APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **22**: 353–374 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1411

The Role of Working Memory Components in Multimedia Comprehension

VALÉRIE GYSELINCK[1]*, ERIC JAMET[2] and VÉRONIQUE DUBOIS[1]

[1]*University Paris Descartes, France*
[2]*Université Rennes 2, France*

SUMMARY

This article argues that conducting experiments involving the ability to control and even manipulate the cognitive load in working memory (WM; storage and or processing load) should make it possible to identify the processes involved during the integration of information coming from multiple sources. Some experiments using the dual-task paradigm are reviewed, and an original experiment using complex multimedia material is presented. Overall, the experiments show that even in cases where subjects have to navigate between different types of information and have to integrate various items of information, the verbal storage component of WM is important in permitting comprehension. Visuospatial WM is in addition involved as soon as visuospatial processing is required. Storage of verbal information does not however depend on the modality of presentation and the classical modality effect appears to depend on individual differences. Some theoretical and practical implications of these results are drawn. Copyright © 2008 John Wiley & Sons, Ltd.

In the context of the development of new technologies for the communication of information, multimedia systems are developing quickly and raise many questions, especially with reference to education. A multimedia system can be defined as a system which requires the integration of different types of information: verbal information presented visually or auditorily (e.g. words, sentences or short texts), pictorial information presented visually in a static or dynamic way (illustrations, photographs, diagrams) and sound information. In addition, systems that allow users to navigate between different sources of information through the use of hypertext structures are often considered to be multimedia systems, even if only one type of information is provided (for example, verbal information presented visually). To what extent does the use of such systems make a positive contribution to the learning process? There is a strong temptation to simply assume that using various information presentation formats, using realistic and vivid presentations and providing multiple possibilities for interaction with a learning system generally results in better learning. The acquisition of information via any technical system remains, however, subject to the constraints of human information processing. Existing models of multimedia comprehension originate from text comprehension models (e.g. Schnotz, 2005) or from research into the effect of illustrations (e.g. Mayer, 2001). One core concept

*Correspondence to: Valérie Gyselinck, Laboratoire Psychologie et Neurosciences Cognitives, Université Paris Descartes, CNRS FRE 2987, 71 av E. Vaillant, 92774 Boulogne-Billancourt cedex, France.
E-mail: valerie.gyselinck@univ-paris5.fr

found in these models is that of working memory (WM) and its limited capacity, which would explain some classical effects observed with multimedia material. We shall start by summarizing the bases of the main models of multimedia comprehension and describe these models. We will then focus on one of the classical multimedia effects, the so-called modality effect. This effect suggests that presenting some information in the visual modality together with information in another modality (auditory) is more beneficial to learning than presenting all the information in the same modality. Some experiments developed within the framework of Baddeley's model of WM with the dual-task paradigm will be presented. We will argue that these studies should be taken further since they may shed light on the modality effect. Finally, an original experiment which tested the modality effect and used the dual-task paradigm to assess the role of WM components in multimedia learning will be presented.

MODELS OF MULTIMEDIA COMPREHENSION

Most of the models of multimedia learning, based on research in educational fields, have taken existing research into language comprehension, and more specifically research into the role of illustrations in text comprehension, as their starting point. Indeed, one of the first and most common integration operations users have to undertake when working with a multimedia system is to process textual information (presented visually or auditorily) and integrate verbal information with pictorial information (e.g. a picture or movie).

Text comprehension

Theories of language comprehension commonly assume that texts or discourses are represented at different levels (Johnson-Laird, 1980, 1983; Kintsch, 1988; Van Dijk and Kintsch, 1983). Van Dijk and Kintsch (1983) considered that, first of all a word-for-word representation is derived from the surface structure of the text. Additionally, a propositional text base is derived. This is a representation which reflects the microstructure and the macrostructure of the text. Moreover, a situation model or a mental model is built. The concepts of situation model and mental model (Johnson-Laird, 1983) share many properties and are often used synonymously. They contain the information that is implicit in the text. The situation model can be seen as the product of the interaction between information provided by the text and knowledge of the world, including knowledge of the reader's objectives and attitudes. Inference processes play a central role in the construction of a mental model, which is both the by-product and the source of inferences. Some authors have gone on to consider that subjects' comprehension of a text and their construction of a mental model should be assessed through their ability to generate elaborative inferences (e.g. Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Tardieu, Ehrlich, & Gyselinck, 1992; Taylor & Tversky, 1992). The construction of a mental model is thought to be optional, depending on knowledge and the aim of the reader. It is, in addition, supposed to be critically dependent on WM capacity.

Baddeley's model of working memory

The model originally proposed by Baddeley and Hitch (1974) has become a dominant conception of WM in cognitive psychology. WM is generally defined as the dynamic

DOI: 10.1002/acp

control and coordination of processing and storage that takes place during the performance of complex cognitive activities such as language processing and visuospatial thinking (Miyake & Shah, 1999). In Baddeley's model (Baddeley, 1986, 1992), WM is thought of as a multicomponent temporary storage and processing system. A distinction is made between two storage subsystems: a verbal WM (VWM) and a visuospatial WM (VSWM). The VWM, that is, the phonological loop, keeps phonological entries active under the control of an articulated process. Visually presented verbal information is also transferred to the phonological loop via a process of articulatory rehearsal. The VSWM or visuospatial scratchpad is responsible for retaining visual and spatial information, and possibly for the formation and manipulation of mental images (Baddeley & Logie, 1999). These subsystems are coordinated by an attentional system of control known as the central executive. This multicomponent model of WM has proved successful in accounting for a wide range of data drawn from previous studies of short-term memory and appears to account for many processes in everyday cognition outside the laboratory environment (Logie, 1995). As far as text comprehension is concerned, several studies have emphasized the role of the central executive in WM (e.g. Oakhill, Yuill, & Parkin, 1986). The capacity of this central system, as measured by performance in a reading span task, is an important factor in the high-level psycholinguistic operations underlying comprehension (e.g. Just & Carpenter, 1992; Daneman & Merikle, 1996). The specific role of the storage systems has however rarely been considered until recent years.

In a new version of his model, Baddeley (2000) proposed the addition of a new component which is conceived of as a temporary store that retains complex information, manipulates it and uses it across a time scale which exceeds that of the presumed capacity of VWM and VSWM. Referred to as the 'episodic buffer', this system is thought to be able to integrate information from various sources under the control of the central executive. It can be thought of as an interface between a set of systems, each involving a different code, which uses a shared multidimensional code. It is in this episodic buffer that multimedia information such as graphical and verbal information could be considered to be temporarily stored and integrated.

The role of illustrations

In the field of research into the role of illustrations, several authors have attempted to explain the facilitative effect of illustrations on memory (see Carney & Levin, 2002 for a review). The most widely known of these studies is probably that conducted by Paivio (1971, 1986) which resulted in the proposal of a dual code theory. According to this theory, at least two coding systems are available: a verbal system and a non-verbal system. Pictures are thought to be memorized as such by the reader who benefits from two memory traces, one in a verbal form and one in a non-verbal form. This theory has proved very useful for explaining a large number of memory effects, such as the image superiority effect and the concreteness effect. On an initial interpretation, Paivio's theory might also be thought to explain the effect of text illustrations on memory. However, more recent research has examined the effects of illustrations on the by-products of comprehension and not just on memory (see, for example, Fletcher & Tobias, 2005 for a review). In this respect, it is not clear how the dual code theory can account for the influence of illustrations on text comprehension. Indeed, it makes no assumptions concerning the construction of mental representations during the reading process, or about the representational levels at which the interaction between the text and the illustrations occurs.

DOI: 10.1002/acp

The beneficial effect of illustrations on comprehension, when not explained in terms of the dual coding view, is often interpreted in terms of the properties of graphics which are thought to have an effect on WM. Some authors consider that illustrations act as an external memory (e.g. Larkin & Simon, 1987). Others are of the opinion that, because diagrams and graphics are usually more concise than equivalent textual statements and because the essential information tends to be perceptually clear, illustrations can reduce the cognitive load associated with complex reasoning tasks (e.g. Marcus, Cooper, & Sweller, 1996). In other words, the advantage of illustrations, as well as of other iconic modes of representation is that they make structural relations more transparent. As such, illustrations are easier to process than the corresponding text statements, thus facilitating the understanding of the situation described (and depicted). This point relates to a third theoretical perspective which is based on the theory of mental models. A specific element of Johnson-Laird's theory, as opposed to that proposed by van Dijk and Kintsch, is the notion of homomorphism to the world: a mental model has a structure which is analogous to that of the situation it represents, and its content corresponds to the objects and events of the world. Therefore, because of its analogical structure, a mental model resembles a mental image of this world.

Several authors mainly originating from the field of language comprehension (Glenberg & Langston, 1992; Gyselinck & Tardieu, 1999; Hegarty & Just, 1993; Kruley, Sciama, & Glenberg, 1994) have suggested interpreting the facilitatory effect of illustrations within the framework of Johnson-Laird's theory (1983). An illustration is also an analogical representation, albeit an external one. When pictures illustrate the content of the text they accompany, they closely mirror the situation described in the text. A picture can thus be viewed as one possible expression of a mental model. Consequently, pictures could act as a transitional step in the process of transforming prose into mental images which subsequently evolve into a mental model. Second, because pictures are able to suggest an appropriate model for a text, they could be used as a guide for the construction of the model. That is to say, they could act as scaffolding on the basis of which readers could easily encode entities and relations from both the text and the picture. A large number of studies investigating spatial representations have confirmed this view (e.g. Bower & Morrow, 1990; Gray-Wilson, Rinck, McNamara, Bower, & Morrow, 1993).

Since recent years, the focus of many searchers in the educational field has moved from the study of text and pictures to the study of multimedia in the learning process. The existing views on the integration of texts and pictures have been taken as a basis of more complex multimedia learning.

COGNITIVE MODELS OF MULTIMEDIA LEARNING

The most influential models that have been proposed in the field of multimedia learning are the cognitive load theory (Paas, Renkl, & Sweller, 2003; Sweller, 1999; Sweller, van Merrienboer, & Paas, 1998), the cognitive theory of multimedia learning (Mayer, 2001, 2005), and the integrated model of text and picture comprehension (Schnotz, 2005).

Cognitive load theory

Sweller (1994, 1999, 2005) has developed a theory, known as the cognitive load theory, which attempts to interpret the limitations of the human information processing system

DOI: 10.1002/acp

within the framework of instructional procedures and learning. Basically, the cognitive load theory is like classical modular theories in that it postulates an architecture composed of a limited capacity WM and an unlimited capacity long-term memory. The limitations of WM are considered to represent limitations to the cognitive resources which a learner may devote to the conduct of a given task. The activation of existing knowledge in long-term memory is not subject to the same limits (Ericsson & Kintsch, 1995; Sweller, 2005). Knowledge is organized into more or less complex schemata. It is the activation of a structured schema in memory, not the sum of its elements that characterizes expertise while simultaneously reducing the memory load imposed on experts in any given domain. These schemata are complex cognitive constructs which allow learners to categorize information in simple, easily retrievable units. With practice and time, the cognitive processes required to complete a task become increasingly automated. Over time, these processes may become fully automatic, thus freeing-up cognitive resources for other activities. In this theory, learning is thought to be based on two mechanisms: schema acquisition and automation (Sweller, 1994).

This theory distinguishes between three sources of cognitive load. The intrinsic cognitive load is related to the difficulty of the task. It is a function of the number of elements to be processed simultaneously in WM as well as of the level of interaction between them, that is to say the number of reciprocal relationships. The second source of cognitive load, the extraneous cognitive load, refers to the supplementary processes required in the case of an inappropriate presentation format. Finally, the germane cognitive load relates to the elaboration of knowledge and corresponds to the efforts involved in the learning process, such as the construction of knowledge schemata or the discovery of solutions to problems. The effects of these sources of cognitive load are additive in any learning situation. As a consequence, the use of an inappropriate presentation format will be especially detrimental in cases where the intrinsic load is high because of the difficulty of the task and/or the lack of domain knowledge. Indeed, the learning process continues to be constrained by the limited capacity of WM. Performance—in terms of time or efficiency—will be impaired compared to situations in which the format of presentation is optimal. In contrast, situations in which the extraneous cognitive load is reduced might permit an increase in the effort devoted to the elaboration of schemata in memory. This, in turn, will progressively reduce the intrinsic load as a result of the increase in expertise, thus freeing-up resources for processes useful to learning (Sweller et al., 1998).

The cognitive theory of multimedia learning

The generative theory of multimedia learning basically relates to how learners integrate verbal and visual information during learning. Learners are viewed as ‘knowledge constructors’ due to the multiple pieces of information they have to integrate. This theory, currently referred to as the ‘cognitive theory of multimedia learning’ (Mayer, 2001, 2005), is based on the idea that there are two separate information processing systems for auditory–verbal and visual–pictorial information. Two ways of conceptualizing the differences between the two channels or information processing systems are usually considered, one based on sensory modalities (auditory vs. visual) and the other based on representation modes (verbal vs. pictorial). Mayer has opted for a compromise, considering both types. In the generative theory, learning consists of establishing referential links between the verbal and pictorial representations. Each of these channels has a limited processing capacity at any given time. Ultimately, meaningful learning requires the

DOI: 10.1002/acp

construction of a mental model of the document on the basis of three processes: selection of the important elements in the presented material, the organization of these elements within a coherent structure and their integration with existing knowledge. Each process operates, at least initially, in WM.

Mayer's conception of WM is based on assumptions taken from Paivio's (1986) dual-coding theory and Baddeley's concept of WM (1992). According to Mayer, WM in the domain of multimedia learning involves two limited capacity stores, that is, an auditory store and a visual store which are analogous to the phonological loop and the visuospatial sketchpad proposed by Baddeley. Visually presented information is processed in visual WM, whereas auditorily presented information is processed in auditory WM. In the case of a text presented visually, it is initially processed in the visual channel, and after it has been converted into sounds, it is processed through the auditory channel. Information is organized independently in each WM space before being fused. WM plays a crucial role in the accomplishment of the three processes that govern multimedia learning. WM is involved in the maintenance of relevant information in each store, in the transformation of the information present in each store into a coherent representation, and in the establishment of connections between visual and verbal representations.

The integrated model of text and picture comprehension

A first version of the integrated model of text and picture comprehension was presented in 1999 (Schnotz, Böckheler, & Grzondziel, 1999) before being extended (Schnotz, 2001; Schnotz, 2005; Schnotz & Bannert, 2003). It was inspired by classical studies of text comprehension, and more specifically the theory of mental models. Readers build a surface representation of the text from which they then extract a semantic propositional representation which, in turn, is used to build a mental model. The propositional representation and the mental model interact with and enrich one other. The mental model is constructed on the basis of the propositional representation and results in the inclusion of new information in this.

As far as illustrations are concerned, perceptual processes are thought to result in a pictorial representation in memory. These are low-level sensory processes up to the point at which the elements are identified and organized in the diagram. The processes responsible for these latter operations are influenced by the reader's cognitive schemata. This stage is followed by a process of semantic integration which is dependent on the knowledge and aims of the individual in question. The construction of a mental model from a diagram is an analogical process of mapping visuospatial relationships to semantic relations. The mental model is different from the perceptual representation because it is more abstract, depends on a task-guided selection process, involves prior knowledge, and does not depend on the modality of presentation.

The model was recently extended to take account of multimodality in the processing of multimedia documents (Schnotz, 2005). The model now explicitly considers the various sensory modalities and memory systems and refers to the models proposed by Atkinson and Shiffrin (1968) and Baddeley (1986, 1992). It is also possible to establish a link with the dual coding theory. The difference between the dual coding theory and Schnotz's model, however, is that verbal and analogical representations are systematically constructed both in the case of verbal and illustrated material.

The integrated model of Schnotz has thus various assumptions in common with the model of Mayer. Both assume a cognitive architecture including a WM system of limited capacity, and they assume different channels for processing and storing of verbal and pictorial information. They nevertheless differ because the integrated model assumes that verbal as well as pictorial information are not necessarily associated with a specific modality, but can be also conveyed by other sensory modalities. Schnotz makes an explicit distinction between representational channels on the cognitive level and sensory channels on the perceptual level.

The three models of multimedia learning presented here all share a conception of WM inspired by the work conducted by Baddeley. The basic characteristics of Baddeley's model, that is, the existence of two distinct, limited capacity peripheral systems, are usually considered to explain the difference in learning efficacy of different types of multimedia learning material, to which we will turn now.

MULTIMEDIA LEARNING AND THE WORKING MEMORY CONSTRUCT

The modality effect

Even though illustrated written documents are relatively effective, they may constrain readers to share their attention between different sources of visual information if these are not comprehensible in isolation. This consequently imposes a high cognitive load on the visual channel (Sweller, van Merrienboer, & Paas, 1998). This phenomenon is referred to as the 'split visual attention effect'. The split visual attention effect (Ayres & Sweller, 2005 for a review) can be reduced if elements in the text are moved to the corresponding locations on the illustration (Chandler & Sweller, 1991, 1992; Kester, Kirschner, & van Merriënboer, 2005; Moreno & Mayer, 1999; Sweller & Chandler, 1991; Sweller, Chandler, Tierney, & Cooper, 1990) or if pop-up windows are integrated in the diagram (Bétrancourt & Bisseret, 1998; Erhel & Jamet, 2006). According to the three theoretical models cited above, in split-attention situations, complex visual research processes create an overload in visual WM and compete with the learning processes for cognitive resources.

To avoid the split visual attention effect, it is possible to use the oral mode to explain a diagram in order to reduce the overload in the visual channel ('the modality effect;' Mayer, 2001; Mayer & Moreno, 1998). In a series of experiments, Mousavi, Low, and Sweller (1995) studied the resolution of geometrical problems in which graphics were presented together with an explanation of the solution. The learning phase consisted of the presentation of examples together with their solution. The graphic was presented together with an orally stated solution, a solution presented in writing or a solution presented in both modalities. After the learning phase, the subjects had to solve a new series of exercises. In the first experiment, the time required to find the solution was shortest when an audiovisual presentation was used in the learning phase (diagram plus oral solution). The joint use of the two modalities in the learning phase was equivalent to the use of the visual modality only. The authors assumed that when the solution is presented visually, even though this solution is also presented in the oral mode, the text is processed visually by the learner and results in the splitting of attention. The same results were obtained in the other experiments reported in the paper as well as in other experiments involving other materials (see Low & Sweller, 2005 for a review).

In a recent meta-analysis of the modality effect, Ginns (2005) compared 43 results. A modality effect has been observed in association with the verbal explanations accompanying graphics and diagrams as diverse as electrical diagrams (Tindall-Ford, Chandler, & Sweller, 1997), geometrical problems (Jeung, Chandler, & Sweller, 1997),

DOI: 10.1002/acp

temperature graphs (Leahy, Chandler, & Sweller, 2003), abacus diagrams (Kalyuga, Chandler, & Sweller, 2000) or illustrations of the fusion of solids (Kalyuga, Chandler, & Sweller, 1999). It has also been observed in the context of more complex multimedia presentations such as animations (Mayer & Moreno, 1998; Moreno & Mayer, 1999), multimedia presentation of the formation of lightening (Craig, Gholson, & Driscoll, 2002) or virtual learning environments in the botanical field (Moreno & Mayer, 2002). In all these experiments, the participants were young adults (university students or apprentices) and the measured performance was the knowledge transfer ability studied in the form of transfer problems, inferences or the application of knowledge to new examples.

Ginns (2005) compared these results in the light of three hypotheses. The first is that the oral presentation of explanations accompanying visual information such as another source of written information, a diagram or an animation will be more propitious for learning than a visual presentation (the modality effect). Second, this effect will depend on the complexity of the material (i.e. the number of links to be established between elements of the document) (e.g. Sweller, 1999) and third, on the learner's ability to control the pace of presentation of the information. These three hypotheses were verified. The modality effect has a mean amplitude (cohen's $d = 0.72$), falls with the complexity of the material and diminishes or even inverts when the presentation is self-paced.

Explaining the modality effect with WM

The modality effect is thus obtained in various instructional situations where verbal information is combined with graphical information. It is generally explained in terms of the better allocation of resources in each WM modality (Kalyuga et al., 1999) as well as through the contiguous processing of the various sources of visual and oral information (Mayer, 2001). However, the load associated with WM is rarely controlled in the various experiments reported here, or is controlled only in terms of the subjective judgments of the participants (e.g. in some of the experiments conducted by Sweller and co-workers). A rating of the participants' perception of the mental effort associated with learning is not, however, a direct measure of memory load and is not independent of the task. In other words, a learner may have the feeling that one task requires a greater cognitive effort than another, but this effort may not be a direct reflection of the actual load in memory. The differences in effort could equally well correspond to differences in the learner's interest or motivation with regard to each task.

In the work of Mayer and co-workers, visual WM is thought to be overloaded when materials are presented in the same mode. However, nothing directly indicates that the visual component of WM is overloaded. Other explanations could therefore be proposed. One is that the simultaneous presentation of verbal information and graphical information stimulates the visual channel at the same time. If the information is presented too quickly, readers might find it difficult to encode it all. In such a case, the problem would not lie in the limits of storage or processing in VSWM, but rather in an earlier stage of the whole process (perceptual). In that respect, the distinction between sensory channels on the perceptual level and representational channels in WM as in Schnotz' model seems more appropriate than the compromise proposed by Mayer. It is also possible that these effects could be explained in terms of the division of attention, thus pointing to a problem relating to executive functions and WM control. Whatever interpretations might be proposed, they are only of interest if integrated within a coherent theoretical construct.

DOI: 10.1002/acp

The fact that we base our interpretations on a precise model of WM and are able to conduct experiments in which the control and even the manipulation of WM memory load (storage and/or processing) are possible, should enable us to specify the processes involved in this stage of the integration of information coming from multiple sources. We therefore think that there is some benefit in using Baddeley's WM model as a framework because it is extremely heuristic, at least in the initial versions to which we will refer here.[1] Specific experimental paradigms, such as the dual-task technique, have been developed on the basis of this model. The WM model proposed by Baddeley and Hitch (1974) originates in part from the dual-task paradigm. The rationale is that if two tasks involve the same cognitive processes or compete for the same limited resources, it will be impossible to perform the two tasks together as well as each task in isolation. In parallel with a primary task, the subject is then asked to perform a secondary task which mobilizes processes which are also thought to be involved in the primary task. The interference value is thus the performance decrement observed in the dual-task condition compared to the single-task condition. We argue that dual-task techniques could provide more direct proof of the involvement of verbal or visual WM or overload phenomena and might thus help us gain an understanding of certain multimedia effects, and in particular the modality effect (Tardieu & Gyselinck, 2003; Tabbers, Martens, & van Merriënboer, 2004).

The dual-task paradigm in complex information processing

Until recently, the use of the dual-task paradigm was rare in experiments designed to investigate the role of the storage components of WM in complex cognitive activities such as text comprehension. It should be of interest for our understanding of the comprehension processes and also for verifying the validity of the WM model in such activities. The on-line use of visuospatial WM during the comprehension of illustrated texts was examined by Kruley et al. (1994). They claimed that illustrations facilitate the construction of a mental model, and that this construction takes place in the VSWM. Texts containing spatial descriptions of an object (e.g. a volcano), a part of an organism (e.g. a leaf) or a mechanical device were presented auditorily. A single picture displaying the structural relationships between the parts of the objects described in the text could be associated with the text. In a first experiment, a concurrent task required the subjects to maintain a dot pattern in VSWM while simultaneously listening to the text and a verification test was performed after the presentation of each sentence with a new configuration. In the control condition, the subjects were not required to remember the visual configuration of dots and simply had to make a judgment concerning it. Text comprehension was tested at the end of the text. The hypothesis was that if processing the pictures involves VSWM, this processing should compete with the maintenance of the configuration (the 'pre-load' material) in WM, thus reducing the benefit of the illustration on comprehension. The results showed that comprehension was facilitated by the presence of the picture and was impaired in the 'pre-load' condition. There was, however, no interaction between the two variables. The interaction hypothesis was nevertheless confirmed by the concurrent task data. Performance was higher in the control condition than in the pre-load condition, and

[1]Although the addition of the episodic buffer (Baddeley, 2000) seems to be a tempting way of explaining multimedia comprehension, we are not, at our current stage of research, fully convinced by it. With this addition, the WM model loses some of the heuristic value which it gained from its simplicity. In general terms, there is no clear way to account for this buffer within an experimental environment. To date, no experimental paradigm has been developed to assess its existence.

DOI: 10.1002/acp

the presence of a picture resulted in a fall-off in performance only in the 'pre-load' condition. When a non-visual concurrent task involving the retention of digits was used, no interaction effect either on comprehension or on concurrent task data was observed. These results were considered to indicate that the interference observed with the configuration of dots was specific to visuospatial processing.

In a series of experiments, Gyselinck, Cornoldi, Dubois, de Beni, and Ehrlich (2002), and Gyselinck, Ehrlich, Cornoldi, de Beni, and Dubois (2000) investigated the involvement of visuospatial WM and the phonological loop in the comprehension of short scientific texts presented on a computer screen and accompanied by illustrations. The same six texts were used in three experiments. Each text summarized a physical concept (static electricity, electrolysis, gas pressure, etc.) in nine sentences which were successively presented on the computer screen. Each sentence was or was not accompanied by an illustration. While reading, the subjects had to perform tasks involving either the visuospatial sketchpad or the phonological loop.

It is now generally accepted that a task such as articulatory suppression (repetition of a series of digits or syllables) competes with the retention of phonological information in VWM, whereas a task such as spatial tapping (continuous tapping of a series of keys or buttons) competes with the retention of spatial information in VSWM (see Farmer, Berman, & Fletcher, 1986 for clear selective interference effects in verbal and spatial reasoning tasks). Articulation is thought to dominate the process of articulatory control and thus prevent its use either for maintaining elements already present in the phonological store or for converting visual elements into a phonological code. The planning and execution of tapping is thought to disrupt the maintenance of information in VSWM.

In Gyselinck et al. (2002), the subjects performed these concurrent tasks together with the texts or with illustrated texts. In the control condition, the subjects had no task to perform. Their comprehension of the phenomena described was tested by means of paraphrase and inference questions. The comprehension performance results revealed that, as expected, the concurrent tapping task eliminated the beneficial effect of the illustrations, while the concurrent articulatory task impaired performance equally in both presentation formats.

To ensure the effect of the concurrent articulatory task was not a general effect due to a decrease in attention or any other general mechanism, an additional experiment was conducted. The presentation of texts alone (text-only) was compared with a format in which the illustrations were presented alone (illustrations-only) together with certain required labels (consequently verbal processing was reduced). The results showed that the concurrent articulatory task selectively impaired comprehension in the text-only format. That is to say, an interference effect was obtained in the text-only format but not in the illustrations-only format. This result suggests that the interference is specifically related to phonological memory. As far as the effect of the concurrent tapping task on the processing of illustrations presented alone is concerned, no significant selective impairment was observed in this format. However, the variation in the mean values pointed in this direction. Thus, processing illustrations presented alone does not seem to involve VSWM to the same extent as processing illustrations presented together with a text. These results show that the processing of pictorial information involves VSWM and that the processing of texts involves VWM.

This conclusion is also supported by the results obtained by Brunyé, Taylor, Rapp, and Spiro (2006). They conducted similar experiments with procedural texts. An advantage of the picture-only and text-plus-picture format over a text-only format was observed in the

DOI: 10.1002/acp

control condition. Interference effects obtained with diverse secondary tasks suggest that the subcomponents of WM are differentially involved depending on the presentation format. The visuospatial secondary task interfered with picture-only processing whereas the articulatory task interfered with text-only processing. In addition, both a visuospatial and a verbal central task exclusively interfered with multimedia processing, suggesting that central resources are necessary for the integration of texts and pictures.

A number of studies have shown that spatial visualization ability influences the comprehension of texts and learning (Friedman & Miyake, 2000; Haenggi, Kintsch, & Gernsbacher, 1995; Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997; Mayer & Sims, 1994). Mayer and Sims (1994) used rotation tasks to examine subjects' spatial ability which is thought to be dependent on both storage and processing of VSWM. The results showed that high spatial ability subjects benefited more than low spatial ability subjects from the concurrent presentation of a text and a visual animation compared to a condition in which these were presented in succession. Individual differences were therefore also considered in the studies conducted by Gyselinck and co-workers. In Gyselinck et al. (2000), the spatial span measured with the Corsi-blocks test (Milner, 1971) determined the subjects' ability to integrate verbal and pictorial information and to benefit from illustrations. Gyselinck et al. (2002) showed that individual differences in storage capacity, as assessed by spatial and verbal spans, define the extent to which the two subsystems of WM—the VSWM and the VWM—are involved in the integration of texts and illustrations, respectively.

In another study, De Beni, Pazzaglia, Gyselinck, and Meneghetti (2005) used the dual-task paradigm to investigate the role of the two WM subsystems in the comprehension and memory of spatial and non-spatial texts. A description of a route and a description of the process of wine production were presented auditorily to subjects. A selective interference effect of the spatial concurrent task on the spatial text and an interference effect of the verbal concurrent task on both the spatial and non-spatial texts were observed. Finally, in a recent study, Gyselinck, De Beni, Pazzaglia, Meneghetti, and Mondoloni (2007) manipulated the instructions given to subjects concerning the processing of visually presented spatial texts. The results confirmed the involvement of VWM and revealed that imagery instructions are more beneficial than repetition instructions and that this advantage vanishes with a concurrent tapping task. Once again, the results were partly dependent on spatial WM capacity as measured by the Corsi Blocks Test. High spatial span subjects benefited from the imagery instruction and specific spatial interference effects were observed, whereas the pattern was less clear with low spatial span subjects.

Overall, these studies demonstrate the involvement of the storage component of VSWM as soon as visuospatial information has to be processed, whether this information is presented visually (as in the case of pictures) or whether it has to be reconstructed from knowledge or by means of imagery procedures (as in the case of spatial texts). They also reveal the involvement of the storage component of VWM in the processing of verbal information, whether it is presented visually or auditorily.

One limitation, however, of the studies reported here in terms of the issues addressed by our research is that even though several types of information have been studied—verbal information (presented orally or visually), verbally presented visuospatial information, and graphical information—this pictorial information has a great degree of informational equivalence with the verbal information, a situation which most certainly does not exhaust the capabilities of multimedia systems. A picture can indeed illustrate the content of a text and be considered as redundant with it, or it can complement the text, showing some

DOI: 10.1002/acp

relationships which are not expressed verbally. In addition, in all the studies described, only one modality was considered, either visual or auditory, but the two have not been combined. This means that we cannot verify whether the modality effect can be explained in terms of an allocation of resources in WM.

WORKING MEMORY AND THE MULTIMEDIA EFFECT

Although it has proven useful for considering the limitations of WM, few experiments have used the dual-task paradigm in the multimedia domain. Brünken, Plass, and Leutner (2003) and Brünken, Plass, and Leutner (2004) have defended the view that using dual-task methodology could allow for the direct assessment of cognitive load during multimedia learning. The effect of the dual-task on the performance on the secondary task is considered a direct measure of the cognitive load induced by the primary task. In Brünken, Steinbacher, and Leutner (2002), the primary task was to learn an anatomy document consisting of an illustration and a visually or auditorily presented text. The secondary task consisted of pressing the space bar whenever the letter 'A' present at the bottom of the screen turned red. Reaction times were longer in the dual-task condition than in a single-task condition, showing an interference effect. They were also longer when the document was visual only than when it was audiovisual. Brünken et al. (2002) concluded that an audiovisual document imposes a lower load on WM than a visual document. The point, however, is that only one document is presented, with alternatively a page in the visual mode and the following in the audiovisual mode, which is a very specific way to compare the two modes of presentation of the same content. In addition, the secondary task presumably calls primarily on visual attentional resources and the observed interference therefore provides little information about the load imposed on the visual channel.

We therefore conducted an experiment to extend previous results indicating the specific involvement of VWM and VSWM in the processing of verbal and visuospatial information to more complex and longer multimedia material. First of all, we expected a modality effect to occur, that is, the presentation of an auditory text should be more beneficial than a visual text when these texts have to be processed together with graphics. In addition, if the cognitive principles involved in multimedia processing relate to the use of available WM capacities, then a concurrent task which overloads VSWM should impair multimedia comprehension more severely in the visual than in the auditory text presentation. With a verbal concurrent task, however, both the visual and the auditory processing of the text should be disrupted as the text has to be processed verbally in both cases.

Method

Participants
Fifty-six first and second year students at a school of engineering (Télécom Paris and related schools) participated voluntarily. Their age ranged from 19 to 23 years.

Material

Multimedia documents
Of several multimedia documents developed on CD-ROM specifically for use by students at the Telecom schools as part of their electronics courses, six topics were selected (e.g.

DOI: 10.1002/acp

'Can inverters be used to design high-performance timers?'). The documents were modified for the purpose of the present experiment. They contained explanatory texts which were combined with diagrams and certain verbal indications intended to direct students' attention to a given point. Two versions were constructed for each topic. In one version, the texts as well as the verbal indications were presented auditorily. In the other version, which was closer to the original, the texts were presented visually and the verbal indications remained auditory. In both versions, static diagrams were also presented. The graphical information was not redundant with the texts but helped explain the concept and both sources of information had to be processed if the participants were to understand the concept.

Each document was divided into four or a maximum of five screen pages, including an introduction and a conclusion. A page could contain two or three graphics or diagrams which could be presented on the left part of the screen with the text on the right part, or in the centre with the text at the top and/or the bottom of the page. Each piece of information of a page (an item of text or a diagram) was presented incrementally until the last piece of information had been presented and then remained on screen for long enough to permit processing (3 seconds for a diagram or graphics and 300 milliseconds per word for text). In the auditory modality, all the texts were recorded by the same speaker, read at a normal rate, and the space previously occupied by the text was left blank in the document. The suitability of the presentation times had been assessed in an earlier study run with the same material (Dubois, 2004). At the end of each page, a jingle was played to make the subject click on an icon to move to the following page. The entire presentation of a document took no more than 10 minutes.

Comprehension questions

Eight triple choice inferential questions were constructed for each document on the basis of the final exercises proposed in the original version of the multimedia document. The false alternative choices were chosen on the basis of the false answers given by subjects in an earlier study in which the documents were assessed for an equivalent level of difficulty (Dubois, 2004).

Design

The design was mixed and involved two factors for the comprehension task data: the presentation modality of texts (visual, auditory) as a between-subjects factor and concurrent task (control, spatial tapping and articulatory) as a within-subjects factor. The order of the concurrent tasks was counterbalanced across subjects, and the questions in the comprehension test appeared in a random order for each document and subject.

Procedure

The subjects were tested individually and the session took about 2 hours. In the first phase they first performed the Corsi Blocks Test (to measure spatial span), and the digit span test. The experimenter sat opposite the subject and pointed to a gradually increasing sequence of blocks arranged on a board between them. After the presentation of each sequence, the subject had to repeat the sequence in the correct order. The subjects also performed a digit span test that measured their phonological memory. They were orally presented with an increasing sequence of digits which they then had to repeat.

In the second phase, the subjects were told that the experiment required them to visualize, read or listen, and understand the notions of electronics presented in the

DOI: 10.1002/acp

sequence, while in some cases performing a concurrent task, and then answer questions. Most subjects were randomly assigned to one or other of the modality groups. However, in order to balance the two groups for the mean scores on the two span tests, some subjects were preferentially assigned to one of the modality groups. The mean scores were: for spatial span, $M = 5.5$, $SD = 0.7$ in the visual format, $M = 5.5$, $SD = 0.8$ in the auditory group; and for digit span, $M = 7.7$, $SD = 1.3$ in the visual group, $M = 7.5$, $SD = 1.5$ in the auditory group. No difference was significant ($t < 1$).

To complete the experiment, each student worked at a computer with a connected headset to hear the jingles, the verbal indications and the texts in the auditory format. The presentation of two successive documents was associated with a concurrent spatial tapping task, a concurrent articulatory task, and a control condition presented in a counterbalanced order. The subjects were trained to perform the concurrent tasks at the correct rate before the experiment was run in that condition.

In the spatial tapping task, they had to move their hands anticlockwise between four buttons on a surface. Tapping was performed using the subject's dominant writing hand, and the subjects had to tap at about 1 second per button with the surface placed below the table, not visible to the subject. In the articulatory task, subjects had to repeat the sequence 'ba, be, bi, bo, bu' at a rate of about 1 syllable per second. The concurrent spatial and articulatory tasks had to be performed for the entire reading/visualization period. In the control condition, no concurrent task was required. The order of presentation of the six documents was partially balanced, so that each document was presented equally often in conjunction with each of the concurrent tasks in each modality of presentation.

At the end of each document, the comprehension questions appeared one at a time in a new random order for each subject. The question appeared in the top left part of the screen and the response choices appeared at the bottom. The subjects had to click with the mouse to give their answers. No time constraint was imposed for answering the questions.

Results

Comprehension question accuracy

For each question, a correct answer was assigned a score of 1 and any false alternative chosen was assigned a score of 0. The total of correct responses to the questions in each concurrent condition was calculated for each subject (maximum 16). The questions resulting in performance below chance level (one-third) in the control condition were not included in the analyses and replaced by the performance obtained by the subject for the same item in the second document of the same condition. Out of 24 questions, 5 required this type of change. Table 1 presents the scores and standard deviations for each modality of presentation in each of the three concurrent task conditions.

An omnibus ANOVA was run with modality (visual vs. auditory) as a between-subjects variable and the concurrent tasks (control vs. spatial tapping vs. articulatory) as a within-subjects variable.

No modality effect was found, even in the control condition, that is, performance was similar whether the format of presentation of the verbal information was visual (75.0% correct responses) or auditory (73.4%) (Fs < 1), contrary to our expectations. An effect of the concurrent tasks was observed ($F(2,54) = 10.86$; $p < .0001$), but there was no interaction with the modality ($F < 1$).

Regarding the contrast between control and concurrent articulatory conditions, the effect of task was significant, $F(1, 54) = 18.72$, $p < .0001$, revealing an interference effect of the

DOI: 10.1002/acp

Table 1. Means and standard deviations for the comprehension questions scores in each of the three concurrent task conditions for the visual and the auditory modality of presentation

	Control	Spatial	Articulatory
Visual			
M	12.00	11.11	10.21
SD	2.49	2.04	2.26
Auditory			
M	11.75	11.18	9.57
SD	2.10	2.23	3.40

articulatory task (74.22% in the control condition and 61.61% in the articulatory task condition). As can be seen in Table 1, this interference effect was, as expected, similar in the two presentation formats—visual and auditory—as there was no interaction between tasks and format ($F < 1$).

Regarding the contrast between the control and concurrent tapping task, the task effect approached significance, $F(1, 54) = 3.67$, $p = .057$. However, contrary to our hypothesis, we observed no interaction between tasks and presentation format ($F < 1$). As Table 1 shows, the concurrent tapping task tended to interfere with comprehension (74.2% in the control condition vs. 69.6% in the tapping task condition) to a similar extent in the two presentation modalities.

Conclusions

In this experiment, a dual-task procedure was used to explore the involvement of WM components in the comprehension of complex multimedia documents. This involvement was expected to explain the classical 'modality effect', which suggests that presenting a text in the auditory channel together with other information (such as graphics) in the visual channel, leads to a lower cognitive load and results in better comprehension than presenting a text in the visual format together with visuospatial information.

Contrary to our expectations, no modality effect was found in the control condition, that is, performance was similar for both the visual and auditory presentation of the verbal information. Nevertheless, the dual-task paradigm we used to induce cognitive load indeed produced interference effects on comprehension performance. The concurrent articulatory task caused a fall-off in performance. This result shows the involvement of the VWM in multimedia learning and suggests that the VWM stores verbal information as efficiently when it is presented visually or auditorily. Contrary to our expectations, however, the concurrent tapping task interfered with comprehension to an equal extent in both the presentation formats.

We have to explain the lack of modality effect, and the fact that no specific interference effect of the tapping task is obtained in the visual modality. At first sight, it would therefore appear that the modality effect does not depend on the VSWM load as the dual-task paradigm indicates. If one accepts the null hypothesis, one possibility would be to question whether the modality effect can be generalized. On the basis of this latter study, we could conclude that there is no advantage in sharing resources between the visual and the auditory channels when dealing with this type of multimedia material. This hypothesis seems to run

DOI: 10.1002/acp

counter to most of the effects observed in various experiments such as those reported in the meta-analysis conducted by Ginns (2005). In addition, the same categories of subjects were tested (young adults) and the same kind of performance was considered (inferencing).

Indeed, Ginns (2005) found that the modality effect depends on the complexity of the material and on the possibility for the learner to control the pace of information presentation. The modality effect can sometimes be reversed when learners have been able to monitor the pace of presentation of the screen-pages (Tabbers, 2002; Tabbers, Martens, & van Merriënboer, 2000, 2004). The authors suggest that, when there is a written presentation, the text is additionally processed in the phonological loop and that a simple explanation of the modality effect in terms of the allocation of WM resources is inadequate. In the studies using a dual-task procedure, indeed, the involvement of VWM was proven, whether the text was visual or auditory. For Tabbers and co-workers, in the absence of any possibility of controlling the incoming information, an oral explanation permits the joint processing of the visual and auditory sources, whereas the exclusive presence of visual sources would result in multiple attentional shifts between sources. However, in the experiment presented here, subjects could only monitor whether they reached the next screen page and were asked to click on the appropriate button as soon as the whole screen page had been displayed. There was no great opportunity to navigate flexibly between different sources of information. It is thus unlikely that the lack of modality effect can be only accounted for by the subjects' control.

Another possibility would be to reject the null hypothesis and to question the homogeneity of the data obtained. Previous studies have demonstrated the importance of taking account of individual differences in storage capacities (Gyselinck et al., 2000, 2002, 2007). High spatial span subjects benefit from illustrations when processing expository texts and from imagery instructions when processing spatial texts and their performance is selectively impaired by a spatial tapping task, whereas low span subjects do not exhibit this pattern of results. In the case of multimedia documents such as those used here, we can thus also predict that the involvement of the storage components of WM depends on the individual capacities of these components. We then considered individual differences in storage capacities. In previous studies utilizing contrasted groups (e.g. Gyselinck et al., 2002), subjects were selected on the basis either of their spatial or their verbal span, depending on the focus of interest. In the present experiment, we wanted to test the modality effect which suggests that the verbal channel and the visuospatial channel can be used separately without overload. We were thus interested in both capacities but wanted to avoid a confound between the two capacities or a confound with a more general capacity. Consequently, in order to ensure that the subjects possessed a high capacity which was specific to a WM component and thus that any effect would be attributable only to one specific storage capacity, we selected subjects with a high span score in one domain (verbal or visual) but not in the other domain.

Unfortunately, only very few subjects (from 5 to 9 subjects in each sub-group) were eligible for analysis and no statistics were calculated. However, the observed performance patterns revealed an intriguing and very interesting result. The patterns for high spatial and high verbal subjects perfectly mirrored one another. Whereas a reverse modality effect was observed for the high spatial subjects, a modality effect was observed for the high verbal subjects. Is it thus possible that the lack of a modality effect for the group as a whole was simply due to the fact that the two inverse effects cancelled each other out? If this was the case, we could conclude that, when faced with this type of material, subjects are confronted with processing difficulties and monitor the various sources of information differently

DOI: 10.1002/acp

depending on their storage capacities. In addition, when we consider the interference induced by the concurrent tasks, high spatial span subjects seem to be sensitive to both concurrent tasks in the auditory format but not in the visual format whereas, in contrast, high verbal span subjects seem to be sensitive to both concurrent tasks in the visual but not in the auditory format. This indicates that depending on her/his storage capacities, a format might be optimal for an individual, but not for the other, therefore reducing the extraneous cognitive load described by Sweller and co-workers. These observations suggest that even though the dual-task paradigm has not enabled us to clearly prove any specific involvement of VWM and VSWM in explaining the modality effect, the explanation may nevertheless relate in some way to the storage components of WM.

CONCLUSIONS AND DISCUSSION

The current article has presented a survey of the most influential models of multimedia learning based on existing models of text comprehension and research into the role of text illustrations. One core concept in the three-presented models is that of a limited WM capacity which may explain some of the classical effects obtained when subjects have to understand and learn on the basis of multimedia material. We have focused on the so-called 'modality effect' which corresponds to the fact that an auditory presentation of verbal information is better than a visual presentation when combined with pictorial information as it makes it possible to reduce the cognitive load in the visual channel. The supposed memory overload is, however, often inferred from poorer performances in one case than in the other, but it has rarely been controlled for and even more rarely directly manipulated. Our argument is that conducting experiments based on a precise model of WM and involving the ability to control and even manipulate the cognitive load in WM (storage and or processing load) should make it possible to identify the processes involved during the integration of information coming from multiple sources. To this end, we chose Baddeley's model (1986, 1992). We then presented a series of experimental studies which use the dual-task paradigm.

Overall, the presented experiments show that VWM is involved in the comprehension of complex verbal information such as texts, whether they are presented orally or visually. Even in cases where subjects have to navigate between different types of information and have to integrate various items of information, the verbal storage component of WM is important in permitting a processing operation that is as complex as comprehending. The results also show that VSWM is involved, insofar as visuospatial information has to be processed, irrespectively of whether comprehension involves visuospatial information that is presented in the form of illustrations of a text or conveyed by the text itself.

These results confirm the view that the WM is concerned primarily with representational channels (verbal and pictorial information) instead of sensory channels (auditory and visual information), and that the compromise of Mayer between these two types of channels is not to be considered at the WM level. In addition, a consideration of the inter-individual differences suggests that the capacity of the WM storage components defines the extent to which subjects are able to benefit from a visuospatial presentation. This capacity also defines the extent to which these components are involved in such complex cognitive activities as comprehending verbal and visuospatial information.

To date, some studies have examined learning preferences, for example, Leutner & Plass (1998) or Plass, Chun, Mayer, & Leutner (1998) who showed that visualizing subjects

DOI: 10.1002/acp

(those who prefer to learn with visual material) memorize and understand a hypertext document presented with graphics and videos better than when only linguistic information is provided. In contrast, verbalizing subjects memorize and understand the hypertext document better when only linguistic information is provided than when graphics and videos are available (see also Mayer & Massa, 2003). It has also been shown that the subjects with the best spatial abilities understand illustrated documents the best, especially when they contain explanatory texts (Hegarty, Kriz, & Cate, 2003) and obtain the best performances in mental animation tasks (Hegarty & Kozhevnikov, 1999; Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997). The relationships between VSWM, executive functions and spatial abilities are strong but complex (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). It nevertheless appears to be the case that individuals' simple storage capacities deserve be considered in future research, even in cases where complex material is processed, given that these capacities, even if they do not explain performance, may at least influence the strategies used and thus modify the obtained performance patterns.

These outcomes have theoretical as well as practical implications. On the theoretical side, the studies on memory have long been conducted with material constructed for experimental purposes, such as lists of words, and with sometimes simplistic tasks which hardly reflect what is going on in everyday situations. Although it might be advantageous at early stages of theory development to study tasks at the less complex end of the continuum, we cannot expect the models developed in the context of simplistic tasks to scale up to models of more complex tasks such as comprehension. The results presented here indeed extend the validity of Baddeley's WM model to a complex cognitive activity, multimedia learning and questions the interest of adding a new component such as the episodic buffer to account for the results obtained. The dual-task paradigm proved in addition useful to assess more directly for cognitive load, even though a shortcoming is to differentiate cognitive load in the form of storing and the various attentional processes. This paradigm was used to measure cognitive load in Brünken et al. (2002), or to induce cognitive load in the experiment reported here, but performance of both the primary and the secondary tasks should be measured in future research. Using such type of methodology should enable researchers to validate empirically theoretical predictions of models of multimedia learning, and explain many multimedia effects. On a more practical side, the consideration of individual differences should be developed, as it could have effects on the cognitive load and consequently on the efficiency of multimedia learning. This might be of relevance to the designers of multimedia documents so that then can adapt their systems to the precise processing constraints of the user.

REFERENCES

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* (Vol. 2). New York: Academic Press.

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

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

Baddeley, A. (1992). Is working memory working? The fifteenth Bartlett lecture. *The Quarterly Journal of Experimental Psychology*, *44A*, 1–31.

DOI: 10.1002/acp

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

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A., & Logie, R. (1999). Working memory: The multiple-component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Bétrancourt, M., & Bisseret, A. (1998). Integrating textual and pictorial information via pop-windows: An experimental study. *Behavior and Information Technology*, *17*, 263–273.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, *247*, 44–48.

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.

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., Plass, J. L., & Leutner, D. (2002). Assessment of Cognitive Load in Multimedia Learning Using Dual-Task Methodology. *Experimental Psychology*, *49*, 109–119.

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

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

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, *94*, 428–434.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology*, *17*, 77–95.

Dubois, V. (2004). Intégration cognitive d'informations multimedia et mémoire de travail. (Cognitive integration of multimedia information and working memory) Unpublished doctoral thesis, University of Paris V.

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

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

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratchpad in working memory. *The Quarterly Journal of Experimental Psychology*, *38A*, 675–688.

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117–133). New York: Cambridge University Press.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, *129*, 61–83.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *4*, 313–331.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, *31*, 129–151.

Gray-Wilson, S., Rinck, M., McNamara, T. P., Bower, G. H., & Morrow, D. G. (1993). Mental models and narrative comprehension: Some qualifications. *Journal of Memory and Language*, *32*, 141–154.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom, and why? In H. van Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Ltd.

DOI: 10.1002/acp

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.
Gyselinck, V., De Beni, R., Pazzaglia, F., Meneghetti, C., & Mondoloni, A. (2007). Working memory components and imagery instruction in the elaboration of a spatial mental model. *Psychological Research*, *71*, 373–382.
Gyselinck, V., Ehrlich, M.-F., Cornoldi, C., de Beni, R., & Dubois, V. (2000). Visuospatial working memory in learning from multimedia system. *Journal of Computer Assisted Learning*, *16*, 166–176.
Haenggi, D., Kintsch, W., & Gernsbacher, M. A. (1995). Spatial situation models and text comprehension. *Discourse Processes*, *19*, 173–199.
Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from texts and diagrams. *Journal of Memory and Language*, *32*, 717–742.
Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*, 684–689.
Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition & Instruction*, *21*, 325–360.
Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, *22*, 411–430.
Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and Individual Differences*, *9*, 19–42.
Jeung, H. -J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, *3*, 329–343.
Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*, 71–115.
Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–372.
Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*, 126–136.
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.
Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*, 163–182.
Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159.
Kruley, P., Sciama, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory and Cognition*, *22*, 261–272.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.
Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*, 401–418.
Leutner, D., & Plass, J. L. (1998). Measuring learning styles with questionnaires versus direct observation of preferential choice behavior in authentic learning situations: The Visualizer/Verbalizer Behavior Observation Scale (VV-BOS). *Computers in Human Behavior*, *14*, 543–557.
Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, U.K.: Lawrence Erlbaum associates, Ltd.
Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambrige handbook of multimedia learning* (pp. 147–158). New York: Cambridge University Press.
Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49–63.
Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University press.
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., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology*, *95*, 833–846.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312–320.
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.
Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.
Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.
Miyake, A., & Shah, P. (1999). Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1–27). New York: Cambridge University Press.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.
Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*, 598–610.
Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.
Oakhill, J. V., Yuill, N., & Parkin, A. J. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading*, *9*, 80–91.
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1–4.
Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, *90*, 25–36.
Schnotz, W. (2001). Signs systems, technologies and the acquisition of knowledge. In J. F. Rouet, J. J. Levonen, & A. Biardeau (Eds.), *Multimedia learning: Cognitive and instructional Issues* (pp. 9–30). Oxford, UK: Elsevier science.
Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.
Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156.
Schnotz, W., Bôckheler, J., & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, *14*, 245–265.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312.
Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER 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., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition & Instruction*, *8*, 351–352.
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., van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.
Tabbers, H. K. (2002). The modality of text in multimedia instructions. Open University of the Netherlands, Heerlen.
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 AECT 2000, Long Beach, California.
Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*, 71–81.

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). Mahwah, NJ: Lawrence Erlbaum Associates.

Tardieu, H., Ehrlich, M.-F., & Gyselinck, V. (1992). Levels of representation and domain-specific knowledge in comprehension of scientific texts. *Language and Cognitive Processes*, *7*, 335–352.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*, 261–292.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, *3*, 257–287.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

DOI: 10.1002/acp