

The Role of Working Memory in Multimedia Instruction: Is Working Memory Working During Learning from Text and Pictures?

Anne Schüler · Katharina Scheiter ·
Erlijn van Genuchten

Published online: 27 May 2011
© Springer Science+Business Media, LLC 2011

Abstract A lot of research has focused on the beneficial effects of using multimedia, that is, text and pictures, for learning. Theories of multimedia learning are based on Baddeley's working memory model (Baddeley 1999). Despite this theoretical foundation, there is only little research that aims at empirically testing whether and more importantly how working memory contributes to learning from text and pictures; however, a more thorough understanding of how working memory limitations affect learning may help instructional designers to optimize multimedia instruction. Therefore, the goal of this review is to stimulate such empirical research by (1) providing an overview of the methodologies that can be applied to gain insights in working memory involvement during multimedia learning, (2) reviewing studies that have used these methodologies in multimedia research already, and (3) discussing methodological and theoretical challenges of such an approach as well as the usefulness of working memory to explain learning with multimedia.

Keywords Multimedia learning · Working memory · Cognitive Theory of Multimedia Learning · Cognitive Load Theory · Dual-task methodology · Working memory capacity

In the last two decades, a lot of educational research has focused on the beneficial effects of using multimedia for learning by referring to the Cognitive Theory of Multimedia Learning (Mayer 2009) and/or the Cognitive Load Theory (Sweller *et al.* 1998). Both theories suggest that multimedia learning can be best explained by paying close attention to how information is processed and stored in the human mind. In particular, working memory is assumed to be crucial in multimedia learning, because all information needs to be processed

A. Schüler (✉) · K. Scheiter · E. van Genuchten
Knowledge Media Research Center, Konrad-Adenauer-Strasse 40, 72072 Tübingen, Germany
e-mail: a.schueler@iwm-kmrc.de

K. Scheiter
e-mail: k.scheiter@iwm-kmrc.de

E. van Genuchten
e-mail: e.genuchten@iwm-kmrc.de

in working memory before it can be stored in long-term memory. Both, the Cognitive Theory of Multimedia Learning and Cognitive Load Theory refer to Baddeley's working memory model (e.g., Baddeley 1999) as their theoretical foundation.

Despite the assumed importance of working memory for multimedia learning, only few studies have empirically tested how it contributes to learning from text and pictures. We agree with Tardieu and Gyselinck (2003) that such research "is necessary in order to validate a theory based on load in working memory" (p. 18). Therefore, the goal of this paper is to stimulate such empirical research by (1) providing an overview of the methodologies that allow investigating working memory components involvement during multimedia learning, (2) reviewing studies that have used these methodologies in multimedia research, and (3) discussing the challenges and usefulness of such an approach.

To better understand how working memory may play a crucial role in multimedia learning, Baddeley's model is first outlined before discussing how it has been applied to multimedia learning.

Baddeley's Working Memory Model

According to Baddeley (1999), working memory is composed of multiple subsystems, namely, the phonological loop, the visuo-spatial sketchpad, and the central executive. More recently, the episodic buffer was added (e.g., Baddeley 2000). Each subsystem has its own limited capacity, which enables the subsystems to act relatively independent from each other. This implies that tasks involving different subsystems can be performed together equally well as separately. Conversely, there will be interference between two tasks involving the same subsystem. In addition, brain research has shown that the subsystems are associated with different brain regions (for overviews, see Baddeley 1998; Smith and Jonides 1997). Together, these findings imply a functional as well as neuropsychological dissociation between the subsystems.

The *phonological loop* (PL) is the subsystem that processes verbal information. Spoken words enter the PL's passive storage unit (i.e., the phonological store) directly, whereas written words have to be first converted from a visual code into an articulatory code. This conversion is done by the PL's subvocal rehearsal process before the words are transferred to the phonological store.

The *visuo-spatial sketchpad* (VSSP) is the subsystem that processes visual and spatial information. The VSSP's visual component deals with visual characteristics of objects (e.g., shape or color; Logie 1995), whereas its spatial component deals with relational or spatial information and the control of movements, for example, arm or eye movements (Lawrence *et al.* 2001; Logie 1995; Logie and Marchetti 1991). The VSSP is likely to be responsible for picture processing.

The *central executive* (CE) is the subsystem that is responsible for (a) monitoring and coordinating the operation of the subsystems and linking them to long-term memory, (b) switching attention between tasks and allocating attention to stimuli, (c) assigning information to one of the subsystems, (d) updating and regulating working memory contents, and (e) coding representations for their time and place of appearance (Baddeley 1996; Smith and Jonides 1999).

Prior to 2000, Baddeley assumed that coordination between the two subsystems was handled by the CE; however, the CE lacked storage capacity for retaining information in different codes. Therefore, the *episodic buffer* was added to the model, which allows temporary storage of multimodal information, and combines information from the PL and

VSSP with each other and with prior knowledge (Baddeley 2000). Like the PL and the VSSP, the episodic buffer is controlled by the CE.

With respect to multimedia learning, there are a couple of implications that can be derived from Baddeley's model. First, limitations of working memory resources should have adverse effects on multimedia learning. Second, these adverse effects should depend on which subsystem is affected by the resource limitations. This implies that limitations in the PL negatively affect the processing of verbal information, whereas limitations in the VSSP negatively affect the processing of pictorial information. Finally, limitations in the CE and episodic buffer should negatively affect the integration of verbal and pictorial information.

The Role of Working Memory in Theories of Multimedia Learning

Two assumptions of Baddeley's model have been incorporated into the Cognitive Theory of Multimedia Learning and Cognitive Load Theory; however, these theories differ in how heavily they rely on these assumptions. The *distinction between a verbal and a visuo-spatial working memory subsystem* is reflected in Cognitive Theory of Multimedia Learning's assumption that humans possess an auditory/verbal channel and a visual/pictorial channel for processing multimedia materials (Mayer 2005). These channels are distinguished on the basis of the sensory mode of representations (visual vs. auditory) and on their presentation code (verbal vs. pictorial). Mayer (2005) argues that the distinction according to the sensory mode of representations is "most consistent with Baddeley's [...] distinction between the visual-spatial sketchpad and the phonological (or articulatory) loop" (p. 34). It is important to note that this interpretation is not necessarily in line with Baddeley's model, since Baddeley distinguishes the PL and the VSSP according to the information's representational code (verbal vs. pictorial) and not according to their mode, as suggested by Mayer (cf. Rummer *et al.* 2010). We will come back to this issue when discussing the explanation underlying one of the most prominent multimedia design effects, that is, the modality effect.

A second assumption based on Baddeley's (1999) model is that the working memory subsystems have *limited resources for processing information* in parallel. Accordingly, Cognitive Theory of Multimedia Learning assumes that each channel can only handle a limited amount of information simultaneously. These limitations in processing capacity provide the main argument for how instruction can be improved by optimizing the cognitive demands that arise during studying multimedia materials.

Whereas Cognitive Theory of Multimedia Learning places equal weight on both assumptions, Cognitive Load Theory emphasizes the limited-capacity assumption at the expense of the dual-channel assumption. In the Cognitive Load Theory, cognitive resources are seen as a unitary construct that are not tied to a specific code of information, that is, verbal or visuo-spatial; instead, a distinction is made between cognitive load resulting from the domain's complexity in interaction with a learner's prior knowledge (intrinsic load) and load resulting from either harmful (extraneous load) or helpful cognitive processes (germane load; Sweller *et al.* 1998). According to the Cognitive Load Theory, cognitive overload occurs when the sum of these load types exceeds available working memory capacity. However, it is unclear how Cognitive Load Theory's assumptions map onto Baddeley's model, in which each subsystem has available its own resources and cognitive overload occurs only within a subsystem. Only for the modality effect, Cognitive Load Theory explicitly distinguishes between two working memory subsystems. This distinction

resembles Cognitive Theory of Multimedia Learning's distinction and will later be discussed in the context of the modality effect.

Neither Cognitive Theory of Multimedia Learning nor Cognitive Load Theory comprise assumptions about the role of the CE (or the episodic buffer) in multimedia learning, despite the fact that at least the Cognitive Theory of Multimedia Learning places emphasis on the fact that meaningful learning occurs only when verbal and pictorial information are integrated with each other and with prior knowledge. According to Baddeley's model, the coordination between the subsystems and long-term memory as well as storage of multimodal information is accomplished by the CE in cooperation with the episodic buffer (cf. Gyselinck *et al.* 2008).

To conclude, theories of multimedia learning assign a pivotal role to working memory in learning from text and pictures; however, the link between these theories and the Baddeley model is loose and appears at least partly inconsistent. Therefore, it is worthwhile to investigate empirically how working memory is involved during multimedia learning. To do so, reliable methods have been developed in cognitive psychology allowing to assess PL, VSSP, and CE contribution in any information-processing task. These methods are outlined next. No established methods exist yet to examine episodic buffer involvement during learning.

Assessing the Contribution of Working Memory Resources to Learning

According to the *capacity approach* to assessing working memory involvement, the capacities of the subsystems are gauged and linked to learning outcomes to test whether inter-individual differences in working memory capacity explain variance in learning outcomes (Andrade 2001). If differences in the capacity of one of the subsystems are associated with different learning outcomes, it can be deduced that this subsystem was involved during learning.

In the *dual-task approach*, participants perform a dual task in addition to learning. This dual task requires information processing in one of the subsystems. If there is interference between the primary learning task and the dual task—as, for instance, indicated by a drop in learning outcomes compared with a control condition without dual task—it can be deduced that this subsystem was involved during learning.

The Capacity Approach

Individual working memory capacity is gauged by measuring how much information can be processed by the subsystem(s) in question, which is denoted as the span of that subsystem. In the following, we make a distinction between simple span and complex span tasks (Carretti *et al.* 2009).

Simple span tasks are designed to gauge the PL and VSSP capacity by measuring the amount of information that can be stored over a short period of time. They require participants to recall sequences of stimuli, such as numbers or spatial configurations. They start with a short sequence (e.g., three stimuli) with three sequences of an equal number of stimuli (i.e., a set). After a sequence has been presented, the participant has to recall the stimuli in the correct order. When recall performance of a set is above a certain threshold, the instructor presents the next set, with each sequence being extended by one stimulus. When recall of a set is below a certain criterion, the task is

terminated. The sequences' length of the last correctly recalled set reflects the capacity of the subsystem.

Complex span tasks are designed to assess CE capacity. Participants are asked to store information in the PL or VSSP while additionally engaging in other cognitive activities (e.g., Daneman and Carpenter 1980). The CE is assumed to be involved during these tasks, because accomplishing them requires executive functions to monitor and coordinate between multiple cognitive tasks. To gauge CE capacity, at least two complex span tasks have to be administered, one involving the CE and PL and the other one involving the CE and VSSP. Then, latent variable analysis can be used to determine CE contribution, based on the variance in performance shared by the two tasks. The variance unique to each task is attributed to the contribution of the two subsystems (cf. Conway *et al.* 2005). Table 1 gives an overview of the tests used to measure individual differences in working memory capacity.

Phonological loop One of the most frequently used simple span tasks to measure PL capacity is the digit span task (Wechsler 1958). In this task, participants have to recall sequences consisting of randomly ordered digits between zero and nine.

Visuo-spatial sketchpad The most common simple span tasks to measure VSSP capacity are the Corsi block task for the spatial component and the Visual Pattern Test for the visual component. In the Corsi block task (Milner 1971; Vandierendonck *et al.* 2004), participants have to remember spatial sequences. The instructor taps cubes that are arranged irregularly on a board in fixed sequences. After presentation, the participant has to recall the sequences by tapping the same cubes in the same order.

Table 1 Tasks to measure working memory capacity

Task	Working memory subsystem	Reliability	Validity	Time (min)	References
Digit span	PL	.80 to .91	.46 to .47	15	Wechsler (1958), Spreen and Strauss (1998), and Waters and Caplan (1996)
Corsi block	VSSP	.38	.27 to .35	15	Milner (1971), Spinnler and Tognoni (1987), and Della Sala <i>et al.</i> (1999)
VPT	VSSP	.73 to .75	.27 to .35	15	Della Sala <i>et al.</i> (1997, 1999)
Reading span task	CE _{PL}	.70 to .80	.23 to .47	15	Daneman and Carpenter (1980), Conway <i>et al.</i> (2005), Kane <i>et al.</i> (2004), and Waters and Caplan (1996)
Listening span	CE _{PL}	.77	.52 to .77	20	Daneman and Carpenter (1980) and Lehman and Tompkins (1998)
Operational span	CE _{PL}	.80	.40 to .73	15	Turner and Engle (1989), Kane <i>et al.</i> (2004), and Conway <i>et al.</i> (2005)
Counting span task	CE _{PL}	.77	.47	— ^a	Engle <i>et al.</i> (1999) and Kane <i>et al.</i> (2004)
Spatial span task	CE _{VSSP}	.72 to .83	.83	— ^a	Shah and Miyake (1996) and Wechsler (1997)

^aNo information available

The Visual Pattern Test, developed by Della Sala *et al.* (1997), gauges the capacity of the visual part of the VSSP by measuring recall performance of abstract visual patterns. These patterns are presented in two-dimensional matrices, in which half of the cells are filled black. After presentation, participants have to recall the pattern by marking the black cells in an empty version of the same matrix.

The correlations between the Corsi block task and the Visual Pattern Test are low (Logie and Pearson 1997). Similarly, Della Sala *et al.* (1999) showed interference between the Visual Pattern Test and a visual dual task but little interference between the Visual Pattern Test and a spatial dual task. They found the opposite pattern for the Corsi block task. These findings indicate that the Visual Pattern Test and the Corsi block task gauge the capacities of independent visual and spatial components of the VSSP.

Central executive Examples of complex memory tasks that involve the CE and the PL (CE_{PL}) are the Reading Span Task (with written stimuli) and the Listening Span Task (with spoken stimuli) constructed by Daneman and Carpenter (1980). In these tasks, participants verify unrelated sentences while remembering the last word of each sentence in the set. The CE_{PL} capacity corresponds to the maximum number of correctly recalled final words while validating sentences. Both tasks show similar correlations with reading and listening comprehension, indicating that general language processes instead of specific reading or listening processes are addressed by these tasks.

A similar CE_{PL} measure is the Operational Span Task (Turner and Engle 1989). The Operational Span Task requires participants to solve mathematical operations while remembering words presented after each equation. At the end of a set, participants recall the presented words.

Finally, CE_{PL} capacity can be gauged by the Counting Span Task. This task involves counting shapes or objects (e.g., blue circles) on a display while ignoring other objects (e.g., blue squares) and storing the total number for later recall. After the presentation of a set, participants have to repeat the stored numbers in the correct order (Engle *et al.* 1999).

The Spatial Span Task (Shah and Miyake 1996) is the only complex span task that involves the CE and the VSSP (CE_{VSSP}). In this task, participants have to mentally rotate letters to decide whether or not these letters are mirrored. In addition to this verification task, participants have to recall the angles of rotation (e.g., 0°, 45°, 90°, etc.) at the end of each set of letters. However, it can be argued that the recall of angles, which are numbers, also involves the PL.

The Dual-Task Approach

Three kinds of dual tasks can be differentiated: secondary tasks, interference tasks, and preload tasks (Andrade 2001; Cocchini *et al.* 2002). The differentiation into three types of tasks is based on (a) the requirements imposed onto the participant, that is, active versus passive processing of information related to the dual task and (b) the point in time the dual tasks are presented, that is, before or during the main task.

The most common dual-task paradigm involves the use of a secondary task in addition to a primary task (e.g., multimedia learning). The secondary task is specifically designed to tap the resources of one working memory subsystem (Andrade 2001). If both tasks rely on the same subsystem, their simultaneous performance will yield interference and as a consequence, performance on either one or both tasks will decrease compared with a

control condition, in which participants perform the two tasks separately. If there is no decrease in performance, it can be concluded that the primary task is processed in a different subsystem than the secondary task. Less common variants of the dual-task paradigm include the use of interference tasks and preload tasks. *Interference tasks* are tasks in which participants are exposed to irrelevant stimuli that are assumed to get access to working memory automatically. Therefore, presenting these stimuli interferes with the primary task and hence reduces performance (Andrade 2001). *Preload tasks* require participants to keep information that is presented before the primary task in working memory while trying to encode the information of the primary task (Cocchini *et al.* 2002). Thus, when performance in the primary task and/or preload task performance are impaired compared with a control condition, it can be concluded that the primary task and the preload task require the same resources.

Whereas secondary tasks and preload tasks require active rehearsal of information by the participant, irrelevant stimuli are assumed to get access automatically to the respective working memory subcomponent. Active rehearsal and automatic access are both assumed to load the respective working memory subsystem.

Phonological loop Articulatory suppression, irrelevant speech, or verbal preload tasks can be used to examine PL involvement. Articulatory suppression is a secondary task to load the PL by requesting participants to articulate syllables, words, or numbers (Murray 1967). The task disturbs subvocal rehearsal and recoding of written information that may be required for the primary task (Gathercole and Baddeley 1993).

To examine PL involvement with an interference task, irrelevant speech can be presented (e.g., Colle and Welsh 1976). According to Baddeley (1999), effects of irrelevant speech on verbal tasks cannot be attributed to distraction, because presenting music without words does not affect performance in the primary task in the same way. To examine PL involvement with a verbal preload task a verbal stimulus, for example, a list of words, is presented before the primary task and has to be recalled after having worked on the primary task (e.g., Cocchini *et al.* 2002).

Visuo-spatial sketchpad Dual tasks such as spatial tapping, interference tasks, and visuo-spatial preload tasks can be used to measure VSSP involvement. Spatial tapping is a secondary task that is assumed to address the spatial component of the VSSP. In this task, participants have to continuously conduct specific movements, which are known to be controlled by the spatial component of the VSSP (e.g., pressing buttons on a hidden keyboard: Della Sala *et al.* 1999; foot tapping: Miyake *et al.* 2004). Spatial tapping has been shown to interfere with spatial tasks (e.g., Farmer *et al.* 1986), but not with visual tasks (Logie and Marchetti 1991). Therefore, when spatial tapping interferes with the primary task, it can be concluded that this task is processed in the spatial component of the VSSP. To measure visual VSSP involvement, no secondary task exists to our knowledge but only interference or preload tasks.

To examine visual VSSP involvement with an interference task, irrelevant visual stimuli are presented (e.g., McConnell and Quinn 2000). Quinn and McConnell (1996) introduced the *Dynamic Visual Noise* technique, in which they showed learners dots, which changed continuously and randomly between black and white. They demonstrated that this technique interfered with visual mnemonics, but not with a spatial task. *Static Visual Noise* did not show these interference effects (see McConnell and Quinn, experiment 1). To

examine VSSP involvement with the use of a visuo-spatial preload task, a visuo-spatial stimulus, for example a visual pattern, is presented before the primary task and has to be recalled after the primary task (e.g., Kruley *et al.* 1994).

Central executive To examine CE involvement, random generation tasks are used. Random generation tasks are secondary tasks, in which participants generate random sequences by naming letters or numbers or by tapping in a *random* order (e.g., Baddeley *et al.* 1998). Attentional processes of the CE are required in this task, because participants have to deliberately avoid stereotypical sequences, such as 1–2–3 or A–B–C. However, as random generation tasks also load the VSSP or the PL (e.g., because letters have to be verbalized), impaired performance can also indicate interference in one of these subsystems. Therefore, similar to the measurement of individual working memory capacity, at least two random generation tasks should be used, loading either the PL or the VSSP. CE contribution can then be determined by means of latent variable analysis. Finally, no established interference tasks or preload tasks for the CE exist to our knowledge.

Advantages and Disadvantages of Capacity and Dual-Task Approach

Implementation The capacity approach is convenient to assess working memory sub-components involvement because capacity measures are easy to administer before or after the actual experiment. However, gauging capacity within the same time slot as the experiment may prolong the experiment substantially, thereby causing fatigue and lack of motivation in the participants. This may become problematic especially when assessing CE involvement, because two complex span measures are required for unambiguous interpretation of the data (i.e., a CE_{PL} and a CE_{VSSP} task).

The dual-task approach is more challenging to implement. The first challenge is that baseline performance of the primary task should be gauged (i.e., the learning task without the dual task). In many basic cognitive psychology experiments this can be accomplished in a within-subjects design. However, with multimedia learning materials, usually no sufficiently comparable and therefore suitable stimuli exist that can be used for a baseline assessment. Hence, a between-subjects design has to be used, in which for every experimental condition investigated in the study two versions exist, one with and one without dual task. Not only does a between-subjects design increase error variance because of individual differences in responding to both, the primary and dual task, but also the number of participants increases in such a design.

The second challenge is that an adequate dual task should be selected. It should be possible to perform the primary task and the secondary task at the same time. For instance, a spatial finger-tapping task cannot be used when the hands are required to provide input during learning. In these cases, interference or preload tasks are better options.

Participant sample There are no specific requirements about sample size arising from the capacity approach unless the aim is to determine CE capacity based on latent variable analyses. In this case, large sample sizes are required (cf. Conway *et al.* 2005). However, the capacity approach requires a sufficient amount of inter-individual variance with regard to participants' working memory capacities. If this is not the case, it will be unlikely that a relationship between working memory capacity and learning outcomes is found. Especially

with rather homogenous samples of, for example, university (psychology) students this problem is likely to occur. The dual-task approach, on the other hand, makes no specific presumptions about sample characteristics; however, it may lead to a considerable increase in sample size necessary to determine baseline performance.

Analysis and Interpretation Data resulting from the capacity approach are easy to analyze, thereby providing a readily estimate of working memory components involvement. However, it is important to note that these data are correlative in nature. This implies that it is possible that a third variable (e.g., intelligence) underlies the observed relationship. Therefore, ideally more than one capacity measure should be used. If correlations between the capacity measures and learning outcomes are found (i.e., pointing toward dissociations between, for instance, PL and VSSP), an involvement of the underlying working memory component will be more likely, because if results are due to a third factor, it should influence *all* correlations between individual working memory measures and learning outcomes in a similar way (cf. Conway *et al.* 2003; Salthouse and Pink 2008).

Moreover, the capacity approach relies on the assumption that working memory capacity constitutes a relatively stable construct (i.e., a personal trait) that has an effect on processing irrespective of the situation. Hence, it presupposes that inter-individual capacity differences are large enough to affect learning outcomes and that effects are not overridden by other variables (e.g., prior knowledge). However, Baddeley (1999, p. 54ff.) acknowledges that, for instance, differences in PL capacity play hardly any role in accounting for differences in reading comprehension of adult readers, even though all information has to pass the PL in standard reading tasks. However, the majority of adult readers has much experience of reading so that lower-level skills and capabilities are unlikely to pose any obstacles; instead, higher-level skills such as the ability to draw inferences based on prior knowledge will become more influential.

The dual-task approach is a more direct approach to measuring working memory components involvement than the capacity approach, because the dual-task approach assesses which aspects of the experimental variations interfere with processes located in specific subsystems. Thus, decreases in task performance can be unambiguously traced back to interference in working memory and hence be interpreted in a causal manner. Moreover, as the dual-task approach focuses on how working memory handles situational demands depending on how many resources are already claimed by other processes, it may also be more sensitive and allows for a more fine-grained measurement of working memory components involvement. In particular, the exact time course of working memory components involvement can be determined over time and as a function of the processes conducted to accomplish the primary task. Therefore, the dual-task approach is tied more closely to the cognitive processes occurring, for instance, during multimedia learning.

To conclude, there are important differences between the validity of the two approaches. Whereas findings from the dual-task approach can be attributed unambiguously to the underlying working memory system, findings from measuring individual working memory capacity have to be interpreted with care, especially if only the capacity of one subsystem has been gauged. On the other hand, individual difference measures are easy to apply; hence, despite their potential drawbacks they are often used in studies in which working memory components involvement is not the main research question.

Assessing Working Memory Involvement During Multimedia Learning

In this section, we address our second goal of this review by discussing studies that examined the general contribution of working memory to multimedia learning. Moreover, we review studies that examined whether working memory components are involved in specific multimedia design effects, namely, the modality effect, to see whether their empirical evidence is in line with what to be expected from multimedia theories.

The literature search was conducted by searching PsychInfo and ERIC for the keywords “working memory,” “multimedia,” “(learning with) text and pictures,” “working memory capacity,” “dual task,” and “individual differences” and examinations of references of individual journal articles. Only studies that used at least one of the before-mentioned, established measures were included into the review, whereas studies investigating working memory components involvement using other types of measurements were excluded (with the exception of one study by Brünken *et al.* 2002, see below).

General Contribution of Working Memory to Multimedia Learning

Evidence from Studies Using the Capacity Approach One of the few researchers, who have examined the influence of working memory capacity on learning using simple span measures, are Gyselinck and colleagues (Gyselinck *et al.* 2002; Gyselinck *et al.* 2000). Gyselinck *et al.* (2000) presented written texts about physics to learners (e.g., static electricity) and varied between subjects whether or not a static picture accompanied the text. They divided learners into groups with either high or low VSSP capacity based on their performance on the Corsi block task. To ensure that PL capacity was comparable across groups, the digit span task was also administered. Learners with high VSSP capacity benefited from static pictures whereas learners with low VSSP capacity did not. This indicates that the VSSP is involved during static picture processing and that sufficient VSSP capacity is required to learn from pictorial representations.

In a follow-up study, Gyselinck *et al.* (2002, experiment 1) again varied between subjects whether or not written texts were accompanied by static pictures. Again, learners with high VSSP capacity benefited more from static pictures than learners with low VSSP capacity. Furthermore, learners with high VSSP capacity were strongly disturbed by a spatial tapping task performed during learning compared with low-capacity learners. A possible explanation is that learners with high VSSP capacity relied more on pictures than learners with low VSSP capacity and, therefore, were more disturbed by the spatial tapping task, leading to impaired learning outcomes.

Gyselinck *et al.* (2002, experiment 2) investigated PL involvement during learning with either only written text or a (labeled) static picture. They compared learners with either high or low PL capacity but equal VSSP capacity. Learners with high PL capacity performed better than learners with low PL capacity in the text-only condition, whereas there were no corresponding differences in the picture-only condition. This indicates that as expected, the PL is involved during text processing but not in picture processing. Furthermore, an articulatory suppression task disturbed text processing of learners with high PL capacity to a large extent, but had no effect on learners with low PL capacity. Again, a possible explanation is that learners with high PL capacity relied more on the text than learners with low PL capacity and, therefore, had less PL capacity available to perform the articulatory task, leading to impaired performance. On the other hand, a spatial tapping task had no

influence on text processing in both the high and low PL capacity groups, indicating that the VSSP was not involved during text processing.

A study using both simple and complex span measures was conducted by Pazzaglia *et al.* (2008). They asked Italian middle school students to learn about Germany's geography from a hypermedia system. After learning, the authors assessed the students' visuo-spatial mental representation of the hypermedia structure (map recognition test) and tested whether information was memorized and correctly integrated (semantic test). The presentation consisted of spoken text, written text, and pictures. Participants performed two simple span tasks, the digit span and Corsi block task, and two complex span tasks, the Listening Span and Dot Matrix Task. In the Dot Matrix Task, participants first verified a visual matrix equation and then viewed a 5×5 matrix containing a dot. After a series of these equations and matrices, participants marked all dots in a single answer 5×5 matrix. This task had been designed by the authors for this study and was assumed to assess the CE_{VSSP} . The results showed that Listening Span Task performance correlated with semantic knowledge acquisition, whereas Dot Matrix Task performance did not. Instead, the role of the CL_{VSSP} clearly emerged in the map recognition test assessing the visuo-spatial mental representation of the hypermedia structure. Neither digit span nor Corsi block task performance was related to map and semantic test performance, which might be because the dependent variables were not pure verbal or visuo-spatial memory measures, but required participants to connect information units into a coherent, global structure. This might explain why only CE_{VSSP} and CE_{PL} capacity correlated significantly with learning outcomes.

Only one complex span task was used in studies by Austin (2009, experiment 2) and Doolittle *et al.* (2009). Participants were assigned to one of three groups: animation and written text, animation and spoken text, or animation, spoken, and written text. Participants learned about lightning (Austin) or the working of a pump (Doolittle *et al.* 2009) and were afterwards tested using a recall and transfer test. In both studies, CE_{PL} capacity was gauged using the Operational Span Task. In Austin's study the CE_{PL} predicted a significant proportion of the variance in transfer test scores. Doolittle *et al.* (2009) used an extreme group design. Participants with high CE_{PL} capacity outperformed low-capacity participants on both tests. These results imply that the CE_{PL} contributed to multimedia learning. Interestingly, there were no interactions between CE_{PL} span and the three multimedia conditions, indicating that the CE_{PL} was involved to the same extent in all three conditions (for similar results, see Doolittle and Mariano 2008; Lusk *et al.* 2009).

Sanchez and Wiley (2006) used a Reading Span and an Operational Span Task to investigate whether CE_{PL} capacity affects the effectiveness of adding a picture to improve text comprehension. They assigned participants to a text-only group, a group with text and conceptually relevant pictures, or a group with text and irrelevant (i.e., seductive) pictures. The authors computed a composite score from the two CE_{PL} tasks and used this score to form extreme groups. Only text comprehension of learners with low CE_{PL} capacity suffered from presenting seductive pictures. A possible explanation is that high-capacity learners are better able to control their attention and ignore irrelevant pictures. This was confirmed by eye-tracking data showing that learners with high CE_{PL} capacity spent less time looking at irrelevant pictures.

Dutke and Rinck (2006) used the Reading Span and the Spatial Span Task to investigate how individual differences affect learning of spatial arrangements, which consisted of five objects (five words or five icons) located at five positions in a 2×3 matrix. For example, one spatial arrangement consisted of the five words *strawberries*, *apples*, *banana*, *pear*, and *pineapple*, which were located on five different positions within the 2×3 matrix. Importantly, the complete spatial arrangement was never shown to participants but had to

be inferred from four adjacent pairs of objects presented on their specific positions within the matrix. As dependent variable, the authors used the time to verify whether complete arrangements or object pairs depicted spatial relations that had been indirectly shown. They divided participants into four groups according to their CE_{PL} and CE_{VSSP} capacity (i.e., low CE_{PL} –low CE_{VSSP} , high CE_{PL} –low CE_{VSSP} , etc.). The results showed that integrating objects (i.e., words or icons) into a complete arrangement was most difficult for participants with low CE_{PL} and low CE_{VSSP} indicating that the CE plays a role in (mentally) integrating information, regardless of the involved subsystem. Furthermore, the authors expected that it would take longer to verify spatial arrangements containing words than verifying spatial arrangements containing icons. The results showed that low CE_{VSSP} capacity led to impaired performance in both cases, probably because of the spatial arrangement of the icons and words; however, low CE_{PL} capacity impaired performance only with spatial arrangements containing words. Thus, the processing of spatial arrangements irrespective of whether they contain words or icons seems to rely on the CE and the VSSP, whereas the processing of spatial arrangements with words seems to additionally rely on the PL.

To summarize, CE_{PL} as well as CE_{VSSP} measures indicate CE involvement during connecting and integrating information, but show diverse results about the processing of verbal and pictorial information. In accordance with the theoretical assumptions CE_{PL} measures are involved during verbal information processing, whereas CE_{VSSP} measures are involved during pictorial information processing (and *not vice versa*). These specific correlations indicate that the relationships between measures and learning outcome can likely be traced back to the underlying subsystem and not to a third variable. The role of the PL and VSSP capacity as measured by simple span tasks is less clear, because only a few studies have been conducted that addressed this issue. Whereas Gyselinck *et al.* (2000; 2002) showed specific correlations between PL and VSSP capacity and text and picture processing, respectively, Pazzaglia *et al.* (2008) observed no relationships, which might, however, also be explained by the dependent variables they used. Table 2 gives an overview over the empirical evidence for the contribution of working memory in multimedia learning by measuring its capacity.

Evidence from studies using the dual-task approach One of the first studies using the dual-task methodology in multimedia learning was conducted by Kruley *et al.* (1994). They addressed the question whether the VSSP and the PL are involved during understanding multimedia material using a preload design. In two experiments, they varied within subjects whether the participants had to keep visuo-spatial or verbal information in working memory during learning (preload condition) or not (control condition). Furthermore, in both experiments they varied within subjects whether or not static pictures accompanied auditory descriptions about causal systems. To investigate the VSSP involvement in multimedia learning (experiment 1), participants in the preload condition had to remember the location of dots presented in 2×2 matrices in working memory during learning. A matrix was presented before each sentence. Learning (i.e., listening to a sentence) was interrupted after each sentence and a second matrix was presented. Participants had to decide whether this matrix matched the matrix shown earlier. In the control condition, participants saw an empty matrix before each sentence; after each sentence, they saw a matrix and decided whether the dots in the matrix were above the centerline of the grid. Thus, no maintenance of visuo-spatial information was required during learning. The authors expected interference between processing of pictures and simultaneously retaining the visuo-spatial matrix, because they assumed that both tasks would load the VSSP. This hypothesis was confirmed for performance in the visuo-spatial matrix task. When no pictures but only

Table 2 Empirical evidence for the contribution of working memory in multimedia learning by measuring working memory capacity

	Verbal information	Pictorial information	Connecting/integrating information
PL	Involvement in processing written and spoken texts (Gyselinck <i>et al.</i> 2002)	No involvement in processing static pictorial information (Gyselinck <i>et al.</i> 2002)	–
VSSP	No involvement in processing written and spoken texts (Gyselinck <i>et al.</i> 2000; Gyselinck <i>et al.</i> 2002)	Involvement in processing static pictorial information (Gyselinck <i>et al.</i> 2000; Gyselinck <i>et al.</i> 2002)	–
CE _{PL}	Involvement in spoken and written text processing (Austin 2009; Doolittle <i>et al.</i> 2009; Dutke and Rinck 2006; Pazzaglia <i>et al.</i> 2008)	No involvement in the processing of static pictorial information (Dutke and Rinck 2006)	Involvement in connecting and integrating information (Austin 2009; Doolittle <i>et al.</i> 2009; Dutke and Rinck 2006; Pazzaglia <i>et al.</i> 2008) Involvement in controlling attention (Sanchez and Wiley 2006)
CE _{VSSP}	No involvement in written text processing (Dutke and Rinck 2006; Pazzaglia <i>et al.</i> 2008)	Involvement in the processing of static pictorial information (Dutke and Rinck 2006) The CE _{VSSP} predicted performance in a map recognition test (Pazzaglia <i>et al.</i> 2008)	Involvement in connecting and integrating information (Dutke and Rinck 2006)

spoken texts were presented, there was no interference. This indicates that both pictures and matrices were processed in the VSSP, but not text.

To investigate PL involvement during multimedia learning (experiment 4), participants retained digits during learning in the preload condition. In the preload condition, four digits were presented before each sentence. After the presentation of a sentence, participants verified whether the order of two test digits was the same as in the original sequence. In the control condition, the test digits and the original sequence were shown between the sentences, so that participants were not required to retain the original sequence. As expected, the verbal preload task impaired text comprehension and performance in the digit task in the text only as well as in the text-picture condition, indicating that the text and the verbal preload task were processed in the PL. Furthermore, as the presentation of a picture had no specific influence on the performance in the digit task, it can be concluded that pictures were not processed in the PL.

Gyselinck *et al.* (2000) conducted a similar study with written instead of spoken text. In addition to measuring individual working memory capacity and varying picture presentation between subjects (see prior section), they varied within subjects whether participants had to keep visuo-spatial information or verbal information (preload conditions) or both (control condition) in working memory during learning. To load the VSSP, participants retained matrices during learning, whereas to load the PL they memorized nonwords. In the control condition, participants retained the position of three nonwords within a 2×2 matrix during learning. In this study, PL and VSSP involvement in multimedia learning using the dual-task methodology was not confirmed. A likely

explanation is that both subsystems were heavily loaded in the control condition, which required information processing in both the PL and VSSP.

Gyselinck *et al.* (2002) used secondary tasks instead of preload tasks. In addition to the between-subject picture presentation (see prior section), they varied within subjects whether participants performed a spatial tapping task, an articulatory suppression task, or no secondary task during learning. The results confirmed the hypothesis that picture processing involves the VSSP and text processing the PL. Specifically, the spatial tapping task impaired comprehension only when texts were accompanied by pictures, but not when only texts were presented. On the other hand, the articulatory suppression task impaired learning within the text-only and the text-picture conditions to a similar extent, without diminishing the advantage of picture presentation.

Whereas the above mentioned studies concentrated on causal tasks (e.g., the functioning of a volcano), Brunyé *et al.* (2006) examined whether working memory components are also involved during learning procedures (e.g., assembling Kinder Egg™ toys). They varied presentation format between subjects, with a written text only, a static picture only, and a written text-static picture group. Furthermore, the authors manipulated between groups whether participants performed a spatial tapping task (VSSP), an articulatory suppression task (PL), a random generation articulatory task (CE_{PL}), or a random generation tapping task (CE_{VSSP}) during learning. In the control group, participants learned without secondary task. As in Gyselinck *et al.* (2002), there was interference between spatial tapping and picture processing as well as between articulatory suppression and text processing, indicating that pictures were processed in the VSSP and text was processed in the PL. Additionally, learners studying text and picture suffered more from the random generation tasks than the text-only or picture-only groups, indicating that the CE plays an important role in multimedia learning, probably because CE resources are needed to integrate texts and pictures.

So far, only dual-task studies have been reported that examined working memory components involvement during learning with text and *static* pictures. Nam and Pujari (2005) investigated whether the VSSP is also involved when dynamic pictures are presented by varying whether animations, static pictures, or no pictures accompanied a written text describing technical systems (e.g., the functioning of a fridge). To examine VSSP involvement, learners conducted a finger-tapping task during learning. No secondary task was performed in the control groups. The interaction between presentation type and secondary task showed that learning was more impaired when combining spatial tapping with either dynamic or static pictures than without pictures. This implies that both static and dynamic pictures are processed in the VSSP. It must be noted, however, that spatial tapping also impaired performance in the text-only condition, indicating that text is also processed in the VSSP under specific circumstances (see for example, De Beni *et al.* 2005).

To summarize, the studies using the dual-task methodology confirm the assumption that the PL and VSSP are involved during multimedia learning: The PL is involved during written and spoken text processing, but not in the processing of static and dynamic pictures; the VSSP is involved during static and dynamic picture processing, but not during the processing of written and spoken text. Only one study by Brunyé *et al.* (2006) investigated CE involvement during multimedia learning using the dual-task methodology. The results confirm the assumption that the CE is involved during connecting and integrating multimedia information (see Table 3 for an overview).

Accordingly, both types of methodologies confirm the assumption that the different working memory subcomponents have specific responsibilities during multimedia learning. This indicates that information-processing theories, such as Baddeley's working memory

Table 3 Empirical evidence for working memory involvement during multimedia learning by applying dual-task methodology

	Verbal information	Pictorial information	Connecting/integrating information
PL	Involvement in processing written and spoken texts (Brunyé <i>et al.</i> 2006; Gyselinck <i>et al.</i> 2002; Gyselinck <i>et al.</i> 2008; Kruley <i>et al.</i> 1994)	No involvement in processing static pictorial information (Brunyé <i>et al.</i> 2006; Gyselinck <i>et al.</i> 2002; Kruley <i>et al.</i> 1994)	No involvement in connecting and integrating information (Brunyé <i>et al.</i> 2006)
VSSP	No involvement in processing written and spoken texts (Brunyé <i>et al.</i> 2006; Gyselinck <i>et al.</i> 2002; Gyselinck <i>et al.</i> 2008; Kruley <i>et al.</i> 1994) Involvement in processing written and spoken texts (Nam and Pujari 2005)	Involvement in processing static and dynamic pictorial information (Brunyé <i>et al.</i> 2006; Gyselinck <i>et al.</i> 2002; Gyselinck <i>et al.</i> 2008; Kruley <i>et al.</i> 1994; Nam and Pujari 2005)	No involvement in connecting and integrating information (Brunyé <i>et al.</i> 2006)
CE _{PL}	–	–	Involvement in connecting and integrating information (Brunyé <i>et al.</i> 2006)
CE _{VSSP}	–	–	Involvement in connecting and integrating information (Brunyé <i>et al.</i> 2006)

model (Baddeley 1999), can be used to derive hypotheses about the processing of multimedia materials as has been suggested by Cognitive Theory of Multimedia Learning and Cognitive Load Theory. However, as discussed earlier these theories appear to be only loosely linked to Baddeley's model. Therefore, it is unclear whether their explanations that are based on working memory processes also hold when looking at more specific multimedia design effects. One effect for which assumptions about its cause refer to working memory processes, is the modality effect, that is, the finding that combining spoken text with pictures yields better learning outcomes than combining written text with pictures (for an overview, see Ginns 2005).

Contribution of Working Memory to the Modality Effect

According to Cognitive Theory of Multimedia Learning and Cognitive Load Theory, the modality effect can be explained by assuming that in the initial processing stages in working memory, written texts and pictures compete for the same resources in the visual part of the visual/pictorial channel (i.e., the VSSP), because both are presented visually. However, with spoken texts, texts are processed in the auditory part of the auditory/verbal channel (i.e., the PL), whereas pictures are processed in the visual part of the visual/pictorial channel. As discussed before, this assumption is not consistent with Baddeley's model (Baddeley 1999), which suggests that both spoken and written texts are solely processed in the PL (i.e., unless they contain visuo-spatial information, cf. de Beni *et al.* 2005). The aforementioned methodologies make it possible to shed light on the working memory explanation of the modality effect: according to Baddeley (1999) the degree to

which the VSSP is involved during multimedia learning should be the same for spoken and for written text, whereas according to Cognitive Theory of Multimedia Learning and Cognitive Load Theory the VSSP should be involved more strongly in multimedia presentations containing written rather than spoken text. Thus, there should be stronger interference between dual tasks and multimedia learning in case of written than in spoken text.

To test this assumption, Brünken *et al.* (2002) asked students to learn from either a written or a spoken multimedia presentation, while responding as fast as possible whenever a letter presented at the top of the computer screen changed color (dual task). Unfortunately, this dual task is likely to induce higher demands related to early (sensory) visual attention processes, because it requires visually monitoring the letter, and hence does not allow investigating working memory involvement in an unambiguous manner. Accordingly and unsurprisingly, this dual task interfered with multimedia learning especially in conditions with written instead of spoken text, because with written text, visual attention is required not only for processing the pictorial but also the verbal input (cf. Cierniak *et al.* 2009, for a discussion of this problem). Furthermore, this kind of dual task does not allow for distinguishing between the load within either the PL and VSSP, which is, however, necessary to validate the specific assumptions made by Cognitive Theory of Multimedia Learning and Cognitive Load Theory. The same is true for studies conducted by Brünken *et al.* (2004) and Seufert *et al.* (2009), which is why we do not describe them in detail here.

A more adequate test of the working memory explanation of the modality effect comes from a study by Gyselinck *et al.* (2008). The authors presented static pictures to learners and varied between subjects whether the pictures were accompanied by either spoken or written text. Within subjects, they varied whether learners performed a spatial finger-tapping task, an articulatory suppression task, or no secondary task during learning. The results showed that articulatory suppression impaired learning compared with the control condition, which indicates that both written and spoken texts are processed in the PL. Additionally, the spatial tapping task impaired learning with written and spoken multimedia materials to the same extent, possibly because learners in both conditions processed static pictures. Most importantly, the authors found no interference between written text and the spatial tapping task, which is not in line with the assumption that written text is initially processed in the visual/pictorial channel as proposed by the Cognitive Theory of Multimedia Learning and Cognitive Load Theory. For alternative explanations of the modality effect that are based on sensory memory rather than working memory processes, the reader is referred to Rummer *et al.* (2010, 2011) and Schmidt-Weigand *et al.* (2010).

The modality effect provides an excellent example of how working memory methodologies can be used to test specific assumptions about the causes underlying multimedia design effects. To do so, it is vital that these explanations are described at a level that allows a clear prediction to be made about underlying working memory mechanisms, which, with the exception of the modality effect, however, hardly exist.

Discussion

Multimedia theories explain multimedia learning against the background of Baddeley's working memory model (Baddeley 1999). Our overview of studies that explicitly investigated working memory contribution to multimedia learning suggests that there is converging evidence that the PL and the VSSP have different responsibilities during processing multimedia instruction, with the PL being responsible for processing verbal

information—irrespective of the modality in which the text is presented—and with the VSSP being responsible for processing pictorial information such as static or dynamic visualizations (e.g., Brunyé *et al.* 2006; Gyselinck *et al.* 2002, 2008; Kruley *et al.* 1994; Nam and Pujari 2005). Relatively little research has focused on how text and pictures are integrated in working memory; however, the first results indicate that the CE plays a major role in this process (e.g., Austin 2009; Brunyé *et al.* 2006; Doolittle *et al.* 2009; Pazzaglia *et al.* 2008). Accordingly, we can conclude based on this review that working memory by and large is working during learning from text and pictures in the way one would expect it based on Baddeley's model.

There is relatively little research yet about working memory components involvement during multimedia learning. One reason could be that the dual-task methodology as the—potentially better suited—way of investigating working memory contribution may be difficult to implement, because the stimuli used in multimedia research are much more complex than stimuli in basic cognitive psychology experiments. The second reason could be that task features of experimental materials are more difficult to control systematically in multimedia learning (e.g., the question of whether spatial information is also contained in the text, which will possibly affect whether VSSP is involved during verbal information processing or not; De Beni *et al.* 2005; Schmidt-Weigand and Scheiter 2011). Finally, it is difficult to generate multiple instances of instructional materials that all share the same features and that can be learned independently from each other, so that they can be used in within-subjects designs for the assessment of baseline performance. To sum up, especially the use of dual-task methodologies imposes certain requirements that may be difficult to meet with multimedia materials, thereby explaining why until now there have been so few studies investigating working memory involvement during multimedia learning.

There are slightly more studies, in which working memory capacities were gauged and related to performance. Because of the correlative nature of this approach, the results have to be interpreted with care. Also, in most studies we reviewed, participants were categorized into high and low-capacity groups on the basis of their performance in the respective tests. The resulting group variable was then incorporated in statistical analyses as a nominal factor (e.g., Gyselinck *et al.* 2000). This dichotomizing procedure often results in rather small group sizes associated with a loss of power in the statistical analyses. Therefore, many authors have suggested using continuous predictors in regression analysis instead (e.g., Irwin and McClelland 2003; MacCallum *et al.* 2002), which allows testing for both main effects of working memory capacity on multimedia learning and interactions between working memory capacity and multimedia design variables (cf., Aiken and West 1991). Accordingly, especially researchers using the working memory capacity approach need to be aware of the limitations in terms of interpreting results as being caused by capacity differences and analyze their data by using adequate statistical procedures.

Whereas the issues discussed until this point involve methodological aspects that need to be considered in future research, more important conclusions can be drawn with respect to refining theories of multimedia learning. Even though the reviewed studies strongly supported the theories' assumption that working memory components are involved during multimedia learning, this review also reveals that there are inconsistencies between Baddeley's working memory model (Baddeley 1999) and the way the model has been used in multimedia research.

First, the theories of multimedia learning refer to Baddeley's model in a rather superficial way, which leads to inconsistencies: Cognitive Load Theory (Sweller *et al.* 1998) promotes a notion of overall working memory load (with the exception of the modality-effect explanation). This is much more in line with the short-term memory concept of Atkinson

and Shiffrin (1968) than with the model of Baddeley, whose major theoretical achievement compared with its theoretical predecessor has been the differentiation of working memory into various subsystems.

Second and more importantly, both Cognitive Theory of Multimedia Learning and Cognitive Load Theory incorporate Baddeley's assumptions about the distinction between working memory subsystems in a way that is inconsistent with Baddeley's model and that cannot be empirically confirmed. In particular, dual-task studies do not provide any support for the assumption that written text is initially processed in the VSSP, because no interference with written text processing could be demonstrated for secondary tasks loading the VSSP (e.g., Brunyé *et al.* 2006; Gyselinck *et al.* 2008). Thus, these empirical results call into question Cognitive Theory of Multimedia Learning's and Cognitive Load Theory's assumption that all visually presented materials (including written text) are initially processed in the VSSP, that is, the visual part of the visual/pictorial channel. This implies that the distinction of the two channels on the basis of sensory mode of representation (i.e., auditory vs. visual) is questionable, unless it refers to sensory rather than working memory. In other words, a distinction of the two channels on the basis of presentation code (i.e., verbal vs. pictorial) is more suitable to explain text and picture processing.

The third inconsistency is that based on Baddeley's model, it can be assumed that the CE is important during multimedia learning, which was confirmed by the studies in this review. This finding can be explained by the fact that the presentation of two information sources (i.e., text and picture) requires the allocation of attention to stimuli, while ignoring others, as well as monitoring and coordinating the subsystems. However, the CE is not explicitly incorporated into Cognitive Theory of Multimedia Learning, but only mentioned briefly as a structure that allocates, monitors, and adjusts limited cognitive resources (Mayer 2005). In earlier versions of the Cognitive Load Theory, the CE is not acknowledged at all. In more recent publications schemas stored in long-term memory are assumed to take over executive functions (e.g., Sweller 2005). Accordingly, in the framework of Cognitive Load Theory executive control is domain specific in that it depends on the schemas acquired so far. This view differs from Baddeley's notion of a domain-unspecific CE. Moreover, according to Baddeley (e.g., Baddeley 1999), the CE is an active control structure that is similar to the Supervisory Attentional System initially proposed by Norman and Shallice (1986) as a system for deliberate action control for novel and complex tasks. Norman and Shallice propose another control mechanism that is, however, limited to the control of routine or semi-automatic, well-learned actions. This contention scheduling „acts through the activation and inhibition of supporting and conflicting schemas“ (Norman and Shallice 1986, p. 3) and may, thus, be more alike to the way Sweller (2005) conceives executive control. Nevertheless, it is important to note that it is not identical to deliberate action control established by the CE or the Supervisory Attentional System, respectively. In short, Sweller and Baddeley have different understandings of the central executive, despite the fact that Baddeley's working memory model builds one of the theoretical foundations of Cognitive Load Theory. Because of its empirically demonstrated importance during multimedia learning, we suggest to more strongly emphasize the role of the CE as conceptualized in Baddeley's working memory model by explicitly embedding it into multimedia theories. Moreover, the ongoing debate about the existence and functioning of the episodic buffer (see Pearson 2006 for an overview) needs to be taken into account in future research, as from a theoretical point of view this subsystem is also involved during multimedia learning (cf. Gyselinck *et al.* 2008). However, currently no established methodologies exist that allow investigating this assumption.

It is important to note that the inconsistencies mentioned here are not just part of an academic debate but have important consequences for educational practice. These

inconsistencies challenge some of the explanations for well-established empirical effects in multimedia research. For instance, if written text is processed in the PL from the beginning on (similar to spoken text), then there is no reason to believe that using spoken rather than written text would free cognitive resources of the VSSP. If other mechanisms than the ones initially suggested were responsible for causing a multimedia design effect, then there might be also other boundary conditions under which this effect can be observed. Thus, recommendations about the design of multimedia materials may need to be changed as a function of modifying explanations at the theory level. Hence, conducting research with the aim of gaining a better understanding of the cognitive mechanisms underlying multimedia learning is worthwhile not only for theory building, but also for improving instructional materials.

Nevertheless, even though we believe that theories of multimedia learning should be grounded in cognitive psychology, some general concerns remain when linking educational theories to basic information-processing theories such as Baddeley's working memory model.

The first concern is that basic information-processing theories do not offer the most useful level of explanation, even if working memory's subsystems are involved during multimedia learning. That is, when we discuss working memory contributions, we typically refer to processes that take up to two seconds, while multimedia learning in real-world educational scenarios spans across much longer timeframes. With longer timeframes, higher-level variables such as motivation or the availability of cognitive and metacognitive learning strategies may become more important to account for differences in learning outcomes. Hence, even though multimedia learning is information processing for explaining differences in achieving educational objectives with multimedia materials, a broader perspective is required. According to this view, working memory is necessary but not sufficient to explain multimedia learning.

The second concern is that pictures might serve different functions in the learning process (see Carney and Levin 2002), which in turn might induce different working memory demands on the learner. Currently, this aspect is rather neglected in multimedia research, however, it cannot be excluded the possibility that the multimedia effects—which are based on assumptions about working memory demands—depend on the specific functions of the presented pictures.

The third concern is that any attempt to ground educational theories in cognitive psychology will always be faced with the challenge to keep up with the insights generated in cognitive psychology. For instance, alternatives to the Baddeley model have been proposed such as Cowan's Embedded Processing Model (see Miyake and Shah 1999, for an overview). However, up to now none of these alternative models has become as widespread across different research communities as Baddeley's model; hence, educational researchers may still be well advised when referring to the latter. Novel insights may also come from neuroscience research. In fact, there is a vast amount of neuroscience research that corroborates the assumptions about working memory (for an overview see Osaka *et al.* 2007). However, so far there are hardly any studies that have utilized the methodologies of neuroscience to examine working memory components involvement during multimedia learning, even though possible links between neuroscience and Cognitive Load Theory have been discussed (cf., for a review Antonenko *et al.* 2010). Finally, especially reading comprehension research has made a theory shift by moving from amodal to modality-specific representations in long-term memory (e.g., Fischer and Zwaan 2008). All these developments may play a role in building comprehensive cognitive theories of multimedia learning in the future. However, accounting for these developments in multimedia research

requires extensive knowledge of both educational as well as cognitive psychology research. At the moment, it seems safe to argue that using the current multimedia learning theories is appropriate for educational research as long as they explain the phenomena of interest. Nevertheless, we need to be aware that we might overlook things, stagnate on the current status quo because new insights are missing, or move into wrong directions, because the theories that we use shape the way we conduct research in the field.

References

- Aiken, L., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.
- Andrade, J. (2001). *Working memory in perspective*. Hove: Psychology Press.
- Antonenko, P. D., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22, 425–438. doi:10.1007/s10648-010-9130-y.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York: Academic.
- Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers & Education*, 53, 1339–1354. doi:10.1016/j.compedu.2009.06.017.
- Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, 49, 5–28. doi:10.1080/027249896392784.
- Baddeley, A. D. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, 8, 234–238. doi:10.1016/S0959-4388(98)80145-1.
- Baddeley, A. D. (1999). *Essentials of Human Memory*. Hove: Psychology Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423. doi:10.1016/S1364-6613(00)015382.
- Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, 51, 819–852. doi:10.1080/027249898391413.
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual task methodology. *Experimental Psychology*, 49, 109–119. doi:10.1027//1618-3169.49.2.109.
- 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. doi:10.1023/B:TRUC.0000021812.96911.c5.
- Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. B. (2006). Learning procedures: The role of working memory in multimedia learning experiences. *Applied Cognitive Psychology*, 20, 917–940. doi:10.1002/acp.1236.
- Carney, R. N., & Levin, R. N. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14, 5–26. doi:10.1023/A:1013176309260.
- Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19, 246–251. doi:10.1016/j.lindif.2008.10.002.
- 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(2), 315–324. doi:10.1016/j.chb.2008.12.020.
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, 30, 1086–1095.
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour*, 15, 17–31. doi:10.1016/S0022-5371(76)90003-7.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science*, 7, 547–552. doi:10.1016/j.tics.2003.10.005.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.

- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Learning and Verbal Behavior*, *19*, 450–466. doi:10.1016/S0022-5371(80)90312-6.
- De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial description. *European Journal of Cognitive Psychology*, *17*, 77–95. doi:10.1080/09541440340000529.
- Della Sala, S., Gray, C., Baddeley, A. D., & Wilson, L. (1997). *Visual Pattern Test: A test of short-term visual recall*. London: Harcourt Assessment.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern spans: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*, 1189–1199. doi:10.1016/S0028-3932(98)00159-6.
- Doolittle, P. E., & Mariano, G. J. (2008). Working memory capacity and mobile multimedia learning environments: Individual differences in learning while mobile. *Journal of Educational Multimedia and Hypermedia*, *17*, 511–530.
- Doolittle, P. E., Terry, K. P., & Mariano, G. J. (2009). Multimedia learning and working memory capacity. In R. Zheng (Ed.), *Cognitive effects of multimedia learning* (pp. 17–33). London: Premier Reference Source.
- Dutke, S., & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction*, *16*, 526–537. doi:10.1016/j.learninstruc.2006.10.002.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology*, *128*, 309–331. doi:10.1037/0096-3445.128.3.309.
- Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology*, *38*(4-A), 675–688.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*, 825–850. doi:10.1080/17470210701623605.
- Gathercole, S., & Baddeley, A. (1993). *Working memory and language processing*. Hove: Erlbaum.
- Gimms, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*, 313–331. doi:10.1016/j.learninstruc.2005.07.001.
- Gyselinck, V., Ehrlich, M.-F., Comoldi, C., de Beni, R., & Dubois, V. (2000). Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning*, *16*, 166–176. doi:10.1046/j.1365-2729.2000.00128.x.
- Gyselinck, V., Cornoldi, C., Dubois, V., de Beni, R., & Ehrlich, M.-F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology*, *16*, 665–685. doi:10.1002/acp.823.
- Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, *22*, 353–374. doi:10.1002/acp.1411.
- Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Market Research*, *40*, 366–371. doi:10.1509/jmkr.40.3.366.19237.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217. doi:10.1037/0096-3445.133.2.189.
- Krueley, P., Sciana, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, *22*, 261–272.
- Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effect of eye and limb movements on working memory. *Memory*, *9*, 433–444. doi:10.1080/09658210143000047.
- Lehman, M. T., & Tompkins, C. A. (1998). Reliability and validity of and auditory working memory measure: Data from elderly and right-hemisphere damaged adults. *Aphasiology*, *12*, 771–785. doi:10.1080/02687039808249572.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Aberdeen: Lawrence Erlbaum Associates.
- Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 105–115). Amsterdam: North-Holland Press.
- Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*, 241–257. doi:10.1080/713752559.
- Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. F. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Developmental Psychology*, *40*(4), 636–651. doi:10.1111/j.1467-8535.2008.00848.x.

- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40. doi:10.1037//1082-989X.7.1.19.
- Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.
- McConnell, J., & Quinn, J. G. (2000). Interference in visual working memory. *The Quarterly Journal of Experimental Psychology*, 53, 53–67. doi:10.1080/027249800390664.
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27(3), 272–277.
- Miyake, A., & Shah, P. (1999). *Models of working memory. Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
- Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, 115, 123–142. doi:10.1016/j.actpsy.2003.12.004.
- Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 21, 263–276. doi:10.1037/h0082978.
- Nam, C. S., & Pujari, A. (2005). The role of working memory in multimedia learning. In *Proceedings of the 11th International Conference on Human-Computer Interaction (HCI '05)*. Las Vegas: Mira Digital Publishing. CD-ROM
- Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum.
- Osaka, N., Logie, R. H., & D'Esposito, M. (2007). *The cognitive neuroscience of working memory*. Oxford: Oxford University Press.
- Pazzaglia, F., Toso, C., & Cacciamani, S. (2008). The specific involvement of verbal and visuospatial working memory in hypermedia learning. *British Journal of Educational Technology*, 39(1), 110–124.
- Pearson, D. G. (2006). The episodic buffer. Implications and connections with visuo-spatial research. In T. Vecchi & G. Bottini (Eds.), *Imagery and spatial cognition* (pp. 139–153). Amsterdam: John Benjamins.
- Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *The Quarterly Journal of Experimental Psychology*, 49, 200–215. doi:10.1080/027249896392865.
- Rummer, R., Schweppe, J., Fürstnerberg, A., Seufert, T., & Brünken, R. (2010). Working memory interference during processing texts and pictures: Implications for the explanation of the modality effect. *Applied Cognitive Psychology*, 24, 164–176. doi:10.1002/acp.1546.
- Rummer, R., Schweppe, J., Fürstnerberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied* (in press).
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, 15, 364–371. doi:10.3758/PBR.15.2.364.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34, 344–355.
- Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. *Computers in Human Behavior*, 27, 22–28. doi:10.1016/j.chb.2010.05.007.
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: New insights from investigating students' viewing behavior. *Applied Cognitive Psychology*, 24, 226–237. doi:10.1002/acp.1554.
- Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning and Instruction*, 19, 28–42. doi:10.1016/j.learninstruc.2008.01.002.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27. doi:10.1037/0096-3445.125.1.4.
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5–42. doi:10.1006/cogp.1997.0658.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657–1661. doi:10.1126/science.283.5408.1657.
- Spinnler, H., & Tognoni, G. (1987). Italian standardization and classification of neuropsychological tests. *Italian Journal of Neurological Science*, 6, 1–120.
- Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests*. New York: Oxford University Press.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.

- Sweller, J., van Merriëboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–285. doi:10.1023/A:1022193728205.
- Tardieu, H., & Gyselinck, V. (2003). Working memory constraints in the integration and comprehension of information in a multimedia context. In H. van Oostendorp (Ed.), *Cognition in a Digital World* (pp. 3–24). London: Lawrence Erlbaum Associates.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. doi:10.1016/0749-596X(89)90040-5.
- Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*, 57–79. doi:10.1348/000712604322779460.
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology*, *49A*, 51–79. doi:10.1080/027249896392801.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1997). *WAIS-III and working memoryS-III: Technical manual*. San Antonio: Harcourt Brace & Company.