Computers in Human Behavior 27 (2011) 22–28

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

The role of spatial descriptions in learning from multimedia

Florian Schmidt-Weigand[a,*,1], Katharina Scheiter[b,1]

[a] Institute of Psychology (FB 4), University of Kassel, Hollaendische Strasse 36, 34109 Kassel, Germany
[b] University of Tuebingen, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

ARTICLE INFO

Keywords:
Multimedia
Animation
Spatial information
Working memory
Redundancy
Conjoint retention

ABSTRACT

In the reported experiment we investigated how spatial information conveyed in an expository text influenced multimedia learning. It was based on a 2×2-design with the degree of spatial information given in the text (high vs. low spatial text) and the presentation format (written text-only vs. written text + animation) as between-subjects factors. As dependent variables learning outcomes as well as self-reported cognitive load were assessed. The results revealed that there was a multimedia effect with regard to learning outcomes only for low spatial text, but not for high spatial text. Moreover, the cognitive load measures showed an overall multimedia effect irrespective of the degree of spatial information conveyed by the text (i.e., higher cognitive load ratings in the text-only conditions). These results can be explained as a special instance of the redundancy effect as well as a consequence of processing interference within visuo-spatial working memory.

1. Introduction

There is a long-standing tradition of educational research showing that combinations of text and visualizations improve learning outcomes compared to learning from text-only (Anglin, Vaez, & Cunningham, 2004; Fletcher & Tobias, 2005). There are two lines of argumentation for explaining this so-called multimedia effect. The first line of argumentation focuses on the cognitive processes that are facilitated when studying visualizations rather than text (i.e., process-oriented; e.g., Larkin & Simon, 1987). The second view is more outcome-oriented in that it is argued that learning from visualizations will yield a richer internal representation of the to-be-learned content compared to learning from text (e.g., Baggett, 1984; Paivio, 1991).

1.1. A process-oriented view on learning from visualizations

Larkin and Simon (1987) have suggested that visualizations are often more computationally efficient than verbal descriptions for accomplishing tasks that require the processing of visuo-spatial properties by reducing the need to search for multiple information elements related to a single idea as this information is grouped in visualizations. Moreover, visualizations may support inference and reasoning processes grounded in perception allowing perceptual processes to replace more demanding logical inferences (Goldstone & Son, 2005; Zhang, 1997). Thus, according to the process-oriented view, visualizations may aid learning in cases where learning is dependent on extracting and reasoning with visuo-spatial information by enabling and facilitating cognitive processes that are less demanding than the processes that would otherwise be required when learning from text.

1.2. An outcome-oriented view on learning from visualizations

According to this view, visualizations support learning, because compared to text their processing yields an additional and qualitatively different representation in memory, thereby facilitating recall of the conveyed information (e.g., Baggett, 1984; Paivio, 1991).

For instance, the *bushiness hypothesis* by Baggett (1984) suggests that knowledge acquired from visual rather than verbal representations will be better accessible in memory because the respective nodes in memory share more associations with other nodes in the semantic network, that is, they are "bushier". Because concepts with more associations can be retrieved more easily from memory, visualizations should facilitate recall of the content.

Similarly, according to the Dual Coding Theory (Paivio, 1991), verbal and non-verbal external information sources are assumed to yield different internal representations. Verbal information results in a symbolic linguistic representation (logogens), whereas non-verbal information is encoded analogically (imagens). Processing verbal and pictorial information compared to text alone is assumed to result in a dual coding of this information, which in turn is better accessible in memory and hence easier to recall. The same assumption has been made in the context of the conjoint-retention hypothesis that specifically deals with spatial information representations (Verdi & Kulhavy, 2002). This hypothesis

* Corresponding author. Tel.: +49 561 8043576; fax: +49 561 8043586.
E-mail address: f.schmidt-weigand@uni-kassel.de (F. Schmidt-Weigand).
[1] These authors contributed equally to this work.

0747-5632/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2010.05.007

states that if spatial information is represented through both, a verbal and a visual representation, it will be better retained than information solely represented in the text.

1.3. Comparing the process- and outcome-oriented view

There are at least three differences between the two views: First, the process-oriented view emphasizes that visualizations aid learning once visuo-spatial reasoning is required, whereas the outcome-oriented view does not limit the benefits of visualizations to a specific learning task. Second, the process-oriented view appears to be better suited to explain the benefits of learning with visualizations with regard to deeper comprehension of the contents as a function of the type of cognitive processes taking place, whereas the outcome-oriented view limits the benefits of visualization to a better recall of information. Despite the fact that deeper understanding rather than recall of information is at the heart of theories of multimedia learning like the Cognitive Theory of Multimedia Learning (CTML, Mayer, 2005), the CTML nevertheless refers to dual coding as the main argument for why visualizations aid learning. Considerably less attention has been paid to the process-oriented view in the multimedia literature (for an exception, see Ainsworth, 2006), despite the fact that it may yield more specific predictions as to when visualizations should be particularly effective for learning. A third difference between the two views that is in the focus of the current paper refers to the question of how information should be distributed across text and visualizations.

According to the outcome-oriented view, redundant information presentations should aid learning compared to a complementary presentation. If the same information is contained both in the text and the visualization, it yields more connections between the visualization and the text (Baggett, 1984) and allows for a dual coding of that information (Paivio, 1991; Verdi & Kulhavy, 2002). Interestingly, this assumption contradicts the so-called redundancy effect in multimedia learning (Sweller, 2005), which suggests that learning will be hindered when identical information is conveyed through multiple representations. It is explained by assuming that redundant information presentations require learners to coordinate the processing of text and pictures in working memory. However, this coordination is unnecessary to arrive at a deeper understanding. Hence, redundancy causes unnecessary, that is, extraneous cognitive load that hinders learning (Sweller, van Merriënboer, & Paas, 1998). There is, however, no evidence for the validity of these processing assumptions.

On the other hand, the process-oriented perspective entails two implications with regard to how information should be distributed across text and pictures and whether redundancy should be avoided or not. Note that this account makes predictions only for situations involving visuo-spatial reasoning, which is why we will limit our discussion to these situations. First, according to the aforementioned analysis visualizations rather than text are better suited to convey visuo-spatial information (Larkin & Simon, 1987). Hence, to optimize information distribution across representations, visuo-spatial information should be conveyed through the visualization and information that is better conveyed through text should be represented verbally (e.g., abstract concepts or logical relations). Accordingly, the CTML (Mayer, 2005) as well as Cognitive Load Theory (CLT, Sweller et al., 1998) both stress that learners will especially benefit from multimedia materials if the text and pictures convey complementary information; however, this complementariness is not specified any further. Rather, it is equated with the request to avoid redundancy.

The second implication suggests that conveying visuo-spatial information through both, text and visualization, may even hinder learning, because it may yield a very specific interference in working memory. To explain this assumption we will outline how visualizations and visuo-spatial verbal descriptions are processed in working memory in the next section.

1.4. The processing of visuo-spatial information in working memory

According to Baddeley (2007), working memory consists of several slave systems, namely the phonological loop (PL), the visuo-spatial sketchpad (VSSP), and the episodic buffer (EB), as well as the so-called central executive (CE). The PL is responsible for the processing of either written or spoken verbal information, whereas the VSSP enables the processing of visual and spatial information. In more recent versions of the working memory model, the EB was added as a system that allows for binding information from the PL and the VSSP and storing it in a multimodal code. The CE is responsible for monitoring and coordinating the operation of the slave systems, deciding on which information is attended to, updating and regulating the contents of working memory, and coding representations in working memory for their time and place of appearance. Each component of working memory is assumed to have its own limited cognitive resources, which enables them to act relatively independent from each other. However, if two concurrent tasks make use of the same working memory component, there will be interference between them.

Using Baddeley's model of working memory, multimedia learning can be described as a process where verbal information is processed in the PL and the visualizations are dealt with in the VSSP (Gyselinck, Jamet, & Dubois, 2008). This seemingly simple assignment of representations to these systems has, however, shown to be wrong for verbal descriptions of spatial information (i.e., spatial texts). For spatial texts, not only the PL, but also the VSSP is involved in their processing, presumably because such texts may evoke mental imagery. For instance, Pazzaglia and Cornoldi (1999) showed that a spatial concurrent task that required processing in the VSSP interfered with the recall of spatial sentences, but not of abstract sentences. Moreover, learners with a low compared to high performance in a well-established VSSP measure (i.e., the Corsi block task, in which subjects have to remember a sequence of targeted movements performed by an experimenter) showed worse recall of spatial texts. Similarly, De Beni, Pazzaglia, Gyselinck, and Meneghetti (2005) found that a spatial concurrent task interfered with the encoding and the retrieval of a spatial text, but not of a non-spatial text (for further evidence, see Deyzac, Logie, & Denis, 2006).

One can derive from these findings that the processing of spatial texts, but not that of non-spatial texts, should interfere with the processing of an accompanying visualization, which likewise demands VSSP resources.

1.5. Hypotheses

The question of what happens if spatial information is conveyed by means of both, text *and* pictures was addressed in the current study on learning from animations, in which we varied experimentally the degree to which the accompanying text contained spatial information. The two aforementioned views on learning from visualizations make contradicting predictions concerning the way in which the type of text (low vs. high spatial text) will moderate the multimedia effect.

According to the *outcome-oriented view* (Baggett, 1984; Paivio, 1991; Verdi & Kulhavy, 2002), one would expect a stronger multimedia effect for text containing a high degree of spatial information (high spatial text) than for text containing a low degree of spatial information (low spatial text). This should be the case, because for high spatial text, visualizations containing the same information can be better linked to the verbal information