirnov (K-S) test was used to test normal distribution in all tests except for bootstrapping analyses. Levene's test was used to test variance homogeneity across groups as assumed in ANOVAs. The Box-M-Test was used to test covariance homogeneity as assumed in MANOVAs or RM-MANOVAs with less than three within variables. Homoscedasticity as assumed in regressional analyses was checked first, by testing the normal distribution of the residuals by means of the K-S test and second, by inspecting the scatter-plot of the residuals plotted against the predicted values. Whenever the distributional assumption was not fulfilled ($p < .05$), it is reported with the analysis of the specific hypothesis test.

7.3.1 Randomization Checks and Exploratory Analyses

Before the statistical analyses of the hypotheses tests were run, it was tested whether the randomized assignment of participants with respect to cognitive learning prerequisites (processing capacity) and perceived task demands was successful. Furthermore, it was tested whether participants in the low-knowledge conditions differed from the participants in the high-knowledge conditions in domain knowledge about physiology and topic knowledge about the domain (manipulation check). It was further investigated exploratory how learners rated their interest in the topic after learning because according to Tobias (1994) different levels in prior knowledge should influence learners' interest. Interest might be a motivational variable important for the expertise reversal effect. Finally, it was tested whether high-knowledge participants with the integrated format studied longer than high-knowledge participants with the separated format, because high-knowledge participants had the option to study shorter than low-knowledge participants. The means and standard deviations of the control variables are shown in Table 24.

Cognitive processing capacity. The IQ-subscale (processing capacity) data from two participants were incomplete because of their delayed appearance at the laboratory and therefore excluded from the analysis. Cronbach's α of the overall scale was .55. A 2 (instructional format) X 2 (prior knowledge) ANOVA on cognitive processing capacity was run. Participants with integrated format did not differ in their processing capacity performance from participants with separated format ($F < 1$). Low prior knowledge participants did not differ from participants with high prior knowledge ($F(1, 51) = 1.42$, $MSE = 50.31$, $p = .24$, $f = 0.17$). There was no interaction effect between instructional format and prior knowledge indicating that none of the experimental groups differed in its mean processing capacity from the other groups ($F < 1$). Because there were no differences between the groups, cognitive processing capacity was not considered any further in the following analyses.

Table 24

Means and standard deviations of the control variables as a function of prior knowledge and instructional format

		Low prior knowledge		High prior knowledge	
		Integrated (*n* = 15)	**Separated** (*n* = 14)	**Integrated** (*n* = 14)	**Separated** (*n* = 14)
Cognitive Processing Capacity [a]	*M*	29.46	28.77	27.43	26.25
	SD	(7.35)	(8.79)	(5.27)	(6.65)
% Knowledge in Physiology	*M*	22.04	19.35	63.29	69.64
	SD	(9.69)	(13.25)	(7.91)	(12.19)
Perc. Task Demands [b]	*M*	2.40	2.29	3.00	2.50
	SD	(1.12)	(1.07)	(1.41)	(0.94)
Interest [b]	*M*	3.47	2.29	4.07	3.71
	SD	(0.83)	(1.27)	(1.14)	(1.14)
Time: Structure Graphic [c]	*M*	180.00	180.00	52.21	58.53
	SD	(0.00)	(0.00)	(25.23)	(21.39)
Time: Process Graphic [c]	*M*	600.00	600.00	362.73	319.06
	SD	(0.00)	(0.00)	(164.66)	(121.28

Note. [a] scale range: 0 to 54, [b] scale range: 1 to 6; [c] time in seconds;
M = mean, *SD* = standard deviation.

Domain knowledge in physiology. Cronbach's α of the prior knowledge test in physiology was .92. To test for group differences in general prior knowledge in physiology, a 2 (instructional format) X 2 (prior knowledge) ANOVA was run. Participants assigned to the low-knowledge conditions performed worse in the test on general physiology than the medicals students assigned to the high-knowledge conditions (F(1, 53) = 249.32, MSE = 119.66, p < .01, f = 2.12), indicating that the distinction in low- and high-knowledge was successful. There was neither an effect of instructional format (F < 1) nor an interaction effect between instructional format and prior knowledge on general knowledge in physiology (F(1, 53) = 2.43, MSE = 119.66, p = .13, f = 0.10).

Perceived task demands. To test whether participants or groups differed with respect to how hard they think it is to learn with the integrated and separated format, their perceived task demands were analyzed. A paired t-test yielded that participants did not differ in how hard they rated the separated and integrated format (t(56) = - 1.64, p = .11).

A correlational analysis yielded that participants who rated that it is harder to learn with the separated format rated that it is easier to learn with the integrated format and vice versa (r(57) = - .27, p = .04) indicting that participants preferred either the integrated or the separated format.

A 2 (instructional format) x 2 (prior knowledge) ANOVA on the perceived task demands of the respective format participants learned with did neither demonstrate a main effect of prior knowledge (F(1, 53) = 1.79, MSE = 1.32, p = .19, f = 0.18), nor a main effect of instructional format (F(1, 53) = 1.02, MSE = 1.32, p = .32, f = 0.14), nor an interaction effect (F < 1), indicating that participants assigned to the low prior knowledge condition with integrated format did not think it was harder to learn with their format than participants assigned to the low-knowledge group with separated format or participants assigned to the high-knowledge groups with integrated or the separated format.

Moreover, partial correlations with prior knowledge and instructional format as control variables were run to get a closer look on the relations between the cognitive load type and motivational measures. The higher participants rated the difficulty of the material (ECL), the higher they rated the perceived task demands of their instructional format (r(53) = .36, p = .01), indicating that perceived task demands an instructional design influences the difficulty ratings of the material after learning.

Interest. A 2 (instructional format) x 2 (prior knowledge) ANOVA tested whether the instructional format and prior knowledge influenced learners' situational interest.

The ANOVA yielded that high-knowledge participants rated the content to be more interesting than low-knowledge participants ($F(1,53) = 12.12$, $MSE = 1.22$, $p < .01$, $f = 0.44$). Furthermore, participants with the integrated format rated the content to be more interesting than participants with the separated format ($F(1,53) = 6.93$, $MSE = 1.22$; $p = .01$, $f = 0.32$). There was no interaction between prior knowledge and instructional format ($F(1,53) = 1.99$, $MSE = 1.22$, $p = .16$, $f = 0.17$).

Partial correlations with prior knowledge and instructional format as control variables revealed that the higher participants rated the difficulty of the content (ICL), the less interesting they rated the content to be ($r(53) = -.27$, $p = .05$) indicating that content difficulty and interest are closely related.

Learning durations. Two 2 (instructional format) X 2 (prior knowledge) ANOVAs on the learning durations for the structure and the process graphic yielded that high prior knowledge participants studied the structure graphic shorter than low prior knowledge participants ($F(1, 53) = 825.16$, $MSE = 268.31$, $p < .01$, $f = 3.89$). The analysis yielded no further effects, indicating that there were no differences in the average learning times between the instructional format conditions (all *F*s < 1). The same result pattern was shown for the learning times of the process graphic. High-knowledge participants studied shorter than low-knowledge participants ($F(1, 53) = 93.23$, $MSE = 10258.18$, $p < .01$, $f = 1.31$). No further effects were revealed (all *F*s < 1). Bonferroni-adjusted comparisons showed that high-knowledge participants with integrated format studied the structure graphic ($p = .31$) as well as the process graphic ($p = .26$) equally long as high-knowledge participants with separated format. Levene's tests indicated that the variances differed between the groups for the structure graphic ($F(3, 53) = 12.57$, $p < .01$) as well as the process graphic ($F(3, 53) = 20.09$, $p < .01$). This is not surprising, because low knowledge participants could not differ in their learning times, whereas high prior knowledge participants could and did. Because the group sizes did not differ, the results of the ANOVAs should be robust despite the violation of the variance homogeneity assumption (Field, 2005).

7.3.2 Learning Outcomes

Cronbach's α of the post-tests were calculated to get information on the different tests. To analyze participants' knowledge gains and learning outcomes a 2 (time of test: pre vs. post) x 2 (prior knowledge: high vs. low) x 2 (instructional format: integrated vs. separated) RM-ANOVA was run for each knowledge test. Time of test was the within subject variable, whereas instructional format and prior knowledge were between subject variables.

The Box-M-tests were significant in all analyses (p < .01) thereby indicating that the assumptions of covariance homogeneity were not fulfilled. These results were mainly caused by the fact that the prior knowledge of high-knowledge participants varied more highly in the pre-tests than in the post-tests, whereas it was the other way round for low-knowledge participants. Although this is a systematic variance pattern not in line with statistical pre-conditions that should be met, RM-ANOVAs are rather robust against such violations. Moreover, using non-parametric tests for analyzing assumed interaction effects are not regarded as more helpful in analyzing the data. Although the interpretability of the following results is not assumed to be severely limited, they should not be easily generalized. Table 25 summarizes the means and standard deviations of all knowledge tests as a function of time of test, instructional format, and prior knowledge.

Table 25

Means and standard deviations of % correct in all knowledge tests as a function of time of test, prior knowledge and instructional format

% correct	Time of test		Low prior knowledge Integrated (*n* = 15)	Low prior knowledge Separated (*n* = 14)	High prior knowledge Integrated (*n* = 15)	High prior knowledge Separated (*n* = 14)
Terminology	*Pre*	*M*	0.56	0.20	89.48	86.71
		SD	(1.56)	(0.74)	(12.14)	(12.76)
	Post	*M*	65.37	66.27	97.22	97.42
		SD	(15.17)	(18.55)	(7.78)	(4.81)
Labeling	*Pre*	*M*	1.67	0.74	72.32	76.64
		SD	(4.47)	(1.55)	(19.10)	(11.03)
	Post	*M*	72.64	54.17	94.79	98.07
		SD	(20.41)	(20.33)	(7.30)	(5.57)
Complex Facts	*Pre*	*M*	1.44	0.49	29.79	28.08
		SD	(2.66)	(0.97)	(9.93)	(7.08)
	Post	*M*	40.76	31.01	56.09	55.68
		SD	(11.38)	(11.20)	(9.59)	(6.41)
Transfer	*Pre*	*M*	2.00	0.54	42.32	42.68
		SD	(4.22)	(1.45)	(11.87)	(15.96)
	Post	*M*	25.00	26.70	55.36	50.98
		SD	(5.71)	(11.48)	(12.19)	(9.17)

Note. STS = secondary task stimulus; *n* = sample size; *M* = mean; *SE* = standard error.

Terminology. Cronbach's α of the terminology post-test was .82 indicating good test validity. The 2 x 2 x 2 RM-ANOVA showed that participants in all four conditions had a significant gain in knowledge on terminology as indicated by the main effect of time of test ($F(1, 53) = 432.13$, $MSE = 91.85$, $p < .01$, $\eta_P^2 = .89$). The interaction between time of test and prior knowledge was significant ($F(1, 53) = 244.94$, $MSE = 91.85$, $p = .01$, $\eta_P^2 = .82$). Bonferroni-adjusted comparisons yielded that high-knowledge participants outperformed low-knowledge participants in the pre-test (all *p*s < .01) and post-test (all *p*s < .01). The interaction between time of test and instructional format was not significant ($F < 1$) neither was the interaction between time of test and prior knowledge and instructional format ($F < 1$), indicating that participants with integrated format did neither differ from participants with separated format in the pre-test (low-knowledge: $p = .91$; high-knowledge: $p = .41$) nor in the post test (low-knowledge: $p = .85$; high-knowledge: $p = .97$). There was neither a split-attention nor an expertise reversal effect. Low-knowledge participants gained about 65% of possible new knowledge on terms. High-knowledge participants raised their outcomes about 10% which reflects 70% of possible new knowledge for them.

Labeling. Cronbach's α of the labeling test was .89 indicating high test validity. The 2 x 2 x 2 RM-ANOVA showed that participants in all four conditions had a significant gain in knowledge on labeling as indicated by the main effect of time of test ($F(1, 53) = 338.43$, $MSE = 148.98$, $p < .01$, $\eta_P^2 = .87$). The interaction between time of test and prior knowledge was significant ($F(1, 53) = 77.43$, $MSE = 148.98$, $p = .01$, $\eta_P^2 = .59$). Bonferroni-adjusted comparisons yielded that high-knowledge participants outperformed low-knowledge learners in the pre-test (all *p*s < .01) and post-test (all *p*s < .01). The interaction between time of test and instructional format was significant ($F(1, 53) = 4.13$, $MSE = 148.98$, $p = .05$, $\eta_P^2 = .07$). The three-way interaction between time of test and prior knowledge and instructional format tended to be significant ($F(1, 53) = 3.27$, $MSE = 148.98$, $p = .01$, $\eta_P^2 = .06$). Bonferroni-adjusted comparisons yielded that low-knowledge participants with integrated format outperformed low-knowledge learners on the post-test only ($p < .01$; pre-test: $p = .83$), whereas high-knowledge participants did not differ from each other on the post-test ($p = .57$; pre-test: $p = .31$), indicating an ordinal expertise reversal effect. Low-knowledge participants with integrated format gained about 70% of possible new knowledge, whereas low-knowledge participants gained only about 55% of possible new knowledge. High-knowledge participants raised their outcomes about 20% which reflects about 85% of possible new knowledge for them.

Complex facts. Cronbach's α of the test about complex facts was .80 indicating good test validity. The 2 x 2 x 2 RM-ANOVA showed that participants in all four conditions had a significant gain in knowledge on labeling as indicated by the main effect of time of test ($F(1, 53) = 385.42$, $MSE = 70.60$, $p < .01$, $\eta_P^2 = .88$). The interaction between time of test and prior knowledge was significant ($F(1, 53) = 6.47$, $MSE = 70.60$, $p = .01$, $\eta_P^2 = .11$). Bonferroni-adjusted comparisons yielded that high-knowledge participants outperformed low-knowledge learners in the pre-test (all *p*s < .01) and post-test (all *p*s < .01). The interaction between time of test and instructional format tended to be significant ($F(1, 53) = 3.05$, $MSE = 70.60$, $p = .09$, $\eta_P^2 = .05$). The three-way interaction between time of test and prior knowledge and instructional format was not significant ($F(1, 53) = 1.10$, $MSE = 70.60$, $p = .30$, $\eta_P^2 = .02$). Bonferroni-adjusted comparisons yielded that low-knowledge participants with integrated format outperformed low-knowledge learners on the post-test only ($p = .02$; pre-test: $p = .68$), whereas high-knowledge participants did not differ from each other on the post-test ($p = .35$; pre-test: $p = .47$), indicating an ordinal expertise reversal effect. Low-knowledge participants with integrated format gained about 40% of possible new knowledge on complex facts, whereas low-knowledge participants gained only about 30% of possible new knowledge. High-knowledge participants raised their outcomes about 25% which reflects about 38% of possible new knowledge for them.

Transfer. Cronbach's α of the transfer test was .83 indicating good test validity. The 2 x 2 x 2 RM-ANOVA showed that participants in all four conditions had a significant gain in knowledge on terminology as indicated by the main effect of time of test ($F(1, 53) = 196.17$, $MSE = 44.86$, $p < .01$, $\eta_P^2 = .79$). The interaction between time of test and prior knowledge was significant ($F(1, 53) = 31.10$, $MSE = 44.86$, $p < .01$, $\eta_P^2 = .37$). Bonferroni-adjusted comparisons yielded that high-knowledge participants outperformed low-knowledge participants in the pre-test (all *p*s < .01) and post-test (all *p*s < .01). The interaction between time of test and instructional format was not significant ($F < 1$) neither was the three-way interaction between time of test and prior knowledge and instructional format ($F(1, 53) = 2.58$, $MSE = 44.86$, $p = .11$, $\eta_P^2 = .05$), indicating that participants with integrated format did neither differ from participants with separated format in the pre-test (low-knowledge: $p = .70$; high-knowledge: $p = .89$) nor in the post test (low-knowledge: $p = .65$; high-knowledge: $p = .25$). Thus, there was neither a split-attention nor an expertise reversal effect. Low-knowledge learners gained about 25% of possible new knowledge. High-knowledge participants raised their outcomes about 10% which reflects about 17% of possible new knowledge for them.

7.3.3 Cognitive Load Measures

Experiment 2 measured cognitive load types by subjective ratings. Furthermore, the cognitive load types were also tried to be measured by participants viewing behavior. In analyzing these measures different statistical analyses were applied. First of all and in accordance with former cognitive load studies, it was probed by means of 2 (prior knowledge: high vs. low) x 2 (instructional format: integrated vs. separated) ANOVAs whether there were mean differences across the groups caused on the subjective ratings and measures of viewing behavior. Furthermore, correlations between the three subjective load type measures were analyzed to investigate how (dis)similar the measures are. Moreover and new in the context of cognitive load research, mediated moderation analyses were used to test whether, and if so, which type of (subjectively rated) cognitive load (and/or interest) mediated the interaction effect of prior knowledge and instructional format demonstrated for knowledge on labeling and complex facts.

7.3.3.1 Subjective ratings: Cognitive load types

In investigating the subjective cognitive load ratings different statistical approaches were used. First, it was tested whether there were mean differences among the groups. Furthermore, relations between the three load types were analyzed to investigate how (dis)similar the measures are in addition to the mean differences. Finally, mediated moderational analyses were used to test whether, and if so, which type of cognitive load and/or interest mediated the expertise reversal effect on knowledge about the labels of the structure parts and about complex facts concerning the physiological processes.

Group Differences. 2 (instructional format) X 2 (prior knowledge) ANOVAs were run for the subjective ratings of intrinsic, extraneous, and germane cognitive load. The means and standard deviations for the cognitive load type ratings are shown in Table 26.

Intrinsic cognitive load. Participants with low prior knowledge rated the difficulty of the learning content higher than participants with high prior knowledge ($F(1, 53) = 86.14$, $MSE = .43$, $p < .01$, $f = 1.21$). Furthermore, participants with integrated format rated the content difficulty lower than students with separated format ($F(1,53) = 6.91$, $MSE = .43$, $p < .05$; $f = 0.22$). Participants' prior knowledge did not moderate their ratings of content difficulty ($F < 1$).

Table 26

Means and standard deviations of the cognitive load ratings as a function of prior knowledge and instructional format

		Low prior knowledge		High prior knowledge	
		Integrated (n = 15)	**Separated** (n = 14)	**Integrated** (n = 14)	**Separated** (n = 14)
ICL	*M*	3.80	4.29	2.21	2.64
	SD	(0.86)	(0.61)	(0.43)	(0.63)
ECL	*M*	2.87	3.29	2.93	3.43
	SD	(0.74)	(0.83)	(1.07)	(1.09)
GCL	*M*	4.93	4.50	4.29	4.5
	SD	(0.59)	(0.65)	(0.83)	(0.76)

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$. PK = prior knowledge.
ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

Extraneous cognitive load. With regard to prior knowledge there were no differences between participants' ratings of the difficulty of the learning materials ($F < 1$). However, participants with separated format tended to rate the difficulty of the material higher than students with integrated format, although the results did not reach statistical significance ($F(1,53) = 3.39$, $MSE = .89$, $p = .07$, $f = 0.25$). Whether participants had high or low prior knowledge did not moderate their difficulty ratings of the learning materials in the instructional format groups ($F < 1$).

Germane cognitive load. There was a trend that participants with low prior knowledge reported to have concentrated more during learning than participants with high prior knowledge, although this result did not reach significance ($F(1,53) = 2.9$, $MSE = .51$, $p = .09$, $f = 0.23$). Whether participants learned with integrated or separated format did not influence their ratings of how much they concentrated during learning ($F < 1$). However, there was a marginally significant interaction effect between prior knowledge and instructional format, although this result did not reach statistical significance ($F(1,53) = 2.9$, $MSE = .51$, $p = .09$, $f = 0.23$). Bonferroni-adjusted comparisons indicated that low-knowledge participants with integrated format reported to have concentrated more than high-knowledge participants with the same instructional format ($p = .02$), whereas low-knowledge participants with separated format did not differ in their ratings of concentration from high-knowledge participants ($p = 1$).

Partial correlations between cognitive load types. Partial correlations with prior knowledge and instructional format as control variables were run, to get a closer look on the relations between the cognitive load type measures. The analyses revealed that the higher the higher participants rated the difficulty of the material (ECL), the higher they rated the content difficulty (ICL; $r(53) = .33$, $p = .01$), the lower they rated their concentration level during learning (GCL; $r(53) = -.29$, $p = .03$), and the higher they had rated the perceived task demands of their instructional format ($r(53) = .36$, $p = .01$). Moreover, the results revealed that the higher participants rated the difficulty of the content (ICL), the less interesting they rated the content to be ($r(53) = -.27$, $p = .05$).

Mediated moderation. In investigating the mechanism of the expertise reversal effect found for knowledge on the labeling test and on the test about complex facts mediated moderational analyses were used by first, estimating multiple regression models and second, estimating the confidence intervals of the mediated moderation by means of bootstrapping. Because an interaction effect between instructional format and prior knowledge on potential mediators (cognitive load types and interest) is a necessary condition of a mediated moderation (see Muller, Judd & Yzerbyt, 2005), germane cognitive load was the only possible mediator according to the results of the above reported 2 x 2 ANOVAs yielding a marginally significant interaction effect on germane cognitive load only. Neither the 2 x 2 ANOVAs on intrinsic or extraneous cognitive load nor on interest showed a significant interaction effect between instructional format and prior knowledge. Hence, germane cognitive load was the only load type probed for mediating the two expertise reversal effects.

First, three types of regression models were estimated to describe the first three steps of mediated moderation for knowledge on labeling and complex facts, respectively: the dependent variable model, the mediator variable model, and the mediational dependent variable model. Because instructional format and prior knowledge were nominal predictors with two categories, they were contrast-coded (cf. Cohen et al., 2003), whereas all continuous predictor variables were centered thereby making 0 to the mean and trying to keep multicollinearity as low as possible because the correlation of one predictor with other predictors maybe reduced (cf. Aiken & West, 1991). All predictors were included at once in all models. The mediational dependent variable models used for testing the mediated moderation did not include intrinsic and extraneous cognitive load and interest as covariates, because otherwise they would have been also included in the mediator variable model (Preacher, Rucker & Hayes, 2007). However, to see whether they were important omitted variables in the mediational dependent variable models, additional expanded dependent variable models were run which included these variables as covariates. Thus, by including these variables it was possible to test whether the parameter estimates of germane cognitive load were similar in both model types. The parameter estimates (B-weights, standard errors, and ß-weights) of the mediator variable model are summarized in Table 27, whereas the parameter estimates of the dependent variable models, the mediational dependent variable models, and the expanded mediational dependent variable models are summarized in Table 28.

Step 1: Simple dependent variable models. In accordance with the above reported ANOVAs, the multiple regression model for *knowledge about labeling* yielded a significant effect of prior knowledge ($t(53) = 8.18$, $p < .01$), a marginally significant effect of instructional format ($t(53) = 1.88$, $p = .07$), and a significant moderation effect of prior knowledge and instructional format ($t(53) = -\ 2.69$, $p < .01$). The multiple regression model for *knowledge about complex facts* yielded a significant effect of prior knowledge ($t(53) = 7.64$, $p < .01$), a marginally significant effect of instructional format ($t(53) = 1.94$, $p = .06$) as well as a marginally significant moderation effect of prior knowledge and instructional ($t(53) = -\ 1.79$, $p < .08$). Although not reaching statistical significance, it is tested whether this marginally significant effect supports the mediated moderation assumption of the expertise reversal effect.

Step 2: Mediator variable model. In accordance with the above reported ANOVA, the regression model for germane cognitive load with instructional format, prior knowledge and instructional format x prior knowledge as predictors yielded no effect of instructional format ($t(49) = 0.58$, *ns*), but a marginally significant effect for prior knowledge ($t(53) = 5.30$, $p = .09$) as well as a marginally significant moderation effect ($t(53) = 5.30$, $p = .09$).

Table 27

Summary of the multiple regression parameter estimates for the GCL mediator variable model

	B	***SE_B***	**ß**
DV: GCL[a]			
Prior Knowledge	-0.16	0.09	-.22+
Instructional Format	0.06	0.09	.06
Instructional Format x Prior Knowledge	-0.16	0.09	-.22$^+$
Model fit	*corr.* $R^2 = .06$		

Note. $^+$ $p < .10$, * $p < .05$, ** $p < .01$. GCL = germane cognitive load. DV = dependent variable.

Step 3a: Mediational dependent variable models. *The* multiple regression model for knowledge about *labeling* with instructional format, prior knowledge, instructional format x prior knowledge, and germane cognitive load ratings as predictors yielded that there was a significant effect of prior knowledge ($t(52) = 8.78$, $p < .01$), a marginally significant effect of instructional format ($t(52) = 1.77$, $p = .08$), still a significant but reduced moderation effect ($t(52) = -2.20$, $p = .03$), and a significant effect of germane cognitive load suggesting germane cognitive load to be a mediator of the expertise reversal effect of knowledge about labeling ($t(52) = 2.26$, $p = .03$). The multiple regression model for knowledge about *complex facts* with instructional format, prior knowledge, instructional format x prior knowledge, and germane cognitive load ratings as predictors yielded a significant effect of prior knowledge ($t(52) = 8.79$, $p < .01$), a marginally significant effect of instructional format ($t(52) = 1.84$, $p = .07$), but no more significant moderation effect ($t(52) = -1.15$, $p = .26$). Moreover, there was a significant effect of germane cognitive load suggesting germane cognitive load to be a mediator of the expertise reversal effect of knowledge about complex facts ($t(52) = 3.19$, $p < .01$).

Step 3b: Expanded mediational dependent variable models. To test whether intrinsic and extraneous cognitive load or interest were important omitted variables and potential simple mediators, multiple regression models including these variables were run. The multiple regression model for *knowledge about labeling* with instructional format, prior knowledge, instructional format x prior knowledge, germane cognitive load, intrinsic cognitive load, extraneous cognitive load, and interest ratings as predictors yielded that there was again a significant effect of prior knowledge ($t(49) = 5.50$, $p < .01$), a significant moderation effect of instructional format and prior knowledge ($t(49) = -2.03$, $p = .05$), and still a significant effect of germane cognitive load ($t(49) = 2.15$, $p = .04$). All other variables in the model (instructional format ($t(49) = 1.63$, $p = .11$), intrinsic cognitive load, extraneous cognitive load, and interest (all ts < 1, *ns*)) did not reach significance, suggesting that omitting the latter three variables from further significance analyses is not problematic. The multiple regression model for knowledge about *complex facts* yielded that there was again a significant effect of prior knowledge ($t(49) = 5.28$, $p < .01$) and still a significant effect of germane cognitive load ($t(49) = 2.55$, $p = .01$). All other variables did not reach significance (instructional format ($t(49) = 1.51$, $p = .14$), moderation ($t(49) = 1.39$, $p = .17$), extraneous cognitive load ($t(49) = -1.58$, $p = .12$), intrinsic cognitive load, and interest (both ts < 1, *ns*), suggesting that omitting the latter three variables from further significance analyses is also not problematic.

Table 28

Summary of the multiple regressions for the dependent variable, the mediational dependent variable, and the expanded mediational dependent variable models on knowledge about labeling and complex facts

	Labeling			Complex facts		
	B	SE_B	*ß*	*B*	SE_B	*ß*
DV-Models						
Prior knowledge	16.51	2.02	.72**	10.00	1.31	.71**
Instructional format	3.80	2.02	.17+	2.54	1.31	.18+
Instructional format x Prior knowledge	- 5.44	2.02	- .24**	- 2.34	1.31	- .17^{+}
Model fit	$R_{corr.}^2$ = .57			$R_{corr.}^2$ = .58		
Mediational DV-Models						
Prior knowledge	17.55	2.00	.76**	10.91	1.24	.77**
Instructional format	3.45	1.95	.15+	2.23	1.21	.16+
Instructional format x Prior knowledge	- 4.40	2.00	- .19*	- 1.43	1.24	- .10
GCL	6.41	2.84	.20*	5.62	1.76	.29**
Model fit	$R_{corr.}^2$ = *0.60*			$R_{corr.}^2$ = .59		
Expanded Mediational DV-Models						
Prior knowledge	19.25	3.50	.84**	11.22	2.12	.79**
Instructional format	3.59	2.20	.16	2.02	1.34	.14
Instructional format x Prior knowledge	- 4.25	2.09	- .19*	- 1.76	1.27	- .12
GCL	6.63	3.08	.21*	4.76	1.87	.24*
ICL	2.53	3.47	.12	- 0.05	2.11	- .01
ECL	- 0.50	2.40	- .02	- 2.31	1.46	- .15
Interest	0.80	1.95	.04	- 0.73	1.18	- .07
Model fit	$R_{corr.}^2$ = .58			$R_{corr.}^2$ = .59		

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$. DV = dependent variable, ICL = Intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load.

Step 3b: Expanded mediational dependent variable models. To test whether intrinsic and extraneous cognitive load or interest were important omitted variables and potential simple mediators, multiple regression models including these variables were run. The multiple regression model for *knowledge about labeling* with instructional format, prior knowledge, instructional format x prior knowledge, germane cognitive load, intrinsic cognitive load, extraneous cognitive load, and interest ratings as predictors yielded that there was again a significant effect of prior knowledge

($t(49) = 5.50$, $p < .01$), a significant moderation effect of instructional format and prior knowledge ($t(49) = - 2.03$, $p = .05$), and still a significant effect of germane cognitive load ($t(49) = 2.15$, $p = .04$). All other variables in the model (instructional format ($t(49) = 1.63$, $p = .11$), intrinsic cognitive load, extraneous cognitive load, and interest (all *t*s < 1, *ns*)) did not reach significance, suggesting that omitting the latter three variables from further significance analyses is not problematic. The multiple regression model for knowledge about *complex facts* yielded that there was again a significant effect of prior knowledge ($t(49) = 5.28$, $p < .01$) and still a significant effect of germane cognitive load ($t(49) = 2.55$, $p = .01$). All other variables did not reach significance (instructional format ($t(49) = 1.51$, $p = .14$), moderation ($t(49) = 1.39$, $p = .17$), extraneous cognitive load ($t(49) = - 1.58$, $p = .12$), intrinsic cognitive load, and interest (both *t*s < 1, *ns*), suggesting that omitting the latter three variables from further significance analyses is also not problematic.

Step 4: Bootstrapping analyses and confidence intervals. The bootstrapping method in combination with confidence intervals was used to test the significance of the described mediated moderation (see step 4) by using the SPSS Macro and the according commands of model 2 (mediated moderation) programmed by Preacher et al. (2007). The bootstrapping analyses estimated the conditional indirect effect of instructional format on labeling and complex factual knowledge, respectively, through germane cognitive load for the low and high prior knowledge participants individually, with prior knowledge assumed as moderating the effect of instructional format on germane cognitive load, while producing a 95 % bootstrap confidence interval for these conditional indirect effects based on 5000 bootstrap samples. For *low-knowledge* participants' learning outcomes on the test about *labeling* the bias corrected 95 % confidence intervals did not include zero ($CI_{lower} = 0.08$, $CI_{upper} = 3.57$) thereby indicating significance. For *high-knowledge* participants' learning outcomes on the test about *labeling*, the bias corrected confidence intervals included zero ($CI_{lower} = - .3.70$, $CI_{upper} = 0.83$) thereby indicating non-significance. Figure 22 shows the path model of the mediated moderation model for the learning outcomes on the labeling test with germane cognitive load as mediator variable.

For *complex factual knowledge* the result pattern was very similar. For *low-knowledge* participants' learning outcomes on the test about *complex facts* the bias corrected 95 % confidence intervals did not include zero ($CI_{lower} = 0.12$, $CI_{upper} = 2.89$) thereby indicating significance. For *high-knowledge* participants' learning outcomes on the test about *complex facts*, the bias corrected confidence intervals included zero ($CI_{lower} = - .2.85$, $CI_{upper} = 0.83$) thereby indicating non-significance. Figure 23 shows the path model of the mediated moderation model for the learning outcomes on the test about complex facts with germane cognitive load as mediator variable.

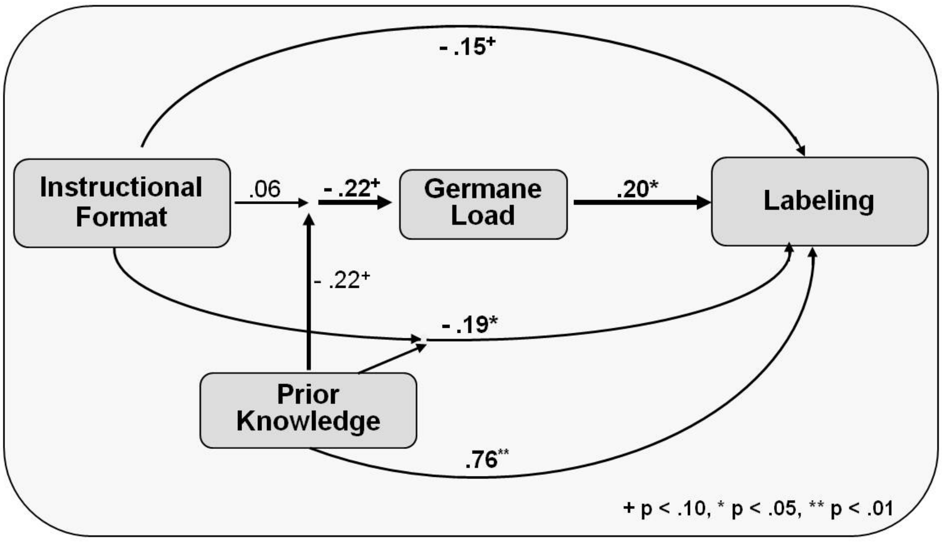


Figure 22. Path model of the mediated moderation on knowledge about labeling

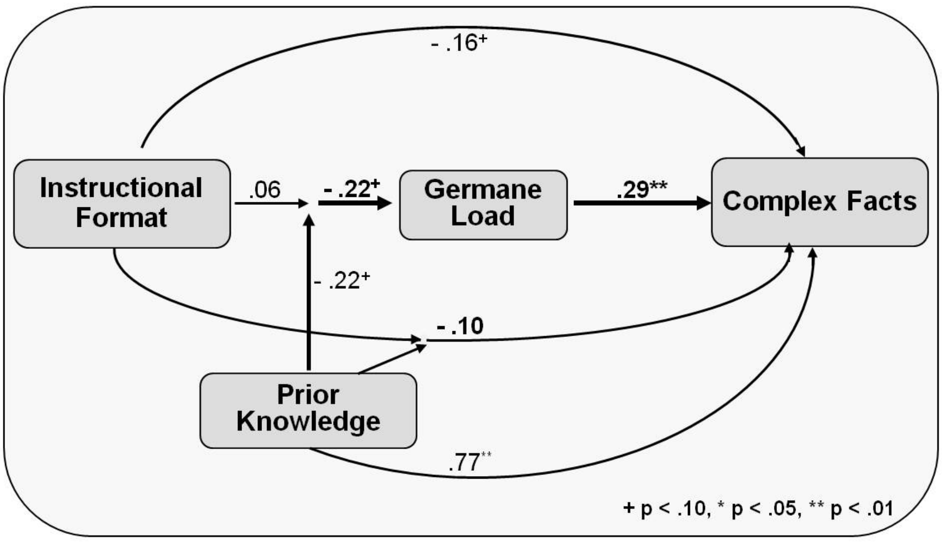


Figure 23. Path model of the mediated moderation on knowledge about complex facts

Final step: Proportions of effects mediated. To summarize, the bootstrapping results indicated that germane cognitive load was a partial mediator of the moderation (interaction) effect of instructional format and prior knowledge on learning outcomes of labeling and complex facts. Germane cognitive load mediated 18.85% of the total expertise reversal effect on knowledge about labeling and 38.43% of the total expertise reversal effect on knowledge about complex facts.

7.3.3.2 Behavioral activities: Measures of viewing behavior

The analysis of participants' viewing behavior concerns three groups of measures differing in temporal and spatial information they are based on: (1) average fixation durations, (2) percentage of dwell time on AOI types, and (3) percentage of spatial transition density between different AOI types per minute which shows in more detail how interrelated the participants processed the different information units or AOIs. The variables were analyzed by 2 (graphic: structure vs. process) x 2 (instructional format: integrated vs. separated) x 2 (prior knowledge: high vs. low) RM-ANOVAs with graphic as within-factor and prior knowledge as well as instructional format as between-factor. Moreover, in order to test exploratory whether the viewing behavior measures are related with subjective cognitive load ratings and with the performance on the four knowledge tests, Pearson's correlations with two-sided significance tests were run for each of the four experimental groups. Because of many non-significant results, only the significant and marginally significant results will be presented.

7.3.3.2.1 Average fixation duration

The means and standard deviations of the average fixation durations are shown in Table 29.

Average fixation duration. The 2 x 2 x 2 RM-ANOVA showed that participants had shorter fixation durations on the structure graphic than on the process graphic ($F(1, 53) = 65.82$, $MSE = 97.81$, $p < .01$, $\eta_P^2 = .55$). The interaction between type of graphic and prior knowledge was significant ($F(1, 53) = 13.90$, $MSE = 97.81$, $p = .01$, $\eta_P^2 = .21$). Bonferroni-adjusted comparisons yielded that high-knowledge participants with integrated format had shorter fixation durations than low-knowledge participants with integrated format ($p < .01$), whereas high-knowledge participants with separated format tended to have shorter fixation durations than low-knowledge participants with separated format ($p = .07$). The interaction between type of graphic and instructional format tended to be significant ($F(1, 53) = 2.90$, $MSE = 97.81$, $p = .09$, $\eta_P^2 = .05$), whereas the three-way interaction between type of graphic, prior knowledge and instructional format was not significant ($F < 1$). Bonferroni-adjusted comparisons yielded that low-knowledge participants with separated format tended to have shorter fixations on the structure graphic than low-knowledge participants with integrated format ($p = .07$).

Table 29

Means and standard deviations of the average fixation durations as a function of prior knowledge, instructional format, and type of graphic

			Low prior knowledge		High prior knowledge	
	Graphic		Integrated (*n* = 15)	Separated (*n* = 14)	Integrated (*n* = 14)	Separated (*n* = 14)
Average fixation duration	*Structure*	*M*	253.06	237.09	226.62	220.36
		SD	(28.44)	(23.36)	(17.15)	(23.04)
	Process	*M*	256.74	249.67	246.71	244.16
		SD	(24.93)	(21.92)	(19.47)	(29.79)

Note. SF cond. = separated format condition.

Correlations. Pearson's correlations with two-sided significance tests were run for each of the four experimental groups and yielded the following results. High-knowledge learners with separated format had higher outcomes on the test on complex facts, the longer their average fixation durations on the structure graphic was ($r(14) = .70$, $p < .01$). However, the longer their average fixation durations on the process graphic, there was a tendency that they had lower outcomes on the transfer test ($r(14) = -.52$, $p = .06$). Low-knowledge learners reported significantly lower ratings in intrinsic cognitive load, the longer their average fixation durations on the process graphic ($r(14) = -.58$, $p = .03$) and similarly they tended to report lower ratings in intrinsic cognitive load, the longer their average fixation durations were on the structure graphic ($r(14) = -.50$, $p = .07$). Notably, these participants also reported significantly higher ratings in germane cognitive load, the longer their fixation durations were ($r(14) = .56$, $p = .04$). There were no correlations for participants with integrated format.

7.3.3.2.2 Dwell times

It was tested how many percentages of their fixation time participants spent on textual and graphical information. The means and standard deviations of participants dwell time in percent on text as well as on graphic with and without labels are shown in Table 30.

Table 30

Means and standard deviations of the dwell times (and summed dwell times) as a function of prior knowledge, instructional format, and type of graphic

	Graphic		Low prior knowledge Integrated (*n* = 15)	Low prior knowledge Separated (*n* = 14)	High prior knowledge Integrated (*n* = 14)	High prior knowledge Separated (*n* = 14)
% dwell time on text-AOIs	*Structure*	*M*	69.99	45.22	67.22	47.13
		SD	(9.90)	(12.44)	(7.23)	(14.09)
	Process	*M*	76.74	68.95	73.43	67.07
		SD	(6.12)	(9.69)	(6.22)	(9.05)
% dwell time on graphic-AOIs	*Structure*	*M*	25.60	31.91	25.12	30.95
		SD	(8.73)	(5.12)	(8.09)	(13.28)
	Process	*M*	15.08	15.80	17.27	16.34
		SD	(5.16)	(5.88)	(5.48)	(6.88)
% dwell time on non-textual AOIs (graphic- and label-AOIs in SF groups	*Structure*	*M*	25.60	44.78	25.12	44.53
		SD	(8.73)	(7.16)	(8.09)	(16.34)
	Process	*M*	15.08	19.22	17.27	19.20
		SD	(5.16)	(7.19)	(5.48)	(7.59)
% of summed means	*Structure*		95.59	90.00	92.34	91.66
% of aummed means	*Process*		92.54	88.17	90.07	86.27

Note. SF = separated format.

Text. The 2 x 2 x 2 RM-ANOVA showed that participants spent more percentages of their fixation time on text information presented on the process graphic than on the structure graphic ($F(1, 53) = 72.64$, $MSE = 78.57$, $p < .01$, $\eta_P^2 = .58$). The interaction between type of graphic and prior knowledge was not significant ($F < 1$) indicating that prior knowledge did not influence how many percentages of fixation time were spent on text information. The interaction between type of graphic and instructional format was significant ($F(1, 53) = 21.35$, $MSE = 78.57$, $p < .01$, $\eta_P^2 = .29$). Bonferroni-adjusted comparisons yielded that low-knowledge participants with integrated format spent more percentages of their time on text than low-knowledge participants with separated format ($p < .01$). The same was true for high-knowledge participants ($p = .04$). The three-way interaction of type of graphic, prior knowledge, and instructional format was not significant ($F < 1$).

Graphic. The 2 x 2 x 2 RM-ANOVA showed that participants spent more percentages of their fixation time on graphical information presented on the structure graphic than on the process graphic ($F(1, 53) = 90.98$, $MSE = 47.13$, $p < .01$, $\eta_P^2 = .63$). The interaction between type of graphic and prior knowledge was not significant ($F < 1$) indicating that prior knowledge did not influence how many percentages of fixation time were spent on graphical information. The interaction between type of graphic and instructional format was significant ($F(1, 53) = 5.77$, $MSE = 47.13$, $p = .02$, $\eta_P^2 = .10$). Bonferroni-adjusted comparisons yielded that low-knowledge participants with integrated format tended to spent less percentages of their time on graphical information on the structure graphic than low-knowledge participants with separated format ($p = .07$), although this result was not significant. The three-way interaction of type of graphic, prior knowledge, and instructional format was not significant ($F < 1$).

If the label-AOIs in the separated format condition were considered as graphic-AOIs because they do not contain any textual information but provide spatial information where a physiological structure is or where a physiological processes happens, the analysis yielded that participants with separated format spent significantly more time on the non-textual AOIs on the structure graphic than low-knowledge participants with integrated format ($p < .01$). The same was true for high-knowledge participants ($p < .01$). Concerning the process graphic only low-knowledge participants with separated format tended to spent more time on non-textual AOIs ($p = .09$) although this result was not significant.

***Correlations*.** Pearson's correlations with two-sided significance tests were run for each of the four experimental groups and yielded the following results. High-knowledge participants reported significantly lower ratings in intrinsic cognitive load, the longer their dwell time on text information on the process graphic ($r(14) = -.61$, $p = .02$). However, these participants had lower outcomes on the test on terms, the longer their dwell time on text information on the process graphic ($r(14) = -.55$, $p = .04$). Complementarily, they reported significantly higher ratings in intrinsic cognitive load ($r(14) = .68$, $p < .01$), and they had higher outcomes on test on terms, the longer their dwell time was on graphical information on the process graphic ($r(14) = .57$, $p = .03$). Low-knowledge participants reported significantly higher ratings in intrinsic cognitive load ($r(14) = .53$, $p = .05$), higher ratings in extraneous cognitive load ($r(14) = .53$, $p = .05$), and lower ratings in germane cognitive load($r(14) = -.73$, $p < .01$), the higher their dwell time was on text information on the process graphic. Complementarily, low-knowledge reported marginally lower ratings in intrinsic cognitive load ($r(14) = -.50$, $p = .07$), marginally lower ratings in extraneous cognitive load ($r(14) = .53$, $p = .06$), and significantly higher ratings in germane cognitive load ($r(14) = -.73$, $p < .01$), the higher their dwell time was on graphical information on the process graphic. Moreover, the higher their dwell time was on graphical information on the process graphic, the marginally higher were their outcomes on the test on complex facts ($r(14) = .49$, $p = .07$), and significantly higher on the transfer test ($r(14) = .65$, $p = .01$). In contrast to these results, high-knowledge participants only tended to report higher ratings in intrinsic cognitive load, the higher their dwell time was on graphical information on the process graphic ($r(14) = -.50$, $p = .07$). They also tended to have higher outcomes on the test on terms, the longer their dwell times on graphical information on the structure graphic ($r(14) = .47$, $p = .09$). In contrast to all other participants, low-knowledge learners with integrated format showed no significant correlation between any cognitive load rating and the dwell times. However, they had significantly lower outcomes on the labeling test($r(15) = -.60$, $p = .03$) and marginally lower outcomes on the test on complex facts ($r(15) = -.45$, $p = .10$), the longer their dwell times on graphical information were on the structure graphic.

7.3.3.2.3 Switching behavior

To investigate in more detail how participants processed the different AOI-units, the percentages of spatial transition density per minute for four types of transitions were analyzed: (1) transitions between different text AOIs (structure graphic: (14*14-14) = 182 transition possibilities; process graphic: (20*20-20) = 380 transition possibilities), (2) transitions between different graphic AOIs (structure graphic: 182 transition possibilities; process graphic: 380 transition possibilities), (3) transitions between non-corresponding text and graphic AOIs (structure graphic: 182 transition possibilities; process graphic: 380 transition possibilities), and (4) transitions between corresponding text and graphic AOIs (structure graphic: 14*2 = 28 transition possibilities; process graphic: 20*2 = 40 transition possibilities). Because the transition possibilities differ between the type of graphic and type of switching according to the different numbers of AOIs, direct comparisons across these different types are not interpretable. The results of all four Box-M-tests of the according 2 x 2 x 2 RM-ANOVAs were significant (all ps $< .01$), indicating that the covariances differed between the groups. The result patterns suggest that the variances of high-knowledge learners are more diverse than those of low-knowledge participants. Although RM-ANOVAs are rather robust against violations against this statistical pre-condition, the results should only be generalized carefully. The means and standard deviations of participants' switching behavior are shown in Table 31.

Table 31

Means and standard deviations of the switching variables as a function of prior knowledge, instructional format, and type of graphic

			Low prior knowledge		High prior knowledge	
	Graphic		Integrated (n = 15)	Separated (n = 14)	Integrated (n = 14)	Separated (n = 14)
text – text switching[a]	*Structure*	*M*	5.31	7.66	7.49	16.37
		SD	(1.56)	(1.51)	(2.87)	(6.18)
	Process	*M*	1.61	1.70	1.81	2.85
		SD	(0.35)	(0.34)	(0.48)	(1.21)
graphic – graphic switching[a]	*Structure*	*M*	3.25	5.06	2.56	7.06
		SD	(1.11)	(0.59)	(1.06)	(3.84)
	Process	*M*	0.61	0.89	0.77	0.88
		SD	(0.24)	(0.25)	(0.54)	(0.38)
non-corresponding text – graphic switching[a]	*Structure*	*M*	6.87	4.46	9.77	6.62
		SD	(1.13)	(1.02)	(3.01)	(1.91)
	Process	*M*	2.44	0.87	3.53	1.12
		SD	(0.40)	(0.15)	(1.38)	(0.44)
corresponding text – graphic switching[a]	*Structure*	*M*	26.02	10.20	62.74	12.45
		SD	(2.69)	(4.57)	(30.93)	(8.30)
	Process	*M*	7.43	2.87	12.98	7.43
		SD	(0.64)	(1.03)	(5.65)	(0.64)

Note. [a] switching behavior in % transition density / min.

Text-text switching. The 2 x 2 x 2 RM-ANOVA showed that participants had a higher text-text switching density per minute on the structure graphic than on the process graphic ($F(1, 53) = 243.62$, $MSE = 6.09$, $p < .01$, $\eta_P^2 = .82$). This difference, however, should not be interpreted any further because of different amounts of AOIs. The interaction between type of graphic and prior knowledge was significant ($F(1, 53) = 26.62$, $MSE = 6.09$, $p < .01$, $\eta_P^2 = .33$). Furthermore, the interaction between type of graphic and instructional format was significant ($F(1, 53) = 29.86$, $MSE = 6.09$, $p < .01$, $\eta_P^2 = .36$), and so was the three-way interaction of type of graphic, prior knowledge, and instructional format ($F(1, 53) = 9.07$, $MSE = 6.09$,

$p < .01$, $\eta_P^2 = .15$). Bonferroni-adjusted comparisons yielded that high-knowledge participants with separated format switched between significantly more different text AOIs per minute than low-knowledge participants with separated format ($p < .01$) on the structure graphic as well as on the structure graphic. Moreover, high-knowledge participants with separated format switched on both graphics between significantly more different text AOIs per minute than high-knowledge participants with integrated format (both *p*s < .01), indicating that high-knowledge participants with separated format showed the highest text switching behavior considering their learning times. Low-knowledge participants with separated format only tended to switch between more text AOIs on the structure graphic than low-knowledge participants with integrated format ($p = .07$).

Correlations. The Pearson's correlations with two-sided significance tests demonstrated that high-knowledge learners with separated format tended to report higher ratings in germane cognitive load, the higher their text-text switching density was on the process graphic, although this effect was not significant ($r(14) = .46$, $p = .09$). Furthermore, these participants tended to have better test outcomes on the transfer test, the higher their text-text switching density was ($r(14) = .52$, $p = .06$). In contrast to these results, low-knowledge participants with separated format tended to report lower ratings of germane cognitive load, the higher their text-text switching density was on the structure graphic, although this result was not significant ($r(14) = -.52$, $p = .06$). These low-knowledge participants also tended to report higher ratings of extraneous cognitive load ($r(14) = .51$, $p = .06$), and higher ratings of intrinsic cognitive load ($r(14) = .49$, $p = .08$), the higher their text-text switching density was on the structure graphic, although these results reached only marginal significance. Moreover, these low-knowledge participants tended to have lower outcomes on the transfer test, the higher their density was on the structure graphic ($r(14) = -.49$, $p = .08$). A similar pattern was found for low-knowledge participants with integrated format who also tended to have lower outcomes on the transfer test ($r(15) = -.45$, $p = .09$) and on the labeling test ($r(15) = -.44$, $p = .10$), the more they switched between many different text AOIs on the process graphic, although these results are not significant. Despite the correlations between learning outcomes and text-text density per minute there were no correlations with subjective cognitive load ratings in the two integrated format groups.

Graphic-graphic switching. The 2 x 2 x 2 RM-ANOVA showed that participants had a higher graphic-graphic switching density per minute on the structure graphic than on the process graphic ($F(1, 53) = 214.65$, $MSE = 2.54$, $p < .01$, $\eta_P^2 = .80$). This difference, however, should not be interpreted more closely because of different amounts of AOIs. The interaction between type of graphic and prior knowledge was significant ($F(1, 53) = 10.61$, $MSE = 2.54$, $p < .01$, $\eta_P^2 = .17$). Furthermore, the interaction between type of graphic and instructional format was significant ($F(1, 53) = 7.14$, $MSE = 2.54$, $p < .01$, $\eta_P^2 = .17$). However, the three-way interaction of type of graphic, prior knowledge, and instructional format was not significant ($F < 1$). Bonferroni-adjusted comparisons yielded that high-knowledge participants with separated format switched between significantly more different graphic AOIs per minute than low-knowledge participants with separated format ($p = .03$) on the structure graphic, and so did high-knowledge participants with integrated format compared with low-knowledge participants with integrated format ($p = .02$). Moreover, low-knowledge participants with separated format switched between significantly more different graphic AOIs per minute on the structure graphic ($p = .04$) as well as on the process graphic ($p = .05$) than low-knowledge participants with integrated format. High-knowledge participants with separated format switched between significantly more graphic AOIs on the structure graphic only ($p = .05$) than high-knowledge participants with integrated format.

Correlations. The Pearson's correlations with two-sided significance tests were run for each of the four experimental groups and yielded the following results. High-knowledge learners with separated format reported lower ratings in germane cognitive load, the higher their graphic-graphic switching density was on the process graphic ($r(14) = -.68$, $p < .01$). At the same time, these participants had higher test outcomes on the test about terms, the higher their graphic-graphic switching density per minute was ($r(14) = .74$, $p < .01$). In contrast to these results, low-knowledge participants with separated format reported significantly higher ratings of germane cognitive load ($r(14) = .69$, $p < .01$) and significantly lower ratings of intrinsic cognitive load ($r(14) = -.66$,$p < .01$), the higher their graphic-graphic switching density was on the process graphic. These low-knowledge participants had higher outcomes on the transfer test ($r(14) = .64$, $p = .01$) and also showed a trend for higher outcomes on the test about terms ($r(14) = .48$, $p < .08$), the higher their graphic-graphic switching density was on the process graphic. Different from low-knowledge participants with separated format low-knowledge participants with integrated format reported significantly higher extraneous cognitive load ratings, the higher their graphic-graphic density per minute was on the process graphic ($r(15) = .51$, $p = .05$). Furthermore, the low-knowledge participants with integrated format had significantly lower outcomes on the labeling test, the higher their graphic-graphic density per minute was on the structure graphic ($r(14) = -.72$, $p < .01$). High-knowledge participants with integrated format tended to report higher intrinsic cognitive load ratings, the higher their graphic-graphic density per minute on the process graphic was, although this result was not statistically significant ($r(14) = .47$, $p = .09$), but showed significantly higher learning outcomes on the transfer test ($r(14) = .55$, $p = .04$), the higher their graphic-graphic density per minute was on the structure graphic.

Non-corresponding text-graphic switching. The 2 x 2 x 2 RM-ANOVA showed that participants had a higher switching density per minute between non-corresponding text-graphic AOIs on the structure graphic than on the process graphic ($F(1, 53) = 358.02$, $MSE = 1.94$, $p < .01$, $\eta_P^2 = .87$). This difference, however, should not be interpreted more closely because of different amounts of AOIs. The interaction between type of graphic and prior knowledge was significant ($F(1, 53) = 12.66$, $MSE = 1.94$, $p < .01$, $\eta_P^2 = .19$). The interaction between type of graphic and instructional format was not significant ($F(1, 53) = 2.29$, $MSE = 1.94$, $p = .14$, $\eta_P^2 = .04$). The three-way interaction of type of graphic, prior knowledge, and instructional format was neither significant ($F < 1$). Bonferroni-adjusted comparisons yielded that all participants with integrated format had a significantly higher switching density per minute between non-corresponding text-graphic AOIs than participants with separated format on both graphic types (all ps $< .01$). Moreover, high-knowledge participants with integrated format switched between significantly more non-corresponding text-graphic AOIs per minute on both graphic types than low-knowledge participants with integrated format (all ps $< .01$), whereas high-knowledge participants with separated format switched only between significantly more non-corresponding text-graphic AOIs on the structure graphic ($p < .01$).

Correlations. The Pearson's correlations with two-sided significance tests yielded the following results. High-knowledge learners with separated format reported higher ratings in germane cognitive load, the higher their switching density per minute between non-corresponding text-graphic AOIs was on the process graphic ($r(14) = .56$, $p = .04$). Different from these results, low-knowledge participants with separated format reported significantly higher ratings of extraneous cognitive load, the higher their switching density per minute between non-corresponding text-graphic AOIs was on the structure graphic ($r(14) = .56$, $p = .04$) as well as on the process graphic ($r(14) = .66$, $p = .01$). Despite these correlations with cognitive load measures there were no correlations with learning outcomes in these two groups with separated format. High-knowledge learners with integrated format tended to have higher outcomes on the labeling test, the higher their switching density per minute between non-corresponding text-graphic AOIs was on the structure graphic

($r(14) = .52, p = .06$), although this result did not reach statistical significance. Different from this positive correlation, low-knowledge participants tended to have lower outcomes on the labeling test, the higher their switching density per minute between non-corresponding text-graphic AOIs was on the process graphic ($r(15) = -.50$, $p = .06$), although this result did neither reach statistical significance. Despite these correlations with learning outcomes there were no correlations with subjective cognitive load ratings in these two integrated format groups.

Corresponding text-graphic switching. The 2 x 2 x 2 RM-ANOVA showed that participants had a higher switching density per minute between corresponding text-graphic AOIs on the structure graphic than on the process graphic ($F(1, 53) = 104.84$, $MSE = 119.32$, $p < .01$, $\eta_P^2 = .66$). This difference, however, should not be interpreted more closely because of different amounts of AOIs. The interaction between type of graphic and prior knowledge was significant ($F(1, 53) = 15.29$, $MSE = 119.32$, $p < .01$, $\eta_P^2 = .22$). Furthermore, the interaction between type of graphic and instructional format was significant ($F(1, 53) = 41.67$, $MSE = 119.32$, $p < .01$, $\eta_P^2 = .44$), and the three-way interaction of type of graphic, prior knowledge, and instructional format was also significant ($F(1, 53) = 13.72$, $MSE = 119.32$, $p < .01$, $\eta_P^2 = .21$). Bonferroni-adjusted comparisons yielded that all participants with integrated format had a significantly higher switching density per minute between corresponding text-graphic AOIs than participants with separated format on both graphic types (all ps $\leq .01$). Moreover, high-knowledge participants with integrated format switched between significantly more corresponding text-graphic AOIs per minute on both graphic types than low-knowledge participants with integrated format (both ps $< .01$), whereas high-knowledge participants with separated format did not differ from low-knowledge participants with separated format.

Correlations. The Pearson's correlations with two-sided significance tests showed that high-knowledge learners with separated format tended to report higher ratings in intrinsic cognitive load, the higher their switching density per minute between corresponding text-graphic AOIs was on the structure graphic ($r(14) = .52$, $p = .06$), although this result did not reach statistical significance. Low-knowledge participants with separated format tended to report higher ratings in extraneous cognitive load, the higher their switching density per minute between corresponding text-graphic AOIs was on the structure graphic ($r(14) = .48$, $p = .09$), although this result was not statistically significant. Despite these correlations with cognitive load measures there were no correlations with learning outcomes in these two groups with separated format. High-knowledge learners with integrated format tended to have higher outcomes on the transfer test, the higher their switching density per minute between corresponding text-graphic AOIs was on the structure graphic ($r(14) = .50$, $p = .07$), although this result was not significant. Similar to these results, low-knowledge participants reported significantly higher ratings in intrinsic cognitive load ($r(15) = .68$, $p < .01$). and tended to report higher ratings in extraneous cognitive load ($r(15) = .50$, $p = .06$), the higher their switching density per minute between corresponding text-graphic AOIs was on the structure graphic. Nevertheless, these participants had higher outcomes on the test on complex facts, the higher their switching density per minute between corresponding text-graphic AOIs was on the process graphic ($r(15) = .58$, $p = .02$).

7.4 Summary and Discussion

The aim of Experiment 2 was to investigate the mechanism underlying the expertise reversal effect concerning spatial contiguity between text and graphic in multimedia learning, because there are seemingly contradictory explanations of this mechanism in literature. In other words, the aim was to find out why less knowledgeable learners benefit from integrated multimedia formats and suffer from separated multimedia formats, whereas more knowledgeable learners do not benefit or even suffer from integrated multimedia formats but perform equally well or even better with separated multimedia formats. Therefore, the study investigated whether high-knowledge learners with integrated format were either seduced to invest less relevant learning resources (germane cognitive load explanation) or whether they were highly loaded by irrelevant coordination processes of redundant information (extraneous cognitive load explanation). Furthermore, it was investigated whether high-knowledge learners with separated format were able to use relevant processing strategies reflected in specific reading/processing behavior (germane cognitive load explanation) compared to low-knowledge learners or whether high knowledge learners with separated format were able to encode and process the graphic very efficiently (extraneous cognitive load explanation). As The assumptions of the intrinsic cognitive load explanation are either in line with the germane or with the extraneous cognitive load explanation but do not differentiate between these two explanations, the results are discussed with a focus on the germane and the extraneous cognitive load explanations.

The experiment showed an ordinal expertise reversal effect. High-knowledge participants with separated text-graphic format performed as well in learning about the physiological processes in the kidney as high-knowledge participants with integrated text-graphic format, whereas low-knowledge participants with integrated format outperformed low-knowledge participants with separated format. Subjective ratings scales were used to measure intrinsic, extraneous and germane cognitive load separately, whereas viewing behavior was used as an objective online-measure of learners' reading or visual processing behavior. The results of the subjective cognitive load ratings support the germane cognitive load explanation suggested by the CTML and CIM. The results of the viewing behavior are also interpreted in favor of the germane cognitive load explanation. The analyses also revealed interesting results about the relation between subjective cognitive load ratings and viewing behavior. The hypotheses and results of the experiment concerning the expertise reversal effect and its underlying mechanism will be summarized and discussed in more detail in the next sections.

7.4.1 Expertise Reversal Effect: Learning Outcomes

According to the expertise reversal effect it was assumed that low-knowledge learners would benefit from integrated formats but suffer from separated formats, whereas high-knowledge learners would either perform equally well in both formats (ordinal interaction) or would even benefit from separated format but suffer from integrated format (disordinal interaction or complete reversal). This performance pattern should appear especially in complex task or tasks demanding inferences but not necessarily in easy, non-complex tasks. This hypothesis was tested by using four different knowledge tests which measured four types of knowledge with increasing complexity in the domain of physiological processes in the nephron as the functional unit of the kidney: (1) technical terms of structures, (2) labeling or mapping the terms to the corresponding depicted structures, (3) complex facts about processes, and (4) transfer tasks demanding complex inferences about the physiological processes. With regard to prior knowledge it was assumed that high-knowledge learners should generally perform better than low-knowledge learners.

In accordance with the prior knowledge assumption high-knowledge participants performed better in all knowledge tests than low-knowledge participants. The gain of new knowledge within the range of possible new knowledge was about the same level for low-knowledge and high-knowledge participants when the amount of possible new knowledge was considered for the high-knowledge participants. Concerning the instructional format, there were no differences between participants with integrated format and participants with separated format in knowledge about the *technical terms* as the easiest knowledge type. This result is in accordance with the assumption that spatial contiguity effects do not need to apply to non-complex information. For the more complex tests on labeling and on complex facts, expertise reversal effects were demonstrated. High-knowledge participants with integrated format performed as equally well as high-knowledge participants with separated format. However, low-knowledge participants with separated format performed worse than low-knowledge participants with integrated format. This result pattern is in accordance with Mayer's (2001) findings that more knowledgeable learners performed equally well but low-knowledge learners suffered from separated formats. These results corroborate hypotheses *H 2.1a* and *H 2.1 b*. However, it differs from Kalyuga et al.'s (1998) results who found a complete reversal in comparing integrated format with graphic only. Thus, hypothesis *H 2.1c* was not supported. Although high-knowledge participants performed equally well in both tests, it must be considered that there was a ceiling effect for these participants in the test about labeling. Hence, the interaction effect on the labeling test might also be caused by the limited possibility for further group differences on this test scale. However, there was also no complete reversal on the test about complex facts, although there was no ceiling effect on this scale. With regard to the most complex test, the transfer test, no expertise reversal effect appeared. Low-knowledge participants with integrated format did not outperform participants with separated format. This result is not in line with most of the previous research findings showing the split-attention effect on transfer tasks with low-knowledge learners. However, Purnell et al. (1991) did neither find a split-attention effect on their test demanding inferences but only on factual knowledge in the domain of geography. Furthermore, Reisslein et al. (2006) did only find an expertise reversal effect on near-transfer tasks but not on far-transfer tasks and raise the question whether there are any inherent underlying differences among different knowledge domains that would give rise to the observed differences in the results. An explanation for the different findings concerning the knowledge tests might be that instructional design effects might progress from factual knowledge to comprehension depending on the complexity of the domain as already outlined during discussing the results of Experiment 1. The finding that high-knowledge learners with integrated format did not perform worse than high-knowledge learners with separated format might also be explained by learning strategies that high-knowledge students might have used when learning with integrated format. According to O'Reilly and McNamara (2007) a complete reversal should only occur, if learners lack the ability to use appropriate reading or processing strategies. As the high-knowledge learners in this study were well educated medical students, they might have been too skilled to suffer substantially from integrated formats. Whether high-knowledge participants used special strategies, is further discussed below in the sections about viewing behavior.

7.4.2 Explanations of the Expertise Reversal Effect: Cognitive Load Measures

According to the seemingly contradicting hypotheses assumed by the CTML (or CIM) and CLT, two patterns concerning the three cognitive load types and the associated processes represented in viewing behavior were hypothesized. Before these hypotheses and the results with regard to the viewing behavior are discussed, the results on the subjective cognitive load and interest ratings are summarized and discussed.

7.4.2.1 Subjective ratings: Cognitive load types and interest

According to the CTML and CLT different hypotheses concerning the cognitive load type pattern can be derived. These hypotheses were investigated by asking participants for their subjective ratings of the content difficulty (intrinsic cognitive load), the difficulty of the materials (extraneous cognitive load), and their level of concentration during learning (germane cognitive load), respectively. Furthermore, participants' were asked about their interest in the content. The measures are discussed with respect to the hypotheses concerning the cognitive load explanations as well as their validity concerning the cognitive load constructs.

Intrinsic cognitive load. According to all explanations it was assumed that high knowledge learners should experience less intrinsic cognitive load than low-knowledge learners independently from the instructional format. In accordance with the assumptions about intrinsic cognitive load (*H 2.2.1a*) it was demonstrated that high-knowledge participants rated the content difficulty lower than low-knowledge participants. However, participants with separated format rated the content difficulty higher than participants with integrated format. This effect of the instructional format contradicts all three explanations assuming that intrinsic cognitive load is not influenced by spatial contiguity. Hence, hypothesis *H 2.2.1b* is to be rejected. Concerning the validity of content difficulty as intrinsic cognitive load that should be independent of instructional format two aspects have to be considered. First, the finding that all participants with the separated format rated the content to be more difficult independently of their prior knowledge suggests that the content difficulty can hardly be perceived or judged without being influenced by the instructional format. This was also shown by the positive correlation between intrinsic and extraneous cognitive load ratings. Second, the finding that high-knowledge participants found the learning content easier than low-knowledge participants fits very well with assumptions about intrinsic cognitive load. Hence, content difficulty shows strong face validity with respect to the intrinsic cognitive load construct consisting of element interactivity and learners' prior knowledge, although the measurement seems to be intrigued by effects of instructional format.

Extraneous cognitive load. Whereas the extraneous cognitive load explanation predicts an interaction effect for extraneous cognitive load, the intrinsic and germane cognitive load explanations predict a main effect of instructional format on extraneous cognitive load. The finding that participants with separated format reported higher ratings in material difficulty (extraneous cognitive load) than participants with integrated format supports the intrinsic and germane cognitive load explanations (*H 2.2.2a*) but contradicts both hypotheses of the extraneous cognitive load explanation (*H 2.2.2b* and *H 2.2.2c*). According to this result high-knowledge learners are not overloaded when learning with integrated formats as suggested by Kalyuga et al. (1998). This is important with respect to Sweller's (in press) assumption that almost all instructional design effects are caused by extraneous cognitive load. Concerning the validity of participants' subjective ratings of how difficult it was for them to learn with the materials an interesting finding is that despite the high correlation with ratings of content difficulty, low-knowledge learners did not find the materials more difficult than high-knowledge learners. Thus, there was no main effect of prior knowledge as was demonstrated for content difficulty. Second, the subjective ratings of the difficulty of the materials did significantly predict knowledge about complex facts, whereas content difficulty did not. Third, the higher participants rated the perceived task demands of the respective instructional format before learning, the more difficult they rated the materials (but not the contents) to be after learning. These findings are in line with the construct of extraneous cognitive load. Thus, material difficulty is interpreted as a more or less valid measure of extraneous cognitive load. Nevertheless, the results do not support the extraneous cognitive load explanation.

Germane cognitive load. According to the germane cognitive load explanation it was assumed that high-knowledge learners with separated format invested more germane cognitive load than high-knowledge learners with integrated format, whereas the reversed pattern was assumed for low-knowledge learners. According to the extraneous and intrinsic cognitive load explanation, however, it was assumed that high-knowledge learners as well as learners with integrated format should not differ in the investment of germane cognitive load from low-knowledge learners and learners with separated format. The results showed that high-knowledge participants tended to rate their concentration level lower than low-knowledge participants. Furthermore, high-knowledge participants with integrated format had the lowest concentration level ratings, whereas low-knowledge participants with integrated format had the highest concentration level ratings. This result is partially in accordance with the germane cognitive load explanation assumed predominantly by researchers of the CTML and CIM, although there was no complete reversal of the concentration ratings. The result contradicts the extraneous and intrinsic cognitive load explanations (*H 2.2.3d*). The fact that participants with higher extraneous cognitive load ratings had lower germane cognitive load ratings suggests that extraneous and germane cognitive load processes may be partially complementary processes. However, the different group results suggest that these two load types are not fully complementary. Moreover, the results of the multiple regression and bootstrapping analyses showed that only germane cognitive load was a partial mediator in the mediated moderation model, even though extraneous cognitive load was related to knowledge about complex facts. These results stress the importance of germane cognitive load in explaining the expertise reversal effect and indicate rather good validity of this measure. Hence, the subjective ratings of concentration level support hypothesis *H 2.2.3b* suggesting an ordinal interaction. Because there was no complete disordinal interaction, hypothesis H *2.2.3a* could not be supported. The fact that germane cognitive load was more predictive for low- than for high-knowledge participants calls for future research that investigates whether learners' estimates of concentration level depends on their expertise in general. Such research should consider the possibility that expertise reversal effects probably only occur, if learners lack the ability to use appropriate reading, or more generally, appropriate processing strategies as suggested by O'Reilly and McNamara (2007).

Interest. According to both theories no differences in interest among the instructional groups were assumed. The first result showed that participants with high prior knowledge rated the content to be more interesting than low-knowledge participants. Furthermore, the interest ratings were negatively correlated with the rated difficulty of the learning content. Both results are in line with the general finding that the higher learners' prior knowledge is, the higher their interest is in the topic (see Tobias, 1994). The second result showed that participants with the integrated format rated the content to be more interesting than participants with the separated format. This result might be a hint that the design of instructions influences learners' motivation and thereby their learning outcomes. However, interest was not a predictor for learning outcomes, when the cognitive load ratings were included in the multiple regression analyses. Moreover, the finding that interest was not correlated with the level of concentration or germane cognitive load indicates that germane cognitive load is not identical with a motivational aspect like interest. Thus, participants' ratings of concentration levels (germane cognitive load) seem to better represent cognitive instead of motivational aspects.

To summarize, the intrinsic and especially the extraneous cognitive load ratings are not in accordance with the extraneous cognitive load explanation, because high-knowledge participants with separated format rated not only the difficulty of the content but also the difficulty of the materials higher than participants with integrated format. Furthermore, the germane cognitive load ratings showed a trend towards an interaction effect and mediated almost 40 % of the interaction effect on knowledge about complex facts. Again, this finding does not support the extraneous cognitive load explanation but is in line with the germane cognitive load explanation.

7.4.2.2 Behavioral activities: Measures of viewing behavior

Concerning the perceptual processing of the learning materials, the CTML (germane cognitive load explanation) provides different assumptions than the CLT (extraneous cognitive load explanation). These different assumptions were investigated by different processing variables determined from participants' viewing behavior during learning: Fixation durations, dwell times on text and on graphic, switching within one representation (text-text or graphic-graphic), and switching between two representations (corresponding and non-corresponding text-graphic switches). Whether and how these measures support the explanations is discussed in the following paragraphs.

Fixation durations. According to findings of former eye tracking studies (e.g., Underwood et al., 2004), it is assumed that more knowledgeable learners should have shorter fixation durations than less knowledgeable learners. This assumption was derived in the context of the CTML as well as CLT. Moreover, it is assumed that if fixation durations represent cognitive load, they might indicate which instructional format causes higher cognitive load. The results showed that high-knowledge participants' average fixation durations were indeed shorter than low-knowledge participants' ones thereby indicating that fixation duration might be sensitive enough to detect differences in intrinsic cognitive load. The results can be interpreted in support of hypothesis *H 2.3.1a*. However, there was no overall effect of instructional format. Nevertheless, low-knowledge participants with separated tended to have shorter fixation durations than low-knowledge participants with integrated format but only on the structure graphic. This result might be interpreted as indication that low-knowledge participants with separated format invested less germane cognitive load than low-knowledge participants with integrated formats because the longer their fixation durations were, the higher their germane cognitive load ratings. Although this result is hypothesized by the germane cognitive load explanation (*H 2.3.1b),* it should be taken cautiously because average fixation durations seem to be a too overall measure which is not specific enough to clearly differentiate between the three cognitive load types. The results might also indicate that fixation durations are too insensitive to measure the cognitive load caused by the instructional format, because the instructional design effects demonstrated in this experiment were in general much smaller than the prior knowledge effects. Further research is needed to test whether stronger instructional design effects would be reflected in fixation durations as demonstrated by Ozcelik et al. (2009) in a study on color coding.

Dwell times on graphic and on text. Concerning the dwell time learners spent on textual and on graphical information the germane cognitive load explanation assumes that learners with separated format should spent more time on text and less time on graphic (text focused processing) than learners with integrated format, whereas the extraneous cognitive load explanation assumes that learners with separated format spent more time on the graphic than learners with integrated format. Furthermore, according to the extraneous cognitive load explanation it is assumed that high-knowledge learners with separated format spent the highest amount of their time on the graphic because of ignoring redundant text information. The finding that all participants with separated format spent more time on graphic is partially in accordance with the extraneous cognitive load explanation (*H2.3.2b*). However, there was no interaction effect assuming that especially high-knowledge learners with separated format concentrate on the graphical information thereby ignoring redundant text information. Moreover, the assumption that low-knowledge learners suffer from visual search processes on the graphic indicating extraneous cognitive load was not supported because low-knowledge participants with separated format not only reported lower ratings in extraneous cognitive load and higher ratings in germane cognitive load but also had higher learning outcomes, the longer they processed the graphical information. Hence, despite the fact that participants with separated format processed the graphical information longer than participants with integrated format, the correlational results contradict the assumptions of the extraneous cognitive load explanation why these learners process graphical information longer than participants with integrated format. These results are neither in line with the germane cognitive load explanation which assumes that low-knowledge learners with separated format should not process the graphic very deeply.

With regard to the dwell time on text, the germane cognitive load explanation was neither supported concerning group differences. Participants with integrated format spent more time on the text information than did participants with separated format, although the germane cognitive load explanation assumes the opposite. Hence, hypothesis *H 2.3.2a* was not supported. Although the text-AOIs sizes were bigger in the integrated format conditions, the difference in size is not regarded as the only causal factor of the instructional format effect because first it is assumed that participants processed the materials purposefully and did not by chance look around on the materials. Second, although the summed percentages of dwell times on text and graphic AOIs were lower for participants with separated format, one has to keep in mind that these learners made longer saccades between text and graphic which are not counted as dwell time. Nevertheless, it cannot be ruled out that the different AOI sizes influenced or even caused the result. Therefore, the results of the dwell times on text should be taken more cautiously and future studies should control for this factor. Despite these concerns, the dwell times showed that participants spent most of their time (about 50% to 75%) on processing the text, also in separated format conditions. Therefore, the assumption of Kalyuga et al. (1998) that high-knowledge learners with separated format would only process the graphic and ignore text was not supported. Whether this result depends on the specific content participants had to learn, however, is not clear. Whereas Kalyuga and colleagues used electrical circuits where the graphic is the main source of information, this dissertation used a complex system (functioning of the nephron) that needs lots of verbal explanations to make the graphic with its depicted processes comprehensible. Nevertheless, the finding that participants concentrated on the text is in line with former eye tracking research which also found that persons concentrate more on text than on pictures (e.g., Hannus & Hyönä, 1999, Rayner et al. 2001). Moreover, low-knowledge participants with separated format reported higher ratings in extraneous cognitive load and lower ratings in germane cognitive load the longer they processed the text information. This result is in accordance with the germane cognitive load explanation which assumes that low-knowledge learners need graphical information to better understand the text information (Mayer & Gallini, 1990). However, the finding that these participants benefited from longer dwell times on the graphic was neither expected by the germane nor by the extraneous cognitive load explanation. To find out more precisely how participants processed both types of representation (text and graphic) the spatial transition density per minute between text-text, graphic-graphic as well as between text and graphic were analyzed in more detail.

Switching behavior within and between representations. As outlined in the hypotheses section, the germane cognitive load explanation assumes that high-knowledge learners with separated format process text information more intensively than low-knowledge learners with separated format (compensation mechanism; Mayer & Gallini, 1990), whereas the extraneous cognitive load explanation predicts that high-knowledge learners with separated format process the graphic more intensively (avoiding redundancy mechanism; Kalyuga et al., 1998). Moreover, according to both explanations, it was assumed first that learners with separated format should show a higher switching density per minute between non-corresponding text-graphic information which, in turn, should be associated with higher extraneous cognitive load. Second, it was assumed that learners with integrated formats should show a more precise integration process of corresponding text and graphic information that is a higher switching density per minute between corresponding text-graphic information which, in turn, should be associated with higher germane cognitive load and lower extraneous cognitive load.

Concerning text-text switching, the results showed that high-knowledge participants with separated format had the highest text-text spatial transition density per minute indicating that these participants switched directly between more different text units than the other participants. This finding is in line with the interaction hypothesis *H 2.3.3a* of the germane cognitive load explanation assuming that high-knowledge learners can benefit from text information alone by applying good reading strategies including imagery. This assumption was further supported by the correlation result that high-knowledge participants tended not only to report higher ratings in germane cognitive load but also tended to have higher outcomes on the transfer test, the higher their spatial transition density of text-text switches was.

Concerning graphic-graphic switching, the results showed that high-knowledge participants with separated format also had the highest graphic-graphic spatial transition density per minute indicating that these participants also switched directly between more different graphic units than the other participants. Although this results is in line with hypothesis *H 2.3.3b* of the extraneous cognitive load explanation, the correlation result that high-knowledge learners with separated format reported lower ratings in germane cognitive load, the higher their graphic-graphic switching was, does not fit to this explanation. Rather, one had expected lower ratings of extraneous cognitive load and as possible consequence maybe higher ratings in germane cognitive load. Hence, it seems too early to take this result as support of the extraneous cognitive load explanation. Moreover, according to the extraneous cognitive load explanation one expects that low-knowledge learners would report higher ratings in extraneous cognitive load, the more they switch between different graphic AOIs indicating visual search. However, low-knowledge learners with separated format reported higher ratings in germane cognitive load and reached higher outcomes on the transfer test, the more they switched between different graphic AOIs. To summarize, the results of the graphic-graphic switching density do not seem to support the extraneous cognitive load explanation.

Concerning non-corresponding text-graphic switching, the results showed that participants with integrated format switched between more different non-corresponding text-graphic units than participants with separated format. This result is not in line with hypothesis *H 2.3.3d* of the extraneous cognitive load explanation which assumes that low-knowledge learners with separated format should show the highest switching density between non-corresponding text-graphic AOIs. Nevertheless, low-knowledge participants with separated format reported high ratings in extraneous cognitive load, the higher their spatial switching density between non-corresponding text-graphic information was. This result is also assumed by the extraneous cognitive load explanation. Notably, high-knowledge participants with separated format reported higher ratings in germane cognitive load, the higher their switching density. This is not assumed by the extraneous cognitive load explanation. However, this result might be interpreted as specific learning strategy of high-knowledge learners. Unfortunately, the switching density per minute between non-corresponding text-graphic units did not correlate with learning outcomes for participants with separated format. Nevertheless, these results suggest that switching between non-corresponding text and graphic units means something different for learners with high and low prior knowledge. Such difference in the same measure of viewing behavior suggests that this processing activity and prior knowledge result in a moderated mediation (in contrast to a mediated moderation as was investigated with regard to the subjective ratings of cognitive load in Experiment 2). Such a moderated mediation effect of viewing behavior and prior knowledge was also demonstrated by Schwonke, Berthold, and Renkl (2009). To sum up in other words, the results also yielded that participants with separated format tended to show a reduced integration processing of text and graphic as assumed by the germane cognitive load explanation.

Concerning corresponding text-graphic switching, the results showed that high-knowledge participants with integrated format showed the highest switching density per minute between corresponding text-graphic units. This overall result can be taken as complementary support for the germane cognitive load explanation assuming lower integration behavior of learners with separated format. This result is in line with recent findings of Holsanova et al. (2009) as well as of Ozcelik et al. (2009) who showed that readers with separated format did not switch very often between text and corresponding graphic. However, the hypothesis *H 2.3.3c* of the germane cognitive load explanation that high-knowledge learners with separated format switch even more between more different corresponding text-graphic units than low-knowledge with separated learners was not supported. Notably, low-knowledge learners with separated as well as with integrated format reported higher ratings in extraneous cognitive load, the higher their switching density between corresponding text-graphic units. This result in combination with the higher ratings of extraneous cognitive load by low-knowledge learners with separated format when switching between non-corresponding text-graphic units might suggest that switching between text and graphic per se is an unfavorable behavior for low-knowledge learners nut not necessarily for high-knowledge learners. This interpretation is not directly in line with the germane cognitive load explanation. Nevertheless, despite the higher ratings in extraneous cognitive load, low-knowledge participants with integrated format benefited from a higher switching density between corresponding text-graphic units as shown by the correlation with higher learning outcomes on complex facts.

The high switching density between corresponding text and graphic information of high-knowledge learners with integrated format can be interpreted in different ways. On the one hand, they may reflect high redundancy and the fact that learners are not able to ignore integrated information (Kalyuga et al. 2003). However, it remains unclear why high-knowledge learners switch more intensively than low-knowledge participants. Such an interpretation would contradict former eye tracking research showing that experienced learners ignore irrelevant information (Canham & Hegarty, 2009; Haider & French, 1999). On the other hand, this way of processing might reflect a kind of superficial processing. In basic research about visual comparison Inamdar and Pomplun (2003) demonstrated that persons who had to find mismatches in a visual comparison task switched more frequently between two columns of figures close to each other than between two columns with a bigger distance in between. Inamdar and Pomplun (2003) suggested that persons use a quick visual strategy without relying on working memory to compare information close to each other, whereas they use their working memory to rehearse more items at once, when information is too far away from each other to apply the quick visual comparison strategy. According to this interpretation high-knowledge learners might show a visually active integration behavior but without a deeper processing of the semantic information. Third, the case might also be that high-knowledge learners switch between corresponding information purposefully to check whether textual and graphical information does really correspond. This might be a strategy of learners with good learning skills to counteract superficial learning as suggested by O'Reilly and McNamara (2007) in reading research. Ozuru et al. (2009) also showed the importance of reading skills in text comprehension. According to this interpretation the advanced medical students with integrated format were probably experienced with all types of text and graphic materials, and thus, had already developed good learning strategies. Hence, they would have used a good learning strategy by switching between corresponding text and graphic information to process the information more thoroughly. This interpretation seems to be the most plausible because there was a trend that these participants had higher learning outcomes on the transfer test. This result contradicts the assumption that they suffered from processing redundant information and just relied on a superficial strategy. This interpretation might also explain why high-knowledge participants with integrated format performed as well as high-knowledge participants with separated format, although high-knowledge learners with integrated format should perform worse than high-knowledge learners with separated format, if there is a complete expertise reversal effect. Notably, the variance of this transition behavior was quite large suggesting that high-knowledge learners differ a lot with respect to this behavior which might be a kind of processing strategy.

To summarize, the overall results of participants' viewing behavior seem to match better with the germane than with the extraneous cognitive load explanation, although not all results were in line with the assumptions of the germane cognitive load explanation. It was demonstrated, however, that learners with separated format tended to process the text and especially the graphic in more isolation, whereas participants with integrated format showed higher processes of integration between text and graphic without suffering from it but instead with benefiting from it. Furthermore, as assumed by the germane cognitive load explanation a higher text-text switching density seemed to be something else for low- than for high-knowledge participants. This is an important result of this dissertation which made such insights possible by the method of triangulation. The combination of eye tracking data, subjective cognitive load ratings, and learning outcomes helped to disambiguate the meaning of only one of these measures of cognitive load. It was shown (even though not for all knowledge tests and for both graphic types) that low-knowledge learners not only reported higher ratings in extraneous and lower ratings in germane cognitive load, the higher their text-text switching density per minute was, but also that they suffered in learning outcomes, whereas high-knowledge participants showed the opposite result pattern. They tended to report higher germane cognitive load ratings and reached higher learning outcomes, the higher their text-text switching density per minute was. Both result patterns are in line with the germane cognitive load explanation. A similar – although unexpected – result pattern was that high-knowledge participants who had higher graphic-graphic switching density per minute reported lower germane cognitive load (although reaching more terminology knowledge), whereas low-knowledge participants reported higher germane cognitive load and reached higher learning outcomes on the transfer test. This result pattern is rather contradictory to the extraneous cognitive load explanation but also not really in line with the germane cognitive load explanation which more or less missed to define the processing of the graphic more explicitly. It must be noted, however, that there were less clear findings for learners with integrated format. It was a rather unexpected result that low-knowledge participants with integrated format reported higher extraneous cognitive load ratings, the higher their switching density between corresponding text-graphic units per minute was. Rather, it was expected that such a processing behavior is correlated with germane cognitive load, however, it was not, even though it correlated with higher learning outcomes. The few findings concerning participants with integrated format need further research in order to find out whether learning with integrated format does not reveal any important processing pattern beyond switching between corresponding text and graphic.

7.4.3 Conclusion

The results of this study seem to justify the conclusion that simple surface characteristics of instructional formats like the spatial contiguity of text and picture influence learners' processing strategies quite strongly. The differences in processing strategies are thought to be related to differences in cognitive processes, and therefore especially to differences in extraneous and in germane cognitive load. Because high-knowledge learners with separated format reported similar subjective difficulty ratings like low-knowledge students with separated format, the instructional demands of separated formats seem to inhibit successful learning in general. However, high-prior knowledge learners seem to be able to compensate these difficulties by applying specific processing strategies as reflected in intensive text processing demonstrated by learners' viewing behavior and higher ratings in germane cognitive load. It seems that high-knowledge learners can apply their prior knowledge during reading the text as claimed by Mayer (2001) (or McNamara and Kintsch (1996)). High-knowledge learners with integrated format did not seem to suffer from high extraneous cognitive load as claimed by Kalyuga (2007). Rather, they seemed to invest germane cognitive load rather intentionally but moderately. Hence, the results better support the germane cognitive load explanation. Because different aspects of prior knowledge (recently acquired vs. acquired long time ago) were not manipulated in this study, it cannot be ruled out that the extraneous cognitive load explanation is more appropriate for different learning situations in which prior knowledge is strongly activated and easily available. Moreover, this study did focus on the expertise reversal effect with regard to spatial contiguity between text and picture. Therefore, the results should not be generalized blindly to other learning situations like for example hypertext research or learning with worked examples without further research. Altogether, it can be said that the measurement of learners' viewing behavior and of their subjective ratings in combination with learning outcomes was a fruitful combination to find out more about the meaning of spatial contiguity between text and graphic in relation with different levels of prior knowledge. The opposing results between high- and low-knowledge participants even suggest that further research might test moderated mediation analyses to investigate how prior knowledge moderates the influence of viewing behavior during learning with integrated and separated formats on learning outcomes.

III. DISCUSSION

Everything has its pros and cons.

(Old German proverb)

8 General Discussion and Future Directions

The following discussion summarizes the contributions and limitations of this thesis with regard to cognitive load explanations and cognitive load measurement. Moreover, directions in future research concerning both issues are suggested. A general conclusion will end the discussion.

8.1 Cognitive Load Explanations

The first strength of this thesis is the thorough analysis of the mechanisms assumed in literature and their empirical evidence with regard to the split-attention as well as the expertise reversal effect. A further strength is the statistical methods that were applied to test the different explanations. The contributions with regard to the explanations are outlined in the next sections. Afterwards follows the discussion of a limitation on the theoretical level that has not gained enough attention by researchers so far.

8.1.1 Testing Different Cognitive Load Explanations

Whereas former reviews concentrated on the effect sizes of the split-attention (Ginns, 2006) and expertise reversal effect (Kalyuga, 2007) and disregarded concurring explanations, this dissertation reviewed the different cognitive load explanations of both effects and their respective empirical evidence. The review on the split-attention effect as well as the review on the expertise reversal effect showed that researchers in the tradition of the CTML or CIM concentrate more on germane cognitive load, whereas CLT researchers assume extraneous cognitive load to be the most important factor in mediating both instructional design effects. In general, CLT researchers approach instructional design from a deficit-oriented perspective focusing on negative aspects like limited cognitive resources and extraneous processing, whereas CTML or CIM researchers approach the topic more from a resource-oriented perspective focusing on active, constructive processes. This dissertation showed that extraneous cognitive load is an important factor, especially for low-knowledge learners, but that it is not the whole story in multimedia learning and that it is probably not the most important factor, especially not for high-knowledge learners. In comparison to former research, one strength of this dissertation is that it demonstrated how little empirical evidence exists in favour of or against the respective explanations and a further strength is that it tested all cognitive load explanations postulated by researchers and not just one.

Concerning the split-attention effect, this dissertation pointed out that several researchers assume that not only a reduction in overall cognitive load (caused by a reduction in extraneous cognitive load) but also an increase in germane cognitive load with integrated formats are responsible for the split-attention effect. However, these few researchers did not measure germane cognitive load individually, and thus, did not provide empirical evidence that supported their assumption. This dissertation is the first study that tested the prevailing extraneous cognitive load explanation and the germane cognitive load explanation explicitly by measuring all three cognitive load types individually, whereas the former studies presented mere post-hoc explanations or interpretations of rather inappropriate cognitive load measures to support the germane cognitive load explanation. The results of this dissertation corroborate the germane cognitive load explanation better by providing subjective ratings of germane cognitive load. This finding is important because the split-attention effect is generally explained by referring to extraneous cognitive load only (Ayres & Sweller, 2005, Mayer, 2009) thereby suggesting that learning with integrated formats is just easy and does not demand learners' cognitive resources. Such implicit messages, however, do not only contradict against cognitive constructivism but might also lead to wrong assumptions about learning in the applied field comprising instructors and learners. Although the germane cognitive load explanation has been only favored by a minority of researchers so far, this dissertation suggests that CLT researchers should try to investigate and explain the split-attention effect from a more constructivist perspective.

Concerning the expertise reversal effect, Kalyuga et al. (2003) assume that extraneous cognitive load caused by redundancy for high-knowledge learners with integrated format inhibits successful knowledge acquisition. Although Kalyuga et al. (1998, 2003) stated that their redundancy explanation is not in line with former explanations of the expertise reversal effect by CTML or CIM researchers, other CLT researchers who adopted the extraneous cognitive load explanation did not consider the competing germane cognitive load explanation any more. Accordingly, none of these studies tried to measure the three load types to really test the different cognitive load explanations. The review on the expertise reversal effect presented in this thesis should already be enough to make CLT researchers aware that high-knowledge learners with integrated formats might not necessarily suffer from too high extraneous cognitive load, but that there is a reasonable explanation suggesting that high-knowledge learners with separated formats engage in germane cognitive load. Researchers investigating the expertise reversal effect within the CLT framework should carefully investigate whether the germane cognitive load explanation might not be a reasonable option compared to the extraneous cognitive load explanation. This would imply that CLT researchers should also try to measure germane cognitive load. On the other hand, researchers favoring the germane cognitive load explanation should not forget to test extraneous cognitive load to provide further evidence on whether extraneous cognitive load is less important for the expertise reversal effect as suggested by the findings of this dissertation. Endeavors in disentangling the different load types and testing the different cognitive load explanations are necessary because whether teachers are told that their advanced students are cognitively overwhelmed by well guided instructions or whether they are cognitively underchallenged (cf. Nückles et al., 2009) has probably very different effects on the communication between teachers and students, and thus, on how teachers motivate their students.

8.1.2 Statistical Methods in Testing Cognitive Load Explanations

Besides the contribution of investigating the different explanations of the split-attention and expertise reversal effect, this dissertation also improved the testing of the explanations by applying more appropriate statistical methods. Although about half of the studies that were reviewed provided a cognitive load measure that was postulated to mediate the instructional design effect on learning outcomes, none of these studies used mediation analyses to test this hypothesis. This dissertation is the first one that tested competing assumptions of the split attention and expertise reversal effect critically by applying mediation analyses. Most of the former instructional design research has not gone beyond claims about the cognitive system because often cognitive load was not measured but if so, it was not analyzed with respect to the meditational role that is assigned to it. Thus, empirical evidence supporting the cognitive load explanations was often not presented, because the prevailing statistical procedures used in investigating both effects were not suited to test the mediating role of cognitive load. That is, cognitive load has been postulated to mediate the split-attention and expertise reversal effect, but even if perceived difficulty was measured, its mediating role was not tested. To summarize, this dissertation applied more appropriate statistical methods in testing whether one or more cognitive load types transmitted the effect of instructional format on learning outcomes. By using this procedure the behavioristic black box was filled not only with mere assumptions but also with empirical data.

8.1.3 Cognitive Load and Working Memory Models

Although the testing of the different cognitive load assumptions was very strict, this dissertation did not investigate whether the cognitive load explanations really fit to the working memory models described by researchers of CLT and the CTML. The cognitive architecture assumed within the CLT as well as within the CTML was described at the beginning of this dissertation to provide the basis of the different explanations of the cognitive load mechanisms. It was pointed out that CLT differs from the CTML by abandoning a cognitive processor like a central executive as assumed in Baddeley's (1996) multi-component model of working memory. CLT assumes that prior knowledge functions as a central executive (cf. Ericsson & Delaney, 1999). Consequently, no working memory resources should be expended during learning for the selection of information according to CLT, whereas learning processes including or mainly basing on learners' prior knowledge should cause germane cognitive load within the framework of the CTML by needing the central executive for these processes. However, neither CLT nor the CTML nor this dissertation has elaborated this difference in the models of the cognitive architectures concerning the cognitive load explanations underlying the expertise reversal effect. Although a germane cognitive load explanation of the expertise reversal effect seems to be reasonable according to the cognitive architecture suggested by the CTML, it seems less clear according to the one presented by CLT. On the other hand, the argument that prior knowledge might lead to redundant processing and thereby to extraneous cognitive load makes more sense within the framework of CLT that assumes that either prior knowledge does not cause cognitive load or, in a poor constellation of instructional format and prior knowledge, causes extraneous cognitive load. Although the methods applied in this dissertation to test the competing cognitive load explanations of the expertise reversal effect seem to be sufficient to differentiate between them, they did not test whether the model of the cognitive architecture suggested by CLT or by the CTML is more appropriate. However, a thorough concept and understanding of the cognitive architecture and its processes are important to fully comprehend a mechanism underlying an instructional design effect. Although this dissertation did not test the models of the cognitive architecture, the results demonstrated that prior knowledge should be considered more early in the architecture and cognitive processing, at least in the CTML, where prior knowledge seems to get first relevant in the later stages of building an integrated model. However, the eye tracking results suggest that prior knowledge is already important in the early stage of information selection. Hence, further research that investigates more basic aspects of the working memory models suggested by CLT and the CTML is needed to draw conclusions concerning the cognitive architecture. Fletcher and Tobias (2005) even demand that both theoretical frameworks have to become more precise concerning the relation between processes of knowledge integration and cognitive load on a theoretical level for further research.

8.2 Cognitive Load Measurement

Whereas the former sections dealt with issues concerning particularly the theoretical assumptions underlying the split-attention and expertise reversal effect, the following sections will concentrate on the measurement issues of cognitive load as the central construct in explaining both effects. First, the combination of different cognitive load or behavioral data will be discussed. Afterwards, it is discussed whether the theoretical constructs of intrinsic, extraneous, and germane cognitive load as assumed by CLT were measured. Moreover, the issues of motivation as well as working memory in relation with subjective ratings will be considered. Finally, future techniques of cognitive load measurement will be discussed.

8.2.1 Applying Different Techniques of Cognitive Load Measurement

In both experiments of this dissertation a combination of objective on-line methods and subjective post-hoc rating scales were used (cf. Brünken et al., 2003). In contrast to former studies that mostly used either only one subjective rating scale or studying times, this thesis combined rather complex measures. In Experiment 1 on the split-attention effect, a direct and objective secondary task performance indicating overall cognitive load was combined with direct subjective ratings scales that attempted to measure intrinsic, extraneous, and germane cognitive load individually. The attempt to measure the three load types individually was hardly used before (cf. Gerjets et al., 2006). In Experiment 2 about the expertise reversal effect, indirect objective viewing behavior was combined with the three direct subjective rating scales already used in Experiment 1. Because each method has its strengths and limitations (see chapter 3) a combination of two measurement techniques enabled more substantiated insights into the mechanisms underlying the instructional design effect in question.

In Experiment 1 it was demonstrated that learners with separated format had better secondary task performance (shorter reaction times) than learners with integrated format. This result is not compatible with the extraneous cognitive load explanation but it is compatible with the germane cognitive load explanation. However, it was only demonstrated for one graphic type and only for one of the secondary task stimuli. The other groups did not differ. Therefore, this result only serves as first evidence of the germane cognitive load explanation. However, the differences in the subjective ratings scales substantiated this interpretation by demonstrating that the extraneous cognitive load ratings as well as the germane cognitive load ratings mediated the split-attention effect. By showing that the subjective rating scales more or less fulfilled the criteria of the three load types (cf. Gerjets et al., 2009) a measurement technique was found that could also be used to investigate how differently knowledgeable learners process and benefit from integrated and separated formats. Experiment 2 investigated the cognitive load explanations of the expertise reversal effect concerning spatial contiguity. Because secondary task performance provides objective measures but reveals nothing about the different cognitive load types or learning strategies, learners' viewing behavior was measured in Experiment 2 instead of secondary task performance. Viewing behavior was regarded as a promising method to disambiguate the different explanations, especially in combination with subjective ratings and learning outcomes. The subjective ratings of extraneous cognitive load did not support the extraneous cognitive load explanation suggested by Kalyuga et al. (2003). Rather, the germane cognitive load ratings showed a trend towards the germane cognitive load explanation suggested by Mayer et al. (1995). Furthermore, it was demonstrated that more extensive text processing behavior (text-text switching) by low-knowledge participants with separated format was related to higher ratings of extraneous but also with lower ratings of germane cognitive load as suggested by Mayer et al. (1995) as well as with lower learning outcomes. The opposite result pattern was demonstrated for participants with high prior knowledge, a result which is also in accordance with the germane cognitive load explanation. Hence, measures of viewing behavior like the switching between different text units provided further insights by showing their relation to the subjective ratings of cognitive load and learning outcomes. Both studies benefited particularly from combining an objective method with the three subjective rating scales. Therefore, I recommend that further instructional design studies that want to gain insights into how cognitive load influences learning outcomes also use complementary measurement techniques.

8.2.2 Subjective Ratings and Motivation

The subjective ratings scales developed in this thesis are characterized by high face validity with respect to the theoretical constructs of intrinsic, extraneous, and germane cognitive load. Nevertheless, subjective ratings of difficulty and concentration may be influenced by motivational aspects that were not considered in this dissertaton. For example, with regard to the germane cognitive load item that asked for the level of concentration one might ask whether the scale represents cognitive aspects influenced only by the instructional format and measurement error or whether the ratings are also influenced by motivational factors. In research on motivation, the level of persons' concentration during a task is often considered to be a cognitive consequence of their intrinsic motivation (e.g., Csikszentmihalyi & Nakamura, 1989; Deci & Ryan, 1985; Guay, Vallerand, & Blanchard, 2000). If this would be true for the results of the presented studies, one might conclude that spatial contiguity influenced participants' intrinsic motivation, which in turn influenced their perception of material difficulty and concentration and cognitive processing. Although learners' ratings of situational interest were not highly related with ratings of concentration level, they were related with difficulty ratings of the materials thereby suggesting that subjective cognitive load ratings seem to be influenced by motivational factors.

Besides intrinsic motivation, further motivational factors like fear of failure and hope of success as two aspects of need for achievement (Rheinberg, 2008) and/or self-efficacy as an aspect of executing control (Bandura, 1998) might also influence subjective ratings. Learners with more hope of success and higher self-efficacy might rate the difficulty of the content and of the material lower than learners with more fear of failure. The level of concentration might also be influenced by these motivational factors. Learners with great fear of failure and test anxiety might invest less concentration or germane processes, because they might be distracted by frightening thoughts about failure. Alternatively, they just rate their level of concentration to be lower, in order to construct an excuse before hand in case they fail in the test situation (Elliot & Harackiewicz, 1996). Hence, whether and how such motivational aspects influence the subjective ratings of items, considered to measure the three cognitive load types, should be investigated more thoroughly in future studies which concentrate on subjective ratings of load type measurement. The knowledge about the different load types mediating an instructional design effect should be as comprehensive as possible. Both studies presented in this thesis suggest that extraneous cognitive load is not the only and maybe not the main cause of the split-attention and the expertise reversal effect as postulated by CLT researchers (e.g., Kalyuga, 2007; Sweller, in press). This finding should be considered when motivational aspects are investigated by testing whether successful formats first motivate learners before the specific cognitive processes are executed (cf. Salomon, 1984). If this would be the case, instructional formats would not directly or at least not solely influence cognitive processes as assumed by researchers of the CTML and CLT, but would first influence learners' motivation, which would in turn influence cognitive processes.

Although there is an increasing demand for considering and investigating motivational issues in relation to the three load constructs of cognitive load (Paas et al., 2005; Schnotz et al., 2009), there is only little research conducted so far which investigates the relation between aspects of cognitive load types, subjective ratings of cognitive load types and motivational aspects. An exception is the approach of investigating intrinsic and extrinsic motivation and learners' cognitive load during multimedia instructions (Zander & Brünken, 2009). Endeavors in investigating motivational aspects should be taken into consideration when investigating subjective ratings of cognitive load and motivational factors. Although this thesis measured situational interest, further motivational aspects like motivation types should be taken into account in future research. The influence of motivation on learning outcomes but also on subjective cognitive load ratings may be investigated more thoroughly.

8.2.3 Subjective Ratings and Working Memory Processes

Despite the predictive validity of the extraneous and germane cognitive load scales, I do not claim that those scales measured extraneous and germane cognitive load in the sense of separate quantitative amounts that fill a limited cognitive container (as suggested by CLT). Rather, the subjective rating scales seemed to measures variables that were related to relevant learning processes as partly substantiated by the measure of participants' viewing behavior. Whether or how these variables use working memory capacity is still debatable. For instance, difficulty ratings of the learning materials might indicate higher cognitive demands because of the need to repeat or reconstruct information during finding corresponding information as suggested by the correlations between extraneous cognitive load and non-corresponding text-graphic switching density of low-knowledge participants with separated format. However, whether these demands are executed or not is still unclear. The fact that low-knowledge learners with separated format reported lower levels of concentration and had lower learning outcomes might indicate that they did not execute these processes, and therefore, did not need extra working memory capacity. Hence, mental integration processes causing extraneous cognitive load as assumed by Sweller and colleagues might not be exerted. On the other hand, it might be the case that participants really repeated or reconstructed but did nevertheless not complete these processes successfully. Both options seem to be responsible for the split-attention effect, thereby making it difficult and unnecessary to refer to a higher working memory load only. Mwangi and Sweller (1998) asked students to self-explain during solving word problems and found that students with separated format made more simple rereads without connecting the information to solutions. On the other side, students with integrated format linked their rereads to solution steps more often, indicating that they processed the material more deeply. Moreover, the study yielded that students with separated format produced more incorrect inferences, whereas students with integrated format produced more correct inferences (Mwangi & Sweller, 1998). This suggests that not the use of more working memory capacity per se leads to lower learning outcomes, but the fact that the related processes are partly not executed, and if processes like inferences are executed, they do not result in correct knowledge. Concerning Experiment 2 of this dissertation it was shown that low-knowledge participants with separated format tended to process text and graphic in more isolation. Hence, they might have lacked to invest their cognitive resources in the appropriate processes of integration, but they were not necessarily overloaded by trying to do so. Moreover, if learners can nevertheless integrate verbal and graphical information by using their prior knowledge, they might not even suffer from separated formats. In this way, a split-attention effect can change into an expertise reversal effect as suggested by authors arguing with the germane cognitive load explanation. In a similar way, the subjective ratings of level of concentration might either indicate the number of elaboration and integration processes as already pointed out and at the same time they might indicate the degree of inhibition of irrelevant thoughts during learning (cf. Engle, Kane, & Tuholski, 1999). The self-explanation results of Mwangi and Sweller (1998) showed that students with separated format made more comments about the problem itself and about their own problem-solving performance showing that they were aware of their difficulties to understand the problems. Such thoughts might distract learners from the learning content itself and might lead them to thoughts about failure which further hinders concentrated learning. Such thoughts might have happened in low-knowledge participants with separated format who switched a lot between different text units. However, a focused processing is needed for successful learning as discussed by Renkl and Atkinson (2007). Unfortunately, there were no correlations between the viewing behavior of participants with integrated format and their subjective ratings of their level of concentration to describe more comprehensively which processing supports their learning and is reflected in subjective ratings. However, low-knowledge participants had higher learning outcomes on complex facts, the more they switched between corresponding text-graphic information. This result supports the assumptions of instructional researchers, even though there were no correlations between behavioral activities and ratings of germane cognitive load. To be able to get further insights into the relation between subjective ratings and learning processes, future studies might use the thinking-aloud method (Ericcson & Simon, 1980, 1993). By applying the thinking-aloud method one might get further insights into processing strategies during integrated and separated text-graphic instructions. Besides the classical think-aloud methodology, cognitive load might be measured by more technically advanced techniques. These are presented in the following.

8.2.4 Future Methods in Cognitive Load Measurement

More and more researchers suggest that cognitive load measurement in instructional design research might benefit either from more elaborated eye tracking measures or from neuroscientific techniques like EEG or fMRI (e.g., Brünken et al., 2003, de Jong, 2009; Whelan, 2007). With regard to the eye tracking technique researchers hope that viewing behavior might reveal processes that are directly linked with cognitive load. Although such relations will probably not be simple, some results of this dissertation suggest that special reading or processing strategies seem to correlate with cognitive load (e.g., extensive text processing operationalized by the spatial switching density between text-text AOIs). By applying the whole potential of the eye tracking methodology, that is the combination of spatial information (where do learners look at) with temporal information (when do they switch and how long do they look at) one might reveal more and better relations between viewing behavior and cognitive load types. However, measures which integrate spatial and temporal information are still rare in multimedia research and definitely need a complementary approach with other cognitive load measures (Hyönö, 2009). This dissertation worked with rather complex measures by applying the spatial density of four different types of switches. However, the potential of even more complex measures like for example the Levenshtein distance or Markov chains that also consider the sequence of different information units (or AOIs) is still not discovered within multimedia learning and might be helpful in clustering different reading types. Diagnosing different reading types might help to identify successful reading strategies and their cognitive processes. Moreover, the eye tracking technique might be used to develop gaze contingent learning environments. That is, only if learners look at pre-defined information areas, new and more advanced information or special prompts might appear automatically. Such environments might help to reduce extraneous cognitive load and increase germane cognitive load.

With regard to neuroscientific techniques different measures might provide helpful insights into cognitive load. Whereas Antonenko (2006) used desynchronization percentage of EEG waves during hypertext learning, Whelan (2007) suggests using event-related fMRI studies to distinguish between different cognitive load types. According to Whelan (2007) intrinsic cognitive load might be localized in the dorsolateral prefrontal cortex according to studies by Banich et al. (2000) and Newman et al. (2002). According to other researchers using the EEG technique, neural efficiency might be an approach to measure intrinsic cognitive load (e.g., Grabner, Stern, & Neubauer, 2003; Grabner, Neubauer, & Stern, 2006). Furthermore, Whelan (2007) assumes that extraneous cognitive load might be tapped by "…architectural constraints in the brain that modulate attention across sensory modalities" (p. 8) according to considerations by Meredith (2001). Concerning the measurement of germane cognitive load, however, Whelan (2007) stays rather skeptical. Nevertheless, there are results suggesting that alpha and theta oscillations might be related with memory processes (e.g., Klimesch, Schack, & Sauseng, 2005). Although it is still unclear whether and which approach is more successful in finding consistent and interpretable brain-related measures, the EEG technique seems to be especially appealing, because this method might be combined with so called passive brain-computer interfaces (BCIs). Passive BCIs might be used to develop adaptive computer-based learning environments that automatically adapt to the cognitive load experienced by the learner and measured on-line via EEG. Concerning the expertise reversal effect, Kalyuga (2006) as well as Kalyuga and Sweller (2005) suggest to measure learners' knowledge several times during learning to adapt the learning environment to learners' changing knowledge level. Learning environments basing on passive BCIs might adapt automatically to changes in learners' knowledge level without disrupting the learning process, if for example, differences in neural efficiency might be detected by the system. Until such learning scenarios become reality, specific research into cognitive load measurement during learning with EEG is needed.

8.3 General Conclusion

This dissertation investigated the spatial contiguity between text and graphic in multimedia learning. The aim was to go one step beyond the mere comparison of learning outcomes from (differently knowledgeable) learners studying either with physically integrated or separated formats by investigating *how* the cognitive system of learners produces the split-attention and expertise reversal effect. The cognitive mechanisms assumed to underlie both effects referred especially either to extraneous or to germane cognitive load as mediating factor. Although extraneous and germane cognitive load are assumed to be differently important in causing the split-attention and expertise reversal effect, former research has not provided much evidence to clarify the role of each cognitive load type. Rather, there exist different explanations of both effects in literature without the evidence to decide which explanation is more appropriate. This dissertation is one of the first approaches trying to disentangle the influences of intrinsic, extraneous, and germane cognitive load on the split-attention and expertise reversal effect concerning spatial contiguity. The results of Experiment 1 and Experiment 2 seem to justify the conclusion that differences in simple surface characteristics of instructional formats, like the contiguity between text and graphic, influence learners' germane processing. Although Horz and Schnotz (2010) deny such relations, the differences in objective (Experiment 1: secondary task performance; Experiment 2: viewing behavior) and subjective cognitive load ratings (Experiment 1 and 2) are thought to be related not only with differences in extraneous but also with germane cognitive load.

Although CLT developed the construct of germane cognitive load about twelve years ago (Sweller et al., 1998), the focus on extraneous cognitive load and on inhibiting mechanisms seems to be still prevailing in CLT (Kalyuga, 2011). Therefore, it is suggested that CLT should also detail positive learning processes resulting in germane cognitive load. The more detailed the assumptions are, the better is the probability to measure cognitive load appropriately. Future research on instructional design characteristics like spatial contiguity between text and graphic and their influence on cognitive load should use the approach of triangulation. This dissertation showed that the combination of subjective and objective measures in relation with learning outcomes has provided empirical support and new insights into how learners cognitively process integrated and separated formats. Focusing on germane cognitive load and related behavioral activities might be especially relevant for instructors who want to support and challenge students instead of just making learning as cognitively "easy" as possible. The same holds to true for learners who want to engage optimally in order to learn successfully. To investigate how important the investment of germane cognitive load in comparison to reductions in extraneous cognitive load is with different instructional design characteristics beyond spatial contiguity, further and more detailed process assumptions as well as their measurement and statistical analyses are needed. Further research into the relations between instructional designs and cognitive load might help to develop learning environments that automatically adapt to learners' individual cognitive load by specifying it by means of gaze contingency displays (Duchowski, Courina, & Murphy, 2004) or/and passive brain-computer interface systems (Zander & Kothe, 2011).

9 References

References marked with one asterisk are part of the review on the explanation of the split-attention effect. References marked with two asterisks are part of the review on the explanation of the expertise reversal effect.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Ainsworth, S. E. (1999). A functional taxonomy of multiple representations. *Computers and Education, 33,* 131-152.

Ainsworth, S. E. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16,* 183-198.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*, 10-14.

Alexander, P. A., & Jetton, T. L. (1996). The role of importance and interest in the processing of text. *Educational Psychology Review, 8*, 89-121.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*, 375-404.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning and Individual Differences, 6*, 379-397.

**Amadieu, F., Tricot, A., & Mariné, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior, 25*, 381-388.

**Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning & Instruction*, 1-11.

Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition, 9*, 237-246.

Antonenko, P. D. (2008). *The effect of leads on cognitive load and learning in a conceptually rich hypertext environment.* Doctoral dissertation, Iowa State University.

Astleitner, H. & Wiesner, C. (2004). An integrated model of multimedia learning and motivation. *Journal of Educational Multimedia and Hypermedia,* 13, 3-21.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory. A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation*. New York, NY: Academic Press.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart & Winston.

Ayres, P. (2006a). Using subjective measures to detect variations of intrinsic load within problems. *Learning and Instruction, 16*, 389-400.

**Ayres, P. (2006b). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287-298.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 134-146). New York: Cambridge University Press.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A*(1), 5-28.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*, 417-423.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85-97.

Bandura, A. (1998). *Self-efficacy: The exercise of control*. New York, NY: Freeman and Company.

Banich, M., Milham, M., Atchley, R. A., Cohen, N. A., Webb, A., Wszalek, T., et al. (2000). The anterior cingular cortex and response competition: Evidence from an fMRI study of overt verb generation. *Cognitive Brain Research, 12*, 298-305.

Bannert, M. (2002). Managing cognitive load - recent trends in cognitive load theory. *Learning and Instruction, 12*, 139 - 146.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology, 33*, 570-585.

Bartlett, R. (1932). *Remembering*. Cambridge, MA: Harvard University Press.

Bernstein, M. R. (1955). Relationship between interest and reading comprehension. *Journal of Educational Researcher, 49*, 283-288.

Bindé, J. (2005). *UNESCO world report 1: Towards knowledge societies.* UNESCO Publishing.

**Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated-interactive elements effect from an accounty class. *Instructional Science, 3,* 277-287.

*Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning & Instruction, 3*, 1-21.

**Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22,* 27-42.

*Bodemer, D., Plötzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*, 325-341.

Borg, G. (1978). Subjective aspects of physical and mental load. *Ergonomics, 21*, 215-220.

Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology, 12*, 249-260.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A*, 525-545.

Bracht, G. (1970). Experimental factors related to aptitude–treatment interactions. *Review of Educational Research, 40*, 627-645.

Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior, 21*, 421-436.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*, 115-132.

Brünken, R., Steinbacher, S., & Leutner, D. (2000). Pädagogisch-psychologische Grundlagenforschung. In D. Leutner & R. Brünken (Eds.), *Neue Medien in Unterricht, Aus- und Weiterbildung* (pp. 37-46). Münster: Waxmann.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*(2), 109-119.

**Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior, 19*, 135-145.

Camp, G., Paas, F., Rikers, R. M. J. P., & Van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control traning: A comparison between performance, mental effort, and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Canham, M., & Hegarty, M. (2009). Effects of knowledge and display design on comprehension of complex graphics. *Learning & Instruction*. Advanced online publication. doi: 10.1016/j.learninstruc.2009.02.014.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology, 95*, 629-640.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*, 5-26.

Carroll, P. J., Young, R. J., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York, NY: Springer.

*Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15*, 345-367.

*Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

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

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151-170.

Charness, N., Reingold, E., Pomplun, M., & Stampe, D. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29*, 1146-1152.

Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NY: Erlbaum.

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.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*, 21-30.

**Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research & Development, 53*, 15-24.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Conway, M. A., Gardiner, J. M., Perfect, T. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126,* 393–413.

Cook, A. E., Zheng, R. Z., & Blaz, J. W. (2009). Measurement of cognitive load during multimedia learning activities. In R. Z. Zheng (Ed.), *Cognitive effects of multimedia learning* (pp. 34-50). Hershey, PA: IGI Global.

**Cooper, G., Tindall-Ford, S., Chandler, P., & Weller, J. (2001). Learning by imagining. *Journal of Experimental Psychology, 7*, 68-82.

Corbalan, G., Kester, L., & Van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaption and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*, 733-756.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York, NY: Irvington Publishers.

Csikszentmihalyi, M., & Nakamura, J. (1989). The dynamics of intrinsic motivation: A study of adolescents. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol.3, pp. 45-71). New York, NY: Academic Press.

Darley, C. F., Klatzky, R. L. & Atkinson, R. C. (1972). Effects of memory load on reaction time. *Journal of Experimental Psychology, 96*, 232-234.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design; Some food for thought. *Instructional Science, 38,* 105-134.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York, NY: Plenum Publishing.

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.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827-1837.

DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*, 189-228.

Dochy, F., Segers, M. & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145-186.

Duchowski, A. T., Courina, N., & Murphy, H. (2004). Gaze-Contingent Displays: A review. *Cyber Psychology & Behavior, 7,* 621-634.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science, 18*, 135-140.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*, 461-475.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York, NY, US: Cambridge University Press.

*Erhel, S., & Jamet, E. (2006). Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning, 22*, 137-147.

Ericsson, A. K., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Ericsson, A. K., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MS: MIT Press.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake, & P. Shah (Ed.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257-297). New York, NY, US: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Eysenck, M. W. (1982). *Attention and arousal: Cognition and performance*. Heidelberg: Springer.

Eysenck, M. W. (2006). *Fundamentals of cognition*. New York, NY, US: Psychology Press.

Fischer, S., Lowe, R. K., & Schwan, S. (2008). Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology, 22*, 1126-1141.

Fisk, A. D., Derrick, W. L., & Schneider, W. (1986). A methodological assessment and evaluation of dual-task paradigms. *Current Psychology, 5*, 315-327.

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

*Florax, M., & Ploetzner, R. (2009). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning & Instruction*. Advanced online publication. doi: 10.1016/j.learninstruc.2009. 02.021.

Folker, S., Ritter, H., & Sichelschmidt, L. (2005). Processing and integrating multimodal material - the influence of color-coding. In B. G. Bara, L. Barslou & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 690-695). Stresa, Italy: Cognitive Science Society.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology*.

Gerjets, P., Cierniak, G., & Scheiter, K. (2008, October). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Workshop "Cognitive Load Theory and its Application to Instructional Design", University of Macau.* Macau, China.

Gerjets, P., & Hesse, F. W. (2004). When are powerful learning environments effective? The role of learner activities and of students´ conceptions of educational technology. *International Journal of Educational Research, 41*, 445-465.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*, 33-41.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Reducing cognitive load and fostering cognitive skill acquisition: Benefits of category-avoiding examples. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 450-455). Mahwah, NY: Erlbaum.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning & Instruction, 16*, 104-121.

Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review, 21*, 43-54.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*, 360-370.

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

**Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*, 229-251.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology, 19*, 183-204.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*, 631-645.

Golenhofen, K. (1997). Wasserhaushalt und Nierenfunktion. In K. Golenhofen (Ed.), *Physiologie* (pp. 377-399). München: Urban & Schwarzenberg.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*, 519-532.

Gopher, D., Chillag, N., & Arzi, N. (1985). *The psychophysics of workload - A second look at the relationship between subjective measures and performance.* Paper presented at the 29th Annual Meeting of the Human Factors Society, Baltimore.

Grabner, R. H., Neubauer, A. C., & Stern, E. (2006). Superior performance and neural efficiency: The impact of intelligence and expertise. *Brain Research Bulletin, 69*, 422-439.

Grabner, R. H., Stern, E. & Neubauer, A. C. (2003). When intelligence loses its impact: Neural efficiency during reasoning in a familiar area. *International Journal of Psychophysiology, 49*, 89-98.

Granit, R. A. (1977). *The purposive brain*. New York, NY: MIT Press.

Greeno, J. G. (1994). Gibbson's affordances. *Psychological Review*, *101*, 336-342.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*, 5-26.

Guay, F., Vallerand, R., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (SIMS)1. *Motivation and Emotion, 24*, 175-213.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 172-190.

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.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*, 92-102.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North Holland.

He, E., & Kowler, E. (1992). The role of saccades in the perception of texture patterns. *Vision Research, 32*, 2151-2163.

Heckhausen, H., & Heckhausen, J. (2008). *Motivation and action*. Cambridge: University Press.

Hegarty, M. (1992a). The mechanics of comprehension and comprehensions of mechanics. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 428-443). New York, NY: Springer.

Hegarty, M. (1992b). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1084-1102.

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

Hick, C., & Hick, A. (2000). Nierenfunktion, Wasser- und Elektrolythaushalt. In C. Hick & A. Hick (Eds.), *Kurzlehrbuch Physiologie* (3rd ed., pp. 199-228). München: Urban & Fischer.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*, 549-571.

Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, Emotion, and Cognition*.

Hogg, N. M. (2007). Measuring cognitive load. In R. A. Reynolds, R. Woods & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 186-192). Hershey, PA: IGI Global.

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

*Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attention guidance. *Applied Cognitive Psychology, 23*, 1215-1226.

**Homer, B. D., & Plass, J. (2010). Expertise reversal for iconic representations in science visualizations. *Instructional Science, 3,* 259-276.

Horz, H. & Schnotz, W. (2010). Cognitive Load in learning with multiple representations. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory.* Cambridge: Cambridge University Press.

Howe, M. L., & Rabinowitz, F. M. (1989). On the uninterpretability of dual-task performance. *Journal of Experimental Child Psychology, 47*, 32-38.

Huppelsberg, J., & Walter, K. (2003). Wasser- und Elektrolythaushalt, Nierenfunktion. In J. Huppelsberg & K. Walter (Eds.), *Kurzlehrbuch Physiologie* (pp. 178-201). Stuttgart: Georg Thieme Verlag.

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

Inamdar, S., & Pomplun, M. (2003). Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks. In R. Alterman & D. Kirsh (Eds.), *25th Annual Conference of the Cognitive Science Society*. Boston, Massachusetts: Cognitive Science Society.

Jäger, A. O., Süß, H.-M., & Beauducel. (1997). *Berliner Intelligenzstruktur-Test, BIS-Test, Form 4*. Göttingen: Hogrefe.

James, L. R., & Brett, J. M. (1984). Mediators, moderators and test for mediation. *Journal of Applied Psychology, 69*, 307-321.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-355.

*Kablan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education, 51*, 660-668.

Kahneman, D. (1973). *Attention and effort.* New Jersey: Prentice-Hall.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.

Kalyuga, S. (2007). Expertise-reversal effect and its implications for learner-tailored instruction. *Educational Psychological Review, 19*, 509-539.

**Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior, 24*, 852-861.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*, 1-19.

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

**Kalyuga, S., Chandler, P. & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.

*Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

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

**Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5-23.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.

Kalyuga, S., & Renkl, A. (2009). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*. Advanced online publication. doi 10.1007/s11251-009-9102-0.

**Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*, 83-93.

Kenny, D. A. (2006). Mediation. Retrieved July 13, 2007, from http://davidakenny.net/cm/mediate.htm.

Kerr, B. (1973). Processing demands during mental operations. *Memory & Cognition, 1*, 401-412.

*Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology, 75*, 71-85.

Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving a troubleshooting skill. *Contemporary Educational Psychology, 31*, 167-185.

Kester, L., Lehnen, C., Van Gerven, P. W. M., & Kirschner, P. A. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior, 22*, 93-112.

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics, 9*, 87-98.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kirschner, P. A., Sweller, P. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experimental, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations. *Experimental Psychology, 52,* 1-10.

**Lambiotte, J. G., & Dansereau, D. F. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *Journal of Experimental Education, 60*, 189-201.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Leahy, W., & Sweller, J. (2004). Cognitive load and the imagination effect. *Applied Cognitive Psychology, 18*, 857-875.

**Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology, 11,* 266-276.

**Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*, 902-913.

Leigh, J. H., & Kinnear, T. C. (1980). On interaction classification. *Educational and Psychological Measurement, 40*, 841-844.

Levin, J.R. , Anglin, G.J., & Carney, R.N. (1987): On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.). *The psychology of illustration. Vol. I Basic Research* (S. 51 – 85) New York, NY: Springer.

Lowe, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. 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.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 89*, 49-63.

*Martin-Michiellot, S., & Mendelsohn, P. (2000). Cognitive load while learning with a graphical computer interface. *Journal of Computer Assisted Learning, 16*, 284-293.

*Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240-246.

Mayer, R. E. (1993). Illustrations that instruct. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 4, pp. 253-284). Hillsdale, NY: Erlbaum.

Mayer, R. E. (1997). Multimedia Learning: Are we asking the right questions? *Educational Psychologist, 32*, 1-19.

Mayer, R. E. (2001). *Multimedia learning* (1st ed.). New York, NY: Cambridge University Press.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*, 125-139.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14-19.

Mayer, R. E. (2005a). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2005b). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). New York, NY: Cambridge University Press.

Mayer, R. E. (2005c). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.

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

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning & Instruction, 20,* 167-171.

**Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology, 11*, 256-265.

Mayer, R. E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187-198.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638-643.

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.

**Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design. Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*, 31-43.

**McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62.

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

**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.

McNamara, D. S. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51). New York, NY: Elsevier Science.

Meredith, M. A. (2001). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research, 14*, 31-40.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review, 63*, 81-97.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309-326.

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

Morris, C. J., & Kimbrell, G. (1972). Performance and attitudinal effects of the Keller method in an introductory psychology course. *Psychological Record, 22*, 523-530.

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.

*Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition & Instruction, 16*, 173-199.

Newman, S. D., Just, M. A., & Carpenter, P. A. (2002). The synchronization of the human cortical working memory network. *NeuroImage, 15*, 31-40.

**Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science*, *38,* 237-258.

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

**Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespeareans text. *Instructional Science, 38,* 217-236.

**Ollerenshaw, A., Aidman, E., & Kidd, G. (1997). Is an illustration always worth ten thousand words? Effects of prior knowledge, learning style and multimedia illustrations on text comprehension. *International Journal of Instructional Media, 24*, 227-238.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121-152.

*Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology, 28*, 29-45.

Ozcelik, E., Karakus, T., Kursun, E., & Kagiltay, K. (2009). An eye-tracking study on how color coding affects multimedia learning. *Computers & Education, 53*, 445-453.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning & Instruction, 19*, 228-242.

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.

Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F., Tuovinen, J. E., Van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research & Development, 53*, 25-34.

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning & Instruction, 16*, 87-91.

Paas, F., Van Gerven, P. W. M., & Tabbers, H. K. (2005). The cognitive aging principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & Van Merrienboër, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor skills, 79*, 419-430.

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

**Pawley, D., Ayres, P., Cooper, M., & Sweller, J. (2005). Translating words into equations: a cognitive load theory approach. *Educational Psychology, 25*, 75-97.

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.

*Pociask, F. D., & Morrsison, G., R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development, 56*, 379-399.

**Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61-86.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

**Potelle, H., & Rouet, J.-F. (2003). Effects of content representation and readers' prior knowledge on the comprehension of hypertext. *International Journal of Human-Computer Studies, 58*, 327-345.

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

Preacher, K. J., & Leonardelli, G. (2007). *Calculation for the Sobel test: An interactive calculation tool for mediation tests.* Retrieved December 16, 2007, from http://www.people.ku.edu/~preacher/sobel/sobel.htm.

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

*Purnell, K. N., Solman, R. T., & Sweller, J. (1991). The effects of technical illustrations on cognitive load. *Instructional Science, 20*, 443-462.

Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Pschological Bulletin, 124*, 372-422.

Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye movements when looking at print advertisements: The goal of the viewer matters. *Applied Cognitive Psychology, 22*, 697-707.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*, 219-226.

Reed, W. M., Burton, J. K., & Kelly, P. (1985). The effects of writing ability and mode of discourse on cognitive capacity engagement. *Research in the Teaching of English, 19,* 283-297.

**Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning & Instruction, 16*, 92-103.

Renkl, A., & Atkinson, R. K. (2007). Interactive learning environments: Contemporary issues and trends. An introduction to the special issue. *Educational Psychology Review, 19*, 235-238.

Rey, G. D. (2009). *E-Learning*. Bern: Huber.

Rheinberg, F. (2008). *Motivation* (6th ed.). Stuttgart: Kohlhammer.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen [FAM: A questionnaire to measure actual motivation in learning-and achievement situations]. *Diagnostica*, 47, 57-66.

*Rose, J. M. (2002). The effects of cognitive load on decision aid users. *Advances in Accounting Behavioral Research, 5*, 115-140.

*Rose, J. M., & Wolfe, C. J. (2000). The effects of system design alternatives on the acquisition of tax knowledge from a computerized tax decision aid. *Accounting, Organizations and Society, 25*, 285-306.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, MJ: Erlbaum.

Rummer, R., Schweppe, J., Scheiter, K., & Gerjets, P. (2008). Lernen mit Multimedia: Die kognitiven Grundlagen des Modalitätseffekts. *Psychologische Rundschau, 59*, 98-107.

**Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 3,* 289-307.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist, 18*, 42-50.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributes. *Journal of Educational Psychology, 78*, 647-658.

Schatz, G. (2005). *Jeff's view on science and scientists.* Oxford: Elsevier Ltd.

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.

Schiefele, U. (1996). Topic interests, text representation, and quality of experience. *Contemporary Educational Psychology, 21*, 3-18.

Schmidt, R. F., & Thews, G. (1995). Funktionen der Nieren. In R. F. Schmidt & G. Thews (Eds.), *Physiologie des Menschen* (26. ed., pp. 737-777). Berlin: Springer.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2009a). A closer look at split visual attention in system-paced and self-paced multimedia instructions. *Learning and Instruction*. Advanced online publication. doi: 10.1016/j.learninstruc.2009.02.011.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2009b). Explaining the modality and contiguity effects: New insights from investigating students` viewing behavior. *Applied Cognitive Psychology*. Advanced online publication. doi: 10.1002/acp.1554.

Schnotz, W. (1997). Zeichensysteme und Wissenserwerb mit neuen Technologien. In H. Gruber & A. Renkl (Hrsg), Wege zum Können. Determinanten des Kompetenzerwerbs (S. 218-235). Bern: Hans Huber.

Schnotz, W. (2005). An integrated model of text and pictre comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. 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.

Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides & P. Nenninger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 69-96). Ashland, OH: Hogrefe & Huber.

Schnotz, W., & Kulhavy, R.W. (Eds.). (1994). *Comprehension of graphics*. Amsterdam: North-Holland.

Schnotz, W., & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychological Review, 19*, 469-508.

**Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technolgy Research & Development, 53*, 47-58.

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

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

**Seufert, T., Jänen, I., & Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Computers in Human Behavior, 23*, 1055-1071.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*, 159-189.

**Shapiro, A. M. (1999). The relationship between prior knowledge and interactive overviews during hypermedia-aided learning. *Journal of Educational Computing Research, 20,* 143–167.

**Shin, E. C., Schallert, D. L., & Savenye, W. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technolgy Research and Development, 42*, 33-46.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.

Sigmund, T. (1988). Teaching strategic text review by computer and interaction with student characteristics. *Computers in Human Behavior, 4*, 299-310.

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology, 59*, 205-216.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*, 347-376.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco: Jossey-Bass.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology, 45*, 1-8.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*, 9-31.

Sweller, J. (2005a). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York, NY: Cambridge University Press.

Sweller, J. (2005b). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159-167). New York, NY: Cambridge University Press.

Sweller, J. (2009). Cognitive bases of human creativity. *Educational Psychology Review, 21*, 11-19.

Sweller, J. (in press). Cognitive load theory: Theoretical underpinnings. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory.* New York, NY: Cambridge University Press.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition & Instruction, 8*, 351-362.

*Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition & Instruction, 12*, 185-233.

*Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176-192.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*, 434-458.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

*Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2000). *Multimedia instructions and cognitive load theory: Split-attention and modality effects.* Paper presented at the National Convention of the Association for Educational Communications and Technology, Long Beach, CA.

*Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*, 424-436.

*Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257-287.

Tobias, S. (1994). Intrest, prior knowlege, and learning. *Review of Educational Research, 64*, 37-54.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*, 133-152.

**Tuovinen, J. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91,* 334-341.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46*, 629-646.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology, 57A*, 165-182.

Van Gog, T., 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.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the Original Construct in Educational Research. *Educational Psychologist, 43*, 16-26.

**Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning & Instruction, 18*, 211-222.

Van Gog, T., & Scheiter, K. (2009). Eye tracking as a tool to study and enhance multimedia learning. *Learning & Instruction. 20*, 95-99.

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: Routledge & Kegan Paul.

*Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*, 1-39.

Wedel, M., Pieters, R., & Liechty, J. (2008). Attention switching during scene perception: How goals influence the time course of eye movements across advertisement. *Journal of Experimental Psychology: Applied, 14*, 129-138.

Welch, J. (1898). On the measurement of mental activity through muscular activity and the determination of a constant of attention. *American Journal of Psychology, 1*, 288-306.

Whelan, R. R. (2006). *The multimedia mind: Measuring cognitive load in multimedia learning.* Doctoral dissertation, New York University, New York.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*, 1-12.

White, A. R. (1993). *Training end users on database queries: the effectiveness of example format and spatial ability on learning outcomes.* Doctoral dissertation, University of California, Los Angeles.

Wickens, C. D. (1991). Processing resources and attention. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 3-34). London: Taylor & Francis.

Windell, D., & Wiebe, E. N. (2007). *Measuring cognitive load in multimedia instruction: A comparison of two instruments*. Paper presented at the Annual meeting of the American Educational Research Association.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376.

Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2010). Observational learning from animated models: Effects of studying-practicing alternation and illusion of control on transfer. *Instructional Science, 38,* 89-104.

Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics, 4,* 213-242.

Yarbus, A. L (1967). *Eye movements and vision* (B. Haigh, Trans.). New York, NY: Plenum Press. (Original work published 1965).

**Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*, 1-21.

Zander, S. & Brünken, R. (2009, August). Do intrinsic and extrinsic aspect of motivation affect cognitive load in different ways? 13th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI). Amsterdam, The Netherlands.

Zander, T., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering, 8, doi*: 10.1088/1741-2560/8/2/025005.

Zijlstra, F. R. H., & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort*. Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

10 APPENDIX

The test materials used in Experiment 1 and 2 are provided in the following.

10.1 Knowledge Tests

The four tests on terminology, labelling, complex facts and transfer were used in Experiment 1 and 2 to measure participants' prior knowledge as well as learning outcomes after instructions.

10.1.1 Terminology Test

Table 32 shows the test items {with correctness of item} of the terminology test.

Table 32

Items of the terminology test

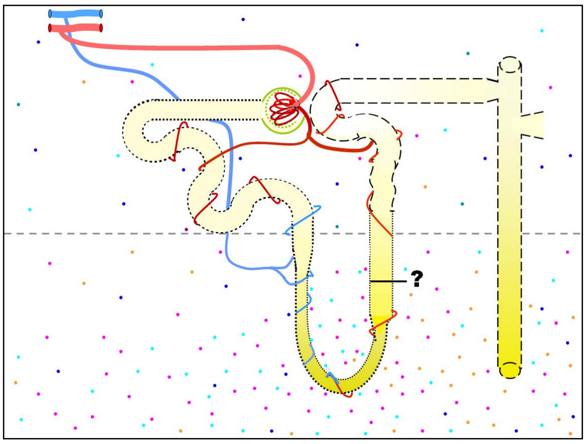
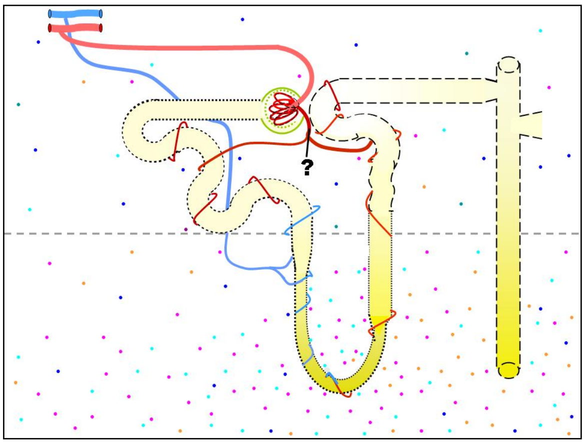
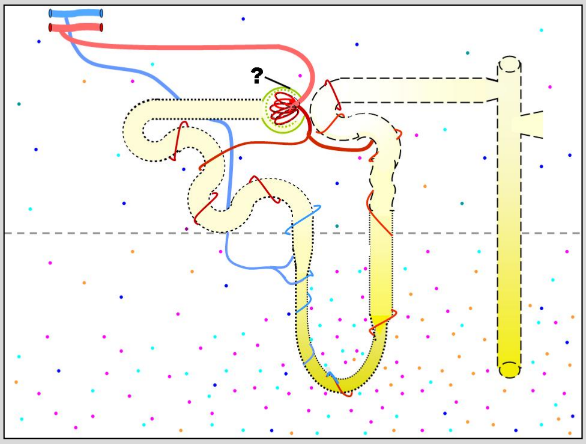
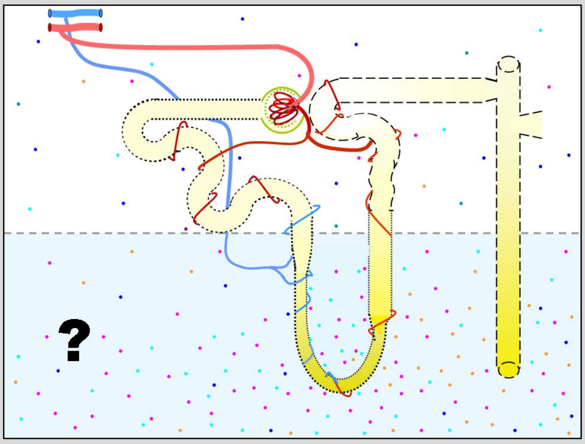
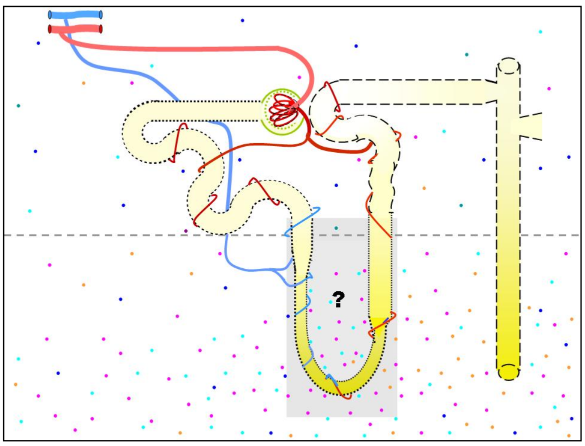
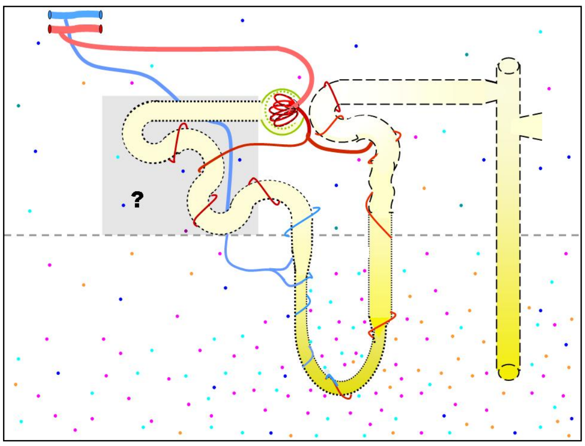
Welcher der folgenden Begriffe bezeichnet eine Struktur, die Bestandteil eines Nephrons oder seiner Umgebung ist? {0 = falsch, 1 = richtig}			
Globulus {0}	Glomulus {0}	Glomerulus {1}	Glemerulus {0}
Interstitium {1}	Interitum {0}	Intrastitium {0}	Interstutium {0}
Vas affernes {0}	Vas afferens {1}	Vas afferrentis {0}	Vas affernis {0}
Harnrohr {0}	Sammelrohr {1}	Harnröhre {0}	Sammelkanal {0}
Bowman-Kapsel {1}	Bowl-Kapsel {0}	Bovvmann-Kapsel {0}	Blowman-Kapsel {0}
distalis Konvultis {0}	distales Konvultum {0}	distales Konvulut {0}	distales Konvult {1}
Helen-Schleife {0}	Hähnle-Schleife {0}	Henle-Schleife {1}	Heinle-Schleife {0}
Tubulus renientis {0}	Tubulus reunionis {0}	Tubulus reuniens {1}	Tubulus reniens {0}
Vas effernis {0}	Vas efferens {1}	Vas efferentis {0}	Vas effernes {0}

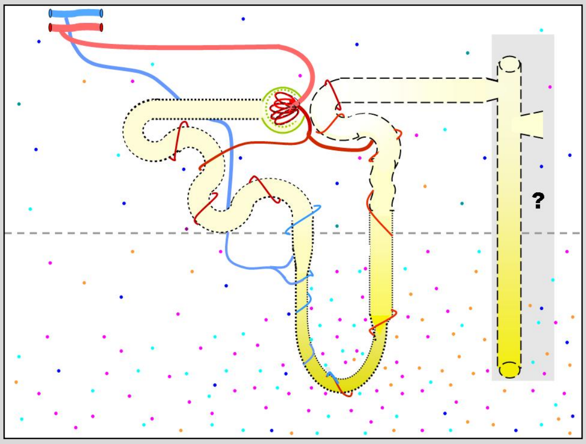
10.1.2 Labeling Test

Table 33 shows the graphics of the labelling test with the correct answer.

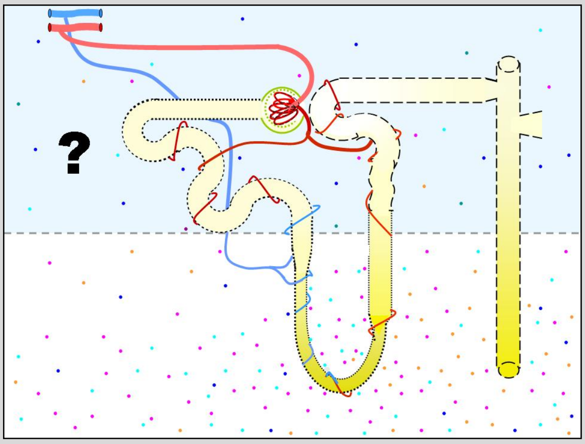
Table 33

Items of the labeling test

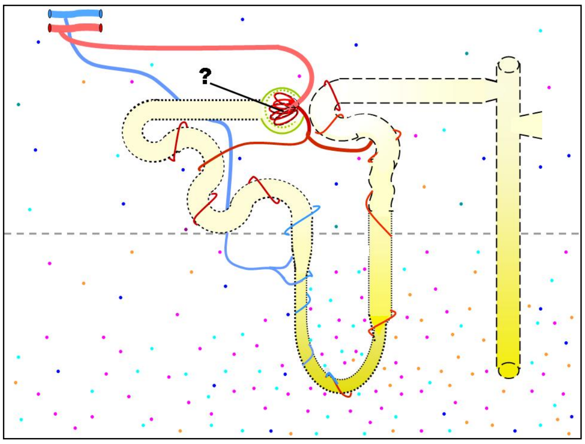
Auf welche Struktur deutet der Strich? / Welche Struktur ist hier Grau/Blau hinterlegt?	
 Aufsteigender Schenkel	 Vas efferens
 Bowman-Kapsel	 Nierenmark
 Henle-Schleife	 Proximales Konvult



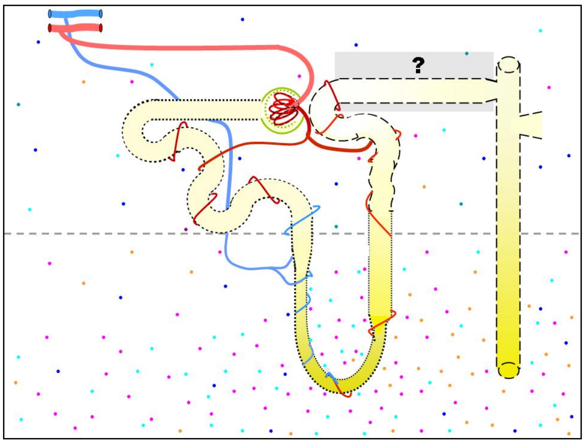
Sammelrohr



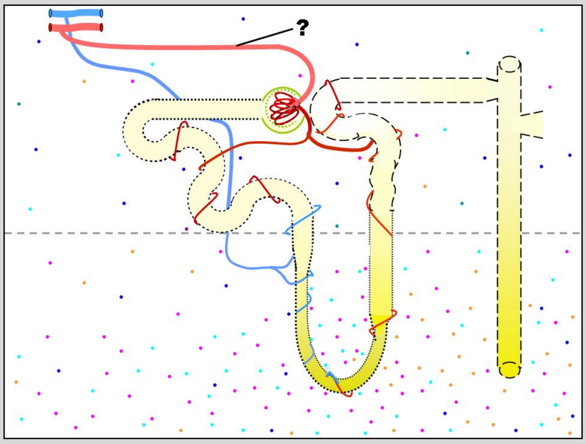
Nierenrinde



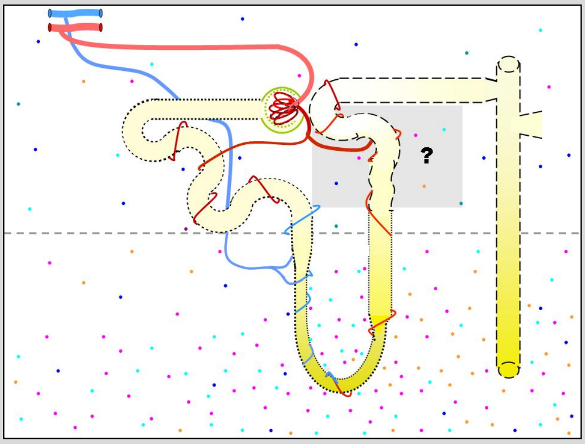
Glomerulus



Tubulus reuniens



Vas afferens



Distales Konvult

10.1.3 Complex Facts Test

Table 34 shows the items of the test on complex facts.

Table 34

Items of the complex facts test

0 = falsch **1 = richtig**	**Welche der folgenden Aussagen ist richtig?**
1	1. Das Vas efferens leitet Blut aus dem Glomerulus in die Kapillaren weiter.
0	2. Das Hormon Adiuretin öffnet Wasserkanäle und sorgt somit im absteigenden Schenkel für einen erhöhten Wasserausstrom.
0	3. Das Hormon Aldosteron senkt die Wasserdurchlässigkeit der Sammelrohrmembran.
1	4. Der ins Nierenbecken abfließende Harn kann maximal die Osmolarität des Interstititums im Nierenmark annehmen.
1	5. Das Hormon Adiuretin öffnet im Sammelrohr und Tubulus reuniens Wasserkanäle.
1	6. Die Membran des aufsteigenden Schenkels ist impermeabel für Wasser.
1	7. Noch ca. 35% des Filtrats gelangen aus dem proximalen Konvult in den absteigenden Schenkel der Henle-Schleife.
1	8. Harnstoff trägt zu ca. 50% zur Aufrechterhaltung der Hyperosmolarität im Interstitium des Nierenmarks bei.
1	9. Die Membran des distalen Konvults ist weitestgehend undurchlässig für Elektrolyte.
1	10. Die Membran des distalen Konvults ist für Harnstoff und Wasser undurchlässig.
0	11. Dem Elektrolytausstrom im proximalen Konvult folgt aus elektrochemischen Gründen ein Wasserausstrom.
1	12. Die Harnkonzentration nimmt im distalen Konvult bis zum Tubulus reuniens kontinuierlich ab.
0	13. Dem aktiven Natriumtransport im proximalen und distalen Konvult sowie im aufsteigenden Schenkel folgt jeweils ein passiver Chloridausstrom.
0	14. Im absteigenden Schenkel der Henle-Schleife findet ein erhöhter Natriumausstrom statt.
0	15. Der primär aktive Natriumtransport im proximalen Konvult ist der Motor zur Harnkonzentrierung.
1	16. Der Hauptanteil (ca. 65%) der Rückresorption lebenswichtiger Stoffe findet im proximalen Konvult statt.
0	17. Im absteigenden Schenkel sinkt die Harnkonzentration kontinuierlich bis zur Spitze der Henle-Schleife.

0	18. Im distalen Konvult ermöglicht das Hormon Aldosteron den Wasserausstrom ins Interstitium.
0	19. Die Membran des aufsteigenden Schenkels ist impermeabel für Chlorid.
1	20. Der Harn im proximalen Konvult hat die gleiche Konzentration wie das Interstitium der Nierenrinde.
0	21. Harnstoff diffundiert entlang seines Konzentrationsgradienten aus der Henle-Schleife zurück in das Sammelrohr im Nierenmark.
0	22. Eine Folge des lumennegativen Potentials, das durch den primär aktiven Natriumtransport im proximalen Konvult entsteht, ist der Kaliumausstrom ins Interstitium.

10.1.4 Transfer Test

Table 35 shows the items of the transfer test.

Table 35

Items of the transfer test

0 = falsch **1 = richtig**	**Welche der folgenden Aussagen ist richtig?**
0	1. Werden in einer Urinprobe Eiweiße gefunden, so ist ein Defekt des Vas efferens sehr wahrscheinlich.
1	2. Wird der erhöhte Natriumtransport im aufsteigenden Schenkel künstlich zum Erliegen gebracht, folgt eine vermehrte Urinausscheidung.
0	3. Der Endharn einer gesunden Niere kann maximal die Konzentration des Interstitiums im Nirenmark annehmen, weil Harnstoff gegen Ende des Sammelrohrs austritt und entsprechend seines Konzentrationsgradienten in die Henle-Schleife zurück diffundiert. Durch diesen Harnstoffausstrom nimmt die Harnkonzentration so ab, dass sie nicht über der des Interstitiums im Nierenmark liegen kann.
1	4. Die Hauptaufgabe der Henle-Schleife ist die Aufrechterhaltung einer hohen Konzentration des Interstitiums im Nierenmark, welche die Bildung eines konzentrierten Harns ermöglicht.
1	5. Glomerulus und Bowman-Kapsel bilden die so genannte Blut-Harn-Schranke.
0	6. Durch eine künstliche Zugabe von Adiuretin ins proximale Konvult kann dort durch vermehrt geöffnete Wasserkanäle die Harnkonzentration gesteigert werden.
0	7. Ein künstlich hervorgerufener Ausfall des primär aktiven Natriumtransports im aufsteigenden Schenkel ist für die endgültige Harnkonzentrierung irrelevant, weil erst im Sammelrohr die endgültige Harnkonzentration stattfindet.
0	8. Die Hauptaufgabe der Henle-Schleife ist die Aufrechterhaltung eines Gegenstrommechanismus', um lebenswichtige Stoffe (z.B. Wasser, Natrium) über die Kapillaren in den Blutkreislauf aufzunehmen.

0	9. Das isotone Interstitium der Nierenrinde ist die Voraussetzung zur Harnkonzentrierung.
1	10. Eine künstliche Zugabe von Adiuretin ins proximale Konvult kann den dortigen Wasserausstrom nicht erhöhen.
1	11. Je länger die Henle-Schleife eines Nephrons ist, desto stärker kann der Harn konzentriert werden.
0	12. Die Filterfunktion der Niere besteht darin, dass sie nur überschüssige Molekularbestandteile (z.B. Harnstoff) aus dem Blut im Glomerulus abfiltriert, um diese dann durch ein vom Blutsystem getrenntes Tubulussystem ins Nierenbecken zu transportieren.
1	13. Die Transportvorgänge im aufsteigenden Schenkel verändern das umliegende Interstitium derart, dass die Harnkonzentration im absteigenden Schenkel durch einen Wasserausstrom zur Spitze hin ansteigt.
1	14. Der aktive Natriumtransport im aufsteigenden Schenkel ist neben der Harnrezirkulation ein Motor zur Aufrechterhaltung einer hohen Konzentration im Interstitium des Nierenmarks.
0	15. Der Wasserausstrom im absteigenden Schenkel verändert das Interstitium derart, dass die Harnkonzentration im aufsteigenden Schenkel durch einen erhöhten Natriumtransport zum distalen Konvult hin abnimmt.
1	16. Das Sammelrohr ist für die endgültige Harnkonzentration unter Kontrolle von Adiuretin verantwortlich.
1	17. Mit Hilfe des Hormons Adiuretin steigt die Harnkonzentration ab dem Tubulus reuniens durch den ermöglichten Wasserausstrom wieder an.
0	18. Die zur Spitze der Henle-Schleife hin steigende Harnkonzentration verstärkt den osmotischen Wasserausstrom im absteigenden Schenkel.
0	19. Im distalen Konvult nimmt die Harnkonzentration bis zum Tubulus reuniens stetig ab, weil die Membran ab dem aufsteigenden Schenkel impermeabel (undurchlässig) für Harnstoff ist.
1	20. Der Endharn einer gesunden Niere kann aus osmotischen Gründen maximal die Konzentration des Interstitiums im Nierenmark annehmen, weil Wasser nur so lange mit Hilfe des Hormons Adiuretin aus dem Sammelrohr strömen kann, bis ein Konzentrationsgleichgewicht zwischen Harn und Interstitium besteht.

10.2 Cognitive Load Items

The following items presented in Table 36 were used in German language to measure participants' cognitive load in Experiment 1 and 2.

Table 36

Cognitive load items

Intrinsic cognitive load	Wie schwierig fandest Du den Lerninhalt?
überhaupt nicht / nur wenig / etwas / ziemlich / sehr / extrem	
Extranoeus cognitive load	Wie schwer ist es Dir gefallen, mit dem gegebenen Material zu lernen?
überhaupt nicht / nur wenig / etwas / ziemlich / sehr / extrem	
Germane cogntive load	Wie sehr hast Du Dich während der Lernphase konzentriert?
gar nicht / nur wenig / etwas / ziemlich / sehr / extrem	

10.3 Control Variables

Control variables were knowledge about physiology, perceived task demands and interest.

10.3.1 Physiology Test

Table 37 shows the test sentences of the physiology test.

Table 37

Items of the physiology test

0 = falsch **1 = richtig**	**Welche der folgenden Aussagen ist richtig?**
1	1. Der Körper eines Erwachsenen besteht zu ca. 2/3 aus Wasser.
1	2. Bei Osmose diffundiert („wandert“) ein Lösungsmittel (z.B. Wasser) von Bereichen mit geringer Konzentration eines gelösten Stoffes in Bereiche mit höherer Konzentration des gelösten Stoffes.

0	3. Die so genannte Natrium-Kalium-Pumpe ist ein Enzym, das unter Energieverbrauch ADH in ADP und Phospat spaltet.
0	4. Bei Diffusion erfolgt der Stofftransport vom Ort der geringen Konzentration zum Ort der höheren Konzentration.
0	5. Die Osmolarität ist ein Maß für die Höhe der chemischen Lösungseigenschaft von molekularen Stoffen.
1	6. Die Natrium-Kalium-Pumpe transportiert Natrium aus der Zelle.
0	7. Eine Aufgabe der Natrium-Kalium-Pumpe ist die Aufrechterhaltung einer erhöhten Natriumkonzentration in der Zelle gegenüber dem Interstitium (Zellzwischenraum).
1	8. Eine Flüssigkeit ist gegenüber einer anderen isotonisch, wenn sie die gleiche Elektrolytkonzentration aufweist.
0	9. Eine wichtige Eigenschaft der Osmose ist ihre Reversibilität (Umkehrbarkeit).
1	10. Die Osmolarität gibt die Anzahl osmotisch aktiver Teilchen pro Liter Lösung an.
0	11. Eine Flüssigkeit ist gegenüber einer anderen hypertonisch, wenn sie eine geringere Elektrolytkonzentration aufweist.
1	12. Die Molarität gibt die Stoffmenge pro Volumen Lösung an.
1	13. Die Natrium-Kaliumpumpe baut einen elektrischen Gradienten über der Zellmembran auf.
1	14. Bei Osmose wird weder Energie freigesetzt noch benötigt.
0	15. Bei Osmose diffundiert ein gelöster Stoff von Bereichen mit höherer Konzentration des gelösten Stoffes in Bereiche mit geringerer Konzentration des gelösten Stoffes.
0	16. Die Durchblutung des Herzens ist pro Gewichtseinheit deutlich höher als die der Niere.
1	17. Unter Diffusion versteht man die Stoffwanderung infolge molekularer Teilchenbewegung zum Ausgleich von Konzentrationsunterschieden.
0	18. Die Molarität eines Stoffes wird in mol/kg angegeben.

10.3.2 Interest

Table 38 shows the item used to measure participants' interest after learning.

Table 38

Interest item

Interest	Wie interessant fandest Du das Lernthema?
	überhaupt nicht / nur wenig / etwas / ziemlich / sehr / extrem

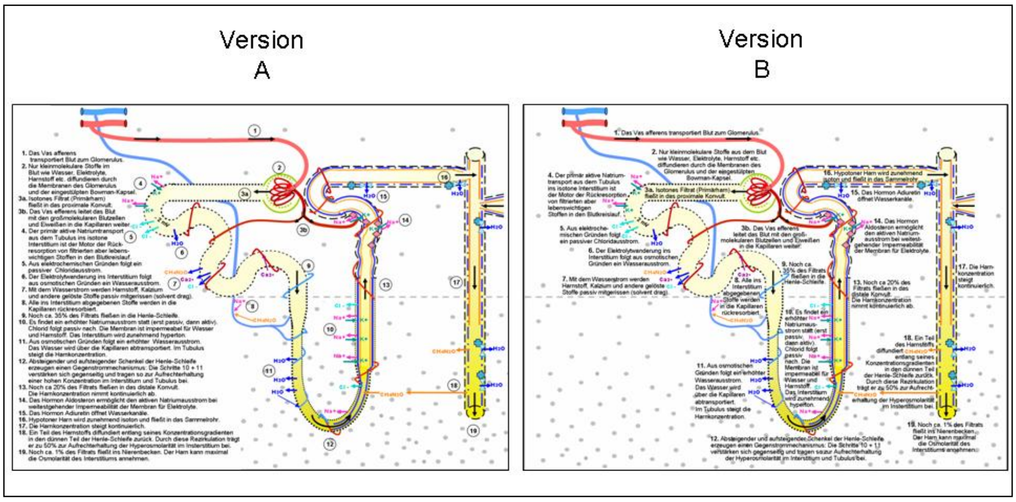
10.3.3 Perceived Task Demands

Table 39 shows the items used to measure participants' perceived task demands in Experiment 1 and 2.

Table 39

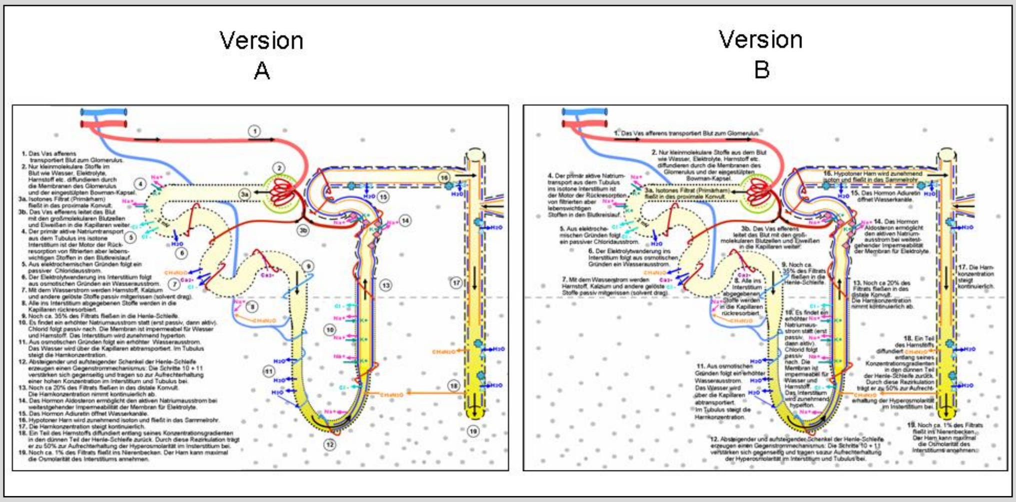
Items of perceived task demands

1. **Wie anstrengend ist es Deiner Meinung nach mit Version A zu lernen?**



nicht anstrengend / nur wenig / etwas / ziemlich / sehr / extrem anstrengend

2. **Wie anstrengend ist es Deiner Meinung nach mit Version B zu lernen?**



nicht anstrengend / nur wenig / etwas / ziemlich / sehr / extrem anstrengend