The Role of the Visuo-Spatial Sketchpad in Multimedia Learning: Do Spatial Text Contents Interfere with Picture Processing?

Anne Schüler (a.schueler@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Strasse 40,
72072 Tuebingen, Germany

Katharina Scheiter (k.scheiter@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Strasse 40,
72072 Tuebingen, Germany

Peter Gerjets (p.gerjets@iwm-kmrc.de)
Knowledge Media Research Center, Konrad-Adenauer-Strasse 40,
72072 Tuebingen, Germany

Abstract

The reported study examined whether the degree of spatial information conveyed through a text influences the effectiveness of multimedia presentations. It was assumed that the processing of spatial text contents might interfere in the visuo-spatial sketchpad with the execution of eye movements, associated with looking at pictures and reading. Accordingly, performance impairments were expected when presenting spatial (rather than visual) text contents along with pictures and, furthermore, when presenting spatial text contents in written instead of spoken form. Fifty-nine students were randomly assigned to four groups, resulting from a 2 × 2 design, with text contents (visual vs. spatial) and text modality (spoken vs. written) as independent variables. Consistent with our assumptions, learners with spatial text contents showed worse recall than those with visual text contents. However, there were no differences between written and spoken spatial text contents. Implications for learning with multimedia are discussed.

Keywords: multimedia; visuo-spatial sketchpad; modality effect; spatial text contents; visual text contents.

Introduction

In the last two decades, a substantial amount of research has been conducted on learning from multimedia, that is, learning from text and pictures.

A theoretical framework for learning with multimedia is provided by the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2005), which assumes that the cognitive system is composed of two channels, an auditory-verbal channel and a visual-pictorial channel. The differentiation of these two channels is partly derived from the working memory model of Baddeley (1992). When comparing CTML to this working memory model, the auditory-verbal channel corresponds to the phonological loop (PL), whereas the visual-pictorial channel corresponds to the visuo-spatial sketchpad (VSSP). According to CTML, text is mainly processed in the auditory-verbal channel (i.e., the PL), whereas pictures are processed in the visual-pictorial channel (i.e., the VSSP). We have argued recently that some aspects of Mayer's comparison between processing channels and working memory structures are problematic in the light of Baddeley's model, in particular some processing distinctions CTML postulates based on the modality of text. (cf. Rummer et al., 2008). However, we will not repeat that argument here but point to another theoretical problem of Mayer's equation between processing channels and working memory structures. Our criticism is based on the fact that since Baddeley's first comprehensive descriptions of his working memory model, there have been numerous new findings that have not yet been incorporated into CTML. In particular, the structure of the VSSP has been further specified. According to our view, these specifications may play an important role for the analyses of multimedia learning. Thus, the aim of this paper is to have a closer look at the information processing in the VSSP and its implications for learning with multimedia.

Information processing in the VSSP

One of the first researchers, who examined the functioning of the VSSP, was Logie (1995). He distinguished two components of the VSSP: a visual component and a spatial component. Whereas the visual component is assumed to deal with information like an object's color or form, the spatial component is assumed to handle information like spatial sequences or spatial configurations. This separation of a visual and a spatial component of the VSSP has been empirically confirmed (e.g., Darling, della Sala, & Logie, 2007; della Sala et al., 1999).

Whereas the research done by Logie and colleagues focused on pictorial stimuli, other researchers have addressed the question whether the VSSP may also be involved in the processing of text. This research suggests that under very specific conditions verbal information will not only be processed in the PL, but also in the VSSP, namely, if it contains information about visual or spatial aspects. Thus, for example, De Beni et al. (2005) showed that text with spatial contents interfered with a spatial secondary task, whereas text with more abstract contents (i.e., text without spatial information) did not interfere with a spatial secondary task. This specific interference between spatial text contents and spatial secondary task indicates that both are processed in the same component of working memory, namely, the spatial component. With regard to the processing of text with visual contents, less empirical evidence is available, but one study by Deyzac, Logie, and Denis (2006) confirms the assumption that visual text contents are processed in the visual component of the VSSP. Thus, if text contains information about spatial or visual configurations it is not only processed in the PL but also in the spatial or visual component of the VSSP, whereas if it contains more abstract information it is not processed in the VSSP but in the PL alone.

Another line of research on the spatial component of the VSSP has also shown that this structure is not only responsible for the processing of spatial information but also for the control of eye movements. Accordingly, several studies demonstrated interferences between the execution of eye movements and the processing of spatial information (e.g., Postle et al., 2006).

To sum up, it can be assumed that the VSSP can be separated into a visual and a spatial component. Besides pictorial information also verbal information can be processed here, namely, if it contains information about visual or spatial aspects. Additionally, the spatial component does not only process spatial information, but also controls a person's eye movements. Figure 1 shows which components of the VSSP are needed to represent combinations of pictures and different types of text contents. Based on this analysis of the VSSP, we can consider its implications for learning with text and pictures.

Implications for multimedia learning

Learning with multimedia means to present text *and* pictures to learners. Pictures are assumed to be processed in the visual and the spatial component of the VSSP, because pictures normally contain visual as well as spatial information (see Figure 1, parts A - C). With regard to text processing, the VSSP can be involved as a function of text contents: Figure 1 (part A) shows that the VSSP is not involved in text processing if abstract text contents are presented. However, if visual text contents are presented, the visual component will be involved (see Figure 1, part B), whereas when spatial text contents are presented, the spatial component is involved (see Figure 1, part C). Furthermore, as the spatial component of the VSSP controls the execution of eye movements looking at pictures and reading written text will result in an additional load of the spatial component. Thus, one might expect interferences between the execution of eye movements and the processing of spatial text contents, because both are processed in the spatial component. As can be seen in Figure 1 (part C) the spatial component might become particularly overloaded when pictures *and* spatial text contents have to be processed and eye movements have to be conducted, for example, in order to read the text or to look at the picture. Two implications result from this analysis: The first one applies to the presentation of pictures together with spatial text contents (either presented spoken or written). The second one is related to the presentation of written instead of spoken text.

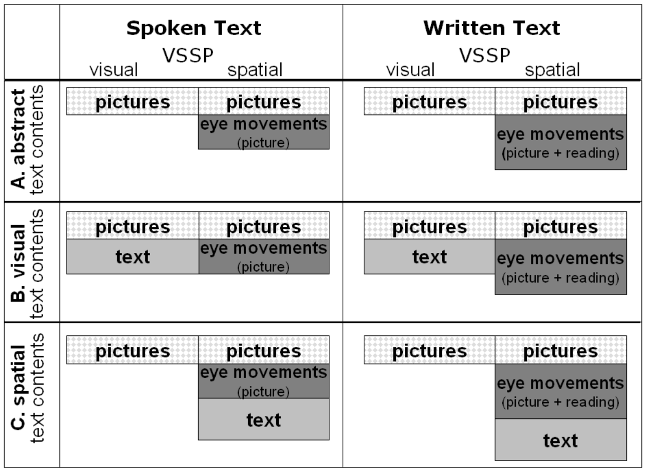


Figure 1. The processing of multimedia material in the VSSP as a function of text contents and text modality. Furthermore, text is always processed in PL.

First Implication: Interference between processing of spatial text contents and looking at pictures. When presenting pictures together with spatial text contents, one would expect interferences in the spatial component of the VSSP, because the processing of spatial picture contents and spatial text contents as well as the control of eye movements take place in the spatial component (see Figure 1, part C). When presenting pictures together with non-spatial text contents, one would expect less interference because the load is distributed more equally (see Figure 1, part A and part B). Accordingly, pictures presented together with spatial text contents should result in worse learning outcome than pictures presented together with non-spatial text contents, that is, abstract, or visual text contents. A study conducted by Scheiter and Schmidt-Weigand (2008) confirms this assumption, by showing that pictures are only helpful for learning when they accompany text with a low degree of spatial information but not text with a high degree of spatial information. Besides this study of Scheiter and Schmidt-Weigand, there is only little empirical evidence for the interplay between text contents and picture processing. One reason for this lack of research might be that it is difficult to compare learning outcomes resulting from learning with different text contents, since the contents might differ with regard to their difficulty, so that differences in text recall might be difficult to interpret. However, if the *same* pictures are presented together with spatial and non-spatial text contents, differences in picture recall would indicate more unequivocally interferences in the VSSP. Accordingly, we will use this method in the current study.

Second implication: Interference between processing of spatial text contents and reading. A second implication of the preceding analysis refers to the modality of the text:

Because eye movements are not only needed for picture inspection, but also for reading, one might expect worse performance with written text than with spoken text when processing spatial text contents. Figure 1 (part C) shows that the spatial component with spoken text and spatial text contents is less loaded than with written presentation of spatial text contents, because more eye movements are required to read the text and to switch between text and picture. This load difference might result in worse learning outcome for written spatial text than for spoken spatial text. With visual text contents or more abstract text contents the difference between written text and spoken text is not expected to be equally harmful, because none of the two text contents is processed in the spatial component and therefore no interference with the control of eye movements is expected. Thus, the spatial component might not be overloaded when presenting written text, which might enable the same learning outcome as for spoken text (see Figure 1, parts A and B). First evidence for that prediction that a "modality effect" (i.e., worse learning outcome for written text than for spoken text) occurs only with spatial text contents was collected in purely text based studies. Studies by Brooks (1967) and Kürschner et al. (2006) used texts that described either spatial relations or contained more abstract information. A modality effect was found only with regard to spatial text contents, but not for more abstract information. Another study by Glass et al. (1985) explicitly examined the influence of text modality on the processing of visual and spatial text contents. Whereas with regard to sentences about spatial relations (e.g., "To turn on a light you move the switch up/down", p. 456) a modality effect occurred, this was not the case with regard to sentences about visual characteristics like color (e.g., "The spots on a giraffe are brown/yellow", p. 456).

In the context of multimedia research a modality effect has been found several times (see Ginns, 2005). In this literature, learners presented with spoken text and picture outperformed learners with written text and picture. The CTML provides a theoretical explanation for this modality effect that has, however, been challenged (for more details see Rummer et al., 2008). From our perspective the structure of the VSSP might be an explanation for this modality effect, at least when text contents about spatial configurations are presented which is the case in most studies.

In the current study we will focus on testing the two predictions that have been derived from our analysis of the structure of the VSSP with regard to text contents and text modality.

Experiment

The aim of the current study was to investigate whether processing spatial text contents interferes with looking at pictures and – additionally – whether processing spatial text contents interferes with reading. To test these hypotheses, we created a multimedia learning environment, where learners were either given text information about visual or spatial features of different artificial fish species. The information was presented either in written or in spoken format. Independently of the available text contents or text modality, all learners received the same pictures showing the artificial fish species described. Our first expectation was that learners with spatial text contents will be less able to recall the text and to recall the pictures than learners with visual text contents. Whereas differences in text recall might potentially be attributed to different text difficulties, differences in picture recall can be attributed unequivocally to interferences in the VSSP because all groups of learners had to recall the same pictures.

Secondly, we expected an interaction between text contents and text modality: Reading written text should interfere with processing spatial text contents, thus we predicted that learners with written spatial text will be less able to recall text and picture contents than learners with spoken spatial text contents. This prediction, however, depends on the presupposition that the VSSP of learners with spoken spatial text contents (Figure 1, left side, part C) is not already too overloaded. Because reading should interfere less severe with visual text contents, we expected no difference in recall of pictures and text between written visual text contents and spoken visual text contents.

Both text materials (spatial and visual contents) additionally contained more abstract contents (e.g., biological facts or behavioral descriptions of the different fish species). This information was the same for both text materials. Our third expectation was that there are no differences between the four groups with regard to recalling these abstract contents, because abstract information should not be processed in the VSSP and, therefore, not interfere with picture processing and the control of eye movements.

Method

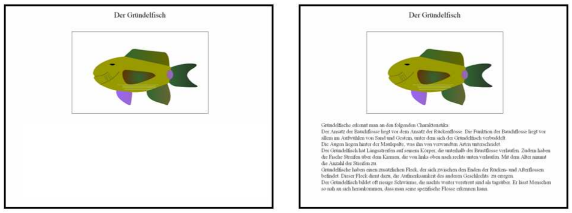
Participants and Design. Fifty-nine students of the University of Tuebingen (43 female, 16 male, average age: *M* = 23.76 years, *SD* = 3.85 years) participated in the study for either payment or course credit. They were randomly assigned to one of four conditions, which resulted from a 2 × 2 design, with text contents (visual vs. spatial contents) and text modality (spoken vs. written text) as independent variables.

Materials. The materials were presented in a computerized learning environment. It comprised an introduction, the learning phase, and a test phase.

In the introduction, learners were asked about their demographic data. Furthermore, they had to learn the names of different body parts of fish (e.g., anal fin, dorsal fin etc.), because these names were used in the subsequent learning materials.

The system paced learning phase consisted of six static pictures and six corresponding texts about artificial fish. The main reason to use artificial fish species instead of real fish species was to avoid influences of prior knowledge on the learning results. Every fish species was presented on a single slide. As mentioned before, the pictures were identical in all groups, whereas the texts differed with regard to contents as a function of the experimental condition. Note that the text lengths of the visual and spatial texts were equivalent and that the pace of presentation was determined by the duration of the spoken text conditions.

The independent variables were varied between groups in the learning phase as follows: Learners with visual text contents received information about visual features of the depicted fish species, that is, the color or form of specific body parts (e.g., “The anal fin has the same light brown color as the dorsal fins”). Learners with spatial text contents received information about spatial features of the fish species, that is, the location of a body part or its spatial relation to other parts (e.g., “The anal fin lies between the two dorsal fins”). Furthermore, both texts contained identical abstract information on biological concepts and facts (e.g., “The fins are used for defense”). In the conditions with spoken text, learners listened to the text while the picture was presented on the screen. In the conditions with written text, the text was presented below the picture (see Figure 2).



Gründelfische erkennt man an den folgenden Charakteristika:
Der Ansatz der Bauchflosse liegt vor dem Ansatz der Rückenflosse. Die Funktion der Bauchflosse liegt vor allem im Aufwühlen von Sand und Gestein, unter dem sich der Gründelfisch verbuddelt.
Die Augen liegen hinter der Maulspalte, was ihn von verwandten Arten unterscheidet.
Der Gründelfisch hat Längsstreifen auf seinem Körper, die unterhalb der Brustflosse verlaufen. Zudem haben die Fische Streifen über dem Kiemen, die von links oben nach rechts unten verlaufen. Mit dem Alter nimmt die Anzahl der Streifen zu.
Gründelfische haben einen zusätzlichen Fleck, der sich zwischen den Enden der Rücken- und Afterflossen befindet. Dieser Fleck dient dazu, die Aufmerksamkeit des anderen Geschlechts zu erregen.
Der Gründelfisch bildet oft riesige Schwärme, die nachts weiter verstreut sind als tagsüber. Er lässt Menschen so nah an sich herankommen, dass man seine spezifische Flosse erkennen kann.

Figure 2: Presentation of text and picture in the spoken (left) and written (right) text groups.

Measures. The test phase consisted of seven types of verification items about the presented fish species. These seven types of verification items were created by considering the following two dimensions: The source of information learners had to remember to verify the item (i.e., text vs. picture vs. picture and/or text) and the contents the item was asking about (i.e., visual vs. spatial vs. abstract). It is important to note that every learner answered every item, thus, items were not varied between groups. Hence, with regard to picture recall, learners with visual text contents also had to recall spatial aspects of the picture, and learners with spatial text contents also had to recall visual aspects of the picture. With regard to the items asking about pure text information we also used the same set of items across all conditions although text contents differed between groups in the learning phase. For example, there were items that ask if specific colors of the fish species were mentioned in the text (e.g., “The anal fin has the same light brown color as the dorsal fins”). If this feature of the fish species was mentioned in the text with visual contents, learners with visual text contents had to accept the item as “true” (hit). However, learners with spatial text contents had to reject the item (correct rejection), because the color was not mentioned in the text with spatial contents. Thus, with regard to text recall it was necessary to code the answers differently in groups with visual and spatial text contents. Based on these differences in coding an unified performance index was computed assuming that hits as well as correct rejection are indicators of text recall.

Furthermore, two items which measured difficulty during learning were presented after the learning phase.

Procedure. Participants were tested individually. They first studied the system paced learning materials. Subsequently, they responded to the item measuring learning difficulty and the verification items. A single experimental session lasted about 60 minutes.

Results

The means and standard deviations in percent for the seven different dependent variables as a function of text contents and text modality are shown in Table 1.

Table 1: Means and standard deviations in percent as a function of text content and text modality.

Test item type		Instructional materials			
		visual text contents		spatial text contents	
source	contents	spoken	written	spoken	written
text	visual	66.67	69.32	61.69	57.79
		(14.85)	(17.53)	(21.74)	(14.53)
	spatial	72.12	71.59	54.54	51.30
		(11.64)	(19.60)	(17.10)	(7.65)
picture	visual	71.67	66.41	55.16	59.13
		(19.17)	(10.91)	(12.22)	(10.59)
	spatial	57.78	54.86	47.62	47.62
		(10.03)	(14.47)	(11.88)	(17.12)
text - picture	visual	75.24	75.00	72.86	84.29
		(12.03)	(14.75)	(26.73)	(21.02)
	spatial	73.33	68.75	53.81	49.05
		(24.69)	(25.27)	(18.20)	(10.97)
text	abstract	80.00	80.01	76.19	76.19
		(21.08)	(21.27)	(19.30)	(22.37)

Because the same pattern of results was expected for recall of picture-based information, recall of text-based information and recall of text- and/or picture-based information, the corresponding variables were analyzed by means of a multivariate analysis of variance with text modality and text contents as between subject factors. As predicted the results showed a significant difference between learners with visual and spatial text contents, *Pillai’s Trace* = .58, $F(6, 50) = 11.51$, $p < .001$. With regard to text modality the main effect was not statistically significant, *Pillai’s Trace* = .06, $F < 1$, that is, learners with spoken text showed the same overall performance as learners with written text. The predicted interaction was not statistically significant, *Pillai’s Trace* = .07, $F < 1$, which indicates that no interference between text contents and text modality appeared. To further investigate for which of the dependent variables the main effect of text contents occurred, univariate two-way ANOVAs were conducted. The results are reported according to the information source, which had to be remembered to answer the items, that is, text, picture, or text and/or picture. Because of space limitations, the statistical details are only reported for significant results.

With regard to items where *text-based* information had to be remembered, a main effect of text contents occurred for both of the item types (i.e., items asking for visual vs. spatial text contents). For items asking for visual information learners with visual text contents (M = 68.04, SD = 16.07) performed marginal significant better than learners with spatial text contents (M = 59.74, SD = 18.25; $F(1, 55) = 3.32$, $p = .07$, $\eta^2 = .06$.) In other words: Learners with visual text contents had more hits than learners with spatial text contents had correct rejections. With regard to items asking for spatial information learners with visual text contents (M = 71.84, SD = 15.98) performed also better than learners with spatial text contents (M = 52.92, SD = 13.10, $F(1, 55) = 23.72$, $p < .001$, $\eta^2 = .30$. In other words: Learners with visual text contents had more correct rejections than learners with spatial text contents had hits. Because hits as well as correct rejections are indicators of text recall, this indicates that learners with visual text contents could recall better "their" text contents than learners with spatial text contents.

With regard to items where *picture-based* information had to be remembered, again two main effects for text contents occurred. Learners with visual text contents (M = 68.95, SD = 15.44) could remember better visual aspects of the picture (like color or form) than learners with spatial text contents (M = 57.14, SD = 11.40; $F(1, 55) = 11.07$, $p < .01$, $\eta^2 = .17$). This indicates that learners who read texts about visual characteristics of the fish also could better remember visual aspects only shown in the picture. Interestingly, learners with visual text contents (M = 56.27, SD = 12.40) could also better remember spatial aspects of the picture than learners with spatial text contents (M = 47.62, SD = 14.46; $F(1, 55) = 6.00$, $p = .02$, $\eta^2 = .10$). These results indicate that learners with visual text contents processed the picture more thoroughly than learners with spatial text contents.

With regard to items that could be answered with *text- and/or picture-based* information (i.e., both information sources could be used to answer the items) a significant main effect was only found for items asking for spatial information: Learners with visual text contents (M = 70.97, SD = 24.68) remembered this information better than learners with spatial text contents (M = 51.43, SD = 14.95; $F(1, 55) = 12.97$, $p > .01$, $\eta^2 = .19$).

Besides the analysis of the recall performance for contents presented in the text and pictures, an additional analysis was conducted with regard to recall of abstract information. Both groups received the same abstract information and we did not expect any differences between groups. This assumption was in line with the analysis, showing no main effects for text contents and text modality and no interaction between text contents and text modality for abstract information (all $F_s < 1$). Thus, this might be seen as an indicator that abstract text contents do not interfere with picture processing in the VSSP (see also Figure 1, A).

With regard to the perceived difficulty of the learning phase no differences were observed between groups, that is, learners with visual and spatial text content as well as learners with spoken and written text evaluated the learning phase as equally difficult.

Discussion

The purpose of the reported study was to examine the hypotheses that spatial text contents might interfere with picture processing and reading due to spatial picture contents and eye movements. These expectations were derived from assumptions about the structure of the VSSP.

We expected worse learning outcomes for combining pictures with spatial text content as compared to visual text contents, because of specific interferences between spatial text processing and the control of eye movements, both of which take place in the spatial component of the VSSP. Furthermore, these interferences between spatial text processing and control of eye movements should be affected by the modality of the presented text: Because reading requires eye movements, it should stronger interfere with the processing of spatial text contents than listening. Hence, we expected a modality effect, that is, worse performance with written as compared to spoken text with regard to spatial text contents, but not with regard to visual text contents. To test these assumptions, text contents (visual vs. spatial) and text modality (spoken vs. written) were varied. Furthermore abstract information was presented to learners. Because abstract contents will not be processed in the VSSP, no differences between the four groups for this type of information were expected.

The first assumption was confirmed in that learners which received pictures together with spatial text contents showed overall worse performance in recalling text-based, picture-based and text- and/or picture-based information than learners which received pictures together with visual text contents. Several univariate ANOVAs confirmed the superiority of learners with visual text contents for nearly all of the dependent variables. Importantly, learners with spatial text contents not only remembered the text, but also the picture worse, which was the same in all conditions. Furthermore, with regard to remembering abstract information, no difference between learners with spatial and visual text contents occurred, which indicates that the interplay between text contents and picture processing is limited to spatial text contents.

With regard to recall of text contents, one might argue that visual text contents, that is, information about color and form, might be easier to process and to remember than spatial text contents, that is, information about spatial relationships or the position of a certain characteristic. Thus, the fact that learners with visual text contents performed better, when they had to remember text-based information might potentially be simply explained by differences in text difficulty and not by interferences in the spatial component of the VSSP. On the other hand, there was no difference with regard to the perceived difficulty of the learning phase between learners which might indicate that the text difficulties were comparable.

The results obtained for picture recall also support the assumption of interferences between processing of spatial text contents, and the processing of pictures. Because pictures were the same in all groups, one would expect the same performance of all learner groups, if text contents had *no* influence on the processing of pictures in the VSSP. As our results show, learners with visual text contents could not only remember visual aspects of the picture but also spatial aspects of the picture better. This supports the assumption that all aspects of the picture are processed more deeply when text contents are not spatial. However, one might argue that also with regard to picture recall text difficulty might play a role: Because spatial text might be more difficult to process, learners might concentrate more on text and neglect the picture. This in turn might result in worse performance for remembering pictures. In spite of the fact that there were no differences with regard to perceived difficulty which contradicts this hypothesis, we will test that assumption by using eye tracking technology to control for time on text and time on picture.

Our second expectation concerning a modality effect for spatial text contents because of interferences between reading and spatial text processing (see Figure 1), was not confirmed: When being presented with spatial text contents, learners with written text showed the same performance level as learners with spoken text. Thus, the often found superiority of spoken text over written text in multimedia learning (see Ginns, 2005) was not replicated in this study. Thus, the results with regard to text modality do neither fit to the assumptions of CTML nor to the more specific assumptions tested in the present study.

The question remains, why in research without picture materials the specific interaction of text contents and text modality was found (e.g., Brooks, 1967; Glass et al., 1985). One explanation might be that learners without picture have to imagine text information in order to achieve an understanding. Maybe the process of imagining spatial configurations interferes with the eye movements associated with reading and therefore leads to a modality effect when being presented with spatial text contents.

To get deeper inside into the interplay of text contents and learning with text and pictures further research is needed that addresses more fine-grained processing aspects (e.g., by means of eye tracking). Currently, a study is conducted where we use the dual task paradigm to examine more accurately whether worse performance of spatial text contents is due to interferences in the spatial component of the VSSP. This approach is in line with our conviction that more basic cognitive research is needed to develop more precise theoretical frameworks for explaining how multimedia learning works.

References

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19*, 289-299.

Darling, S., della Sala, S., & Logie, R. H. (2007). Behavioural evidence for separating components within visuo-spatial working memory. *Cognitive Processing, 8*, 175-181.

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.

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.

Deyzac, E., Logie, R. H., & Denis, M. (2006). Visuospatial working memory and the processing of spatial descriptions. *British Journal of Psychology, 97*, 217-243.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning & Instruction, 15*, 313-331.

Glass, A., Millen, D., Beck, L., & Eddy, J. (1985). Representation of images in sentence verification. *Journal of Memory and Language, 24*, 442-465.

Kürschner, C., Seufert, T., Hauck, G., Schnotz, W., & Eid, M. (2006). Konstruktion visuell-räumlicher Repräsentationen beim Hör- und Leseverstehen. *Zeitschrift für Psychologie, 214*, 117-132.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, England: Erlbaum.

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* Cambridge: Cambridge University Press.

Postle, B. R., Idzikowski, C., della Sala, S., Logie, R. H., & Baddeley, A. (2006). The selective disruption of spatial working memory by eye movements. *The Quarterly Journal of Experimental Psychology, 59*, 100-120.

Rummer, R., Schweppe, J., Scheiter, K., & Gerjets, P. (2008). Lernen mit Multimedia: Die kognitiven Grundlagen des Modalitätseffekts. *Psychologische Rundschau, 59*, 98-107.

Scheiter, K., & Schmidt-Weigand, F. (2008). The influence of spatial text information on learning with visualizations. *Proceedings EARLI Special Interest Group Text and Graphics: Exploiting the opportunities - Learning with textual, graphical, and multimodal representations* (pp. 123-126). Tilburg, NL: Tilburg University