Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations?

Anne Schüler[*], Jana Arndt, Katharina Scheiter

Knowledge Media Research Center, Tuebingen, Germany

ABSTRACT

We investigated whether individuals construct either two interconnected mental representations or a single mental representation that merges information from two representations when processing multimedia materials. Individuals memorized text-picture stimuli in four different versions that differed in the specificity of information contained in either text or pictures: general pictures/general sentences, general pictures/specific sentences, specific pictures/general sentences, and specific pictures/specific sentences. Afterwards, individuals decided whether they had previously seen the specific or the general version of the sentences and the pictures. Across two experiments, individuals more frequently falsely recognized the specific sentences after having seen general sentences/specific pictures. This indicates that individuals had integrated the specific picture information with the general sentence information into a single mental representation. No such effects were observed for picture recognition. The implications of these results for multimedia learning are discussed.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In the past decades, it has been demonstrated repeatedly that people learn more deeply from text and pictures than from text alone (for overviews see Levie & Lentz, 1982; Mayer, 2009). Multimedia learning theories explain this finding by assuming that text and pictures are integrated with each other (e.g., Mayer, 2009; Schnotz & Bannert, 2003). The resulting mental representation is assumed to better enable learners to recall and apply the learnt content to novel situations. However, despite its importance for multimedia learning (e.g., Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Seufert, 2003), the precise nature of the integration process has never been examined empirically under conditions of multimedia learning. This was the aim of the studies reported here. From a theoretical perspective, knowing more about the kind of mental representations constructed during processing multimedia materials is pivotal to understand how people process text and pictures. In addition, as Renkl (2013) points out, "educational research is typically not only theoretically but also practically grounded" (p. 318). Thus,

knowing more about the way people learn from text and pictures allows us to derive recommendations for the design of multimedia materials that are grounded in cognitive theory (e.g., Mayer, 2009; for an analogous example see also Rummer, Schweppe, Fürstenberg, Scheiter, & Zindler (2011) and related follow-up research by Leahy & Sweller (2011)).

In the reported studies we contrasted two views concerning the integration of text and pictures. According to the first view two separate mental representations are constructed (i.e., one from the text and one from the picture), which are interconnected (i.e., integrated) via referential connections. We will call this view the "two-mental-representations" view, which goes back to dual coding theory (Paivio, 1990). According to the second view a single mental representation is constructed in which the information from the two representations is merged (i.e., integrated). We will call this view the "single-mental-representation" view, which goes back to theories on mental model construction (Johnson-Laird, 1983; Kintsch, 1988).

In the following, we will first introduce the assumptions regarding the integration process made by the two most prominent theories of multimedia learning, namely, the cognitive theory of multimedia learning (Mayer, 2009) and the integrated model of text and picture comprehension (Schnotz & Bannert, 2003), before describing the aforementioned views in more detail.

* Corresponding author. Knowledge Media Research Center, Schleichstraße 6, 72076 Tuebingen, Germany. Tel.: +49 7071 979341.

E-mail address: a.schueler@iwm-kmrc.de (A. Schüler).

1.1. Theoretical assumptions of multimedia theories regarding text-picture integration

The cognitive theory of multimedia learning (Mayer, 2009) posits that when text and pictures are presented together, learners construct a verbal representation from the text as well as a pictorial representation from the picture, which are then integrated. Concerning the integration process, Mayer (1997) assumes that learners build one-to-one correspondences (i.e., referential connections) between the verbal and the pictorial mental representation. This assumption is theoretically grounded in dual coding theory (Paivio, 1990) and appears to be in line with the two-mental-representations view. On the other hand, Mayer (2009, p. 79) states that making connections between the verbal and pictorial mental model "involves a change from having two separate representations − a pictorial model and a verbal model − to having an integrated representation." Furthermore, in Mayer's diagrammatic depiction of the cognitive theory of multimedia learning, the resulting mental model is represented by means of a single element in working memory (see for example, Mayer, 2009, p. 77). This assumption seems to better correspond with the single-mental-representation view.

The integrated model of text and picture comprehension (Schnotz & Bannert, 2003) assumes that individuals construct a propositional representation from the text and a mental model from the picture. The propositional representation may trigger the construction of the mental model; furthermore, the mental model can be used to read off information in a propositional format. As the authors point out, constructing a propositional representation as well as a mental model seems to be, at first glance, similar to dual coding (Schnotz & Bannert, 2003). However, in contrast to dual coding theory, the authors do not assume that the mental model simply adds an additional pictorial code that yields a quantitative advantage compared to a single code. Instead, the mental model is a qualitatively different representation, which incorporates also information from text. Thus, the assumptions of Schnotz and Bannert (2003) are more in line with the single-mental-representation view than with the two-mental-representations view.

To conclude, both theories agree that the integration of text and pictures is crucial to learning from multimedia. However, whether the integration process results in two mental representations, which are interconnected with each other (Mayer, 1997), or a single representation, which merges information from text and picture (Mayer, 2009; Schnotz & Bannert, 2003), is still an open question. Therefore, the aim of the studies reported in the current paper was to examine the precise nature of this integration process under conditions of multimedia learning. In the following, we will describe the two different views concerning the integration of text and pictures in more detail.

1.2. Construction of mental representations when processing text and pictures

According to the *two-mental-representations view*, text and pictures support the construction of two mode-specific mental representations, which are interconnected via referential connections. This view is grounded in dual coding theory (Paivio, 1990). According to this theory, the cognitive system is composed of a verbal system, which deals with language, and a nonverbal system, which deals with the analysis of pictorial information and the generation of mental images. One system can trigger activity in the other system by means of referential connections between representations stored in them. For example, when a picture is presented, the corresponding verbal

representation can be activated via activation spreading. Dual coding theory assumes that dual coding of information yields better recognition and recall performance because if one code is forgotten the other code may still be accessible in memory (e.g., Paivio, 1965).

According to the two-mental-representations view based on dual coding theory, text-picture presentations are more likely to lead to two mental representations interconnected via referential connections than presenting either text or pictures, which can explain why people learn more deeply from multimedia materials.

On the other hand, according to the *single-mental-representation view*, text and pictures support the construction of a single mental representation that contains verbal as well as pictorial information. This view is grounded in theories proposing that during information processing, individuals construct comprehensive mental representations known as situation models (Kintsch, 1988) or mental models (Johnson-Laird, 1983; Zwaan & Radvansky, 1998). According to these theories a single mental representation is constructed, which merges information derived from different information sources and prior knowledge (for empirical evidence, see for example Allen, Hitch, & Baddeley, 2009; Elsley & Parmentier, 2009; Maybery et al., 2009). It is assumed that this mental representation comprises all the information from the different external representations (e.g., Baddeley, Allen, & Hitch, 2011; Brunyé & Taylor, 2008; Payne & Baguley, 2006; Zwaan & Radvansky, 1998). This implies that by integrating information across non-redundant information sources, the resulting mental representation becomes more specific. For example, when a picture of a car is presented and a sentence additionally states that the car has ski-racks on it, it is assumed that the mental representation will then contain information about a car with ski-racks (see Pezdek, 1977). Conversely, it is not assumed that the information from one external representation overwrites the information conveyed by another external representation. Thus, the mental representation will not only contain the picture information (i.e., a car without ski-racks).

To conclude, regarding text-picture processing, mental model research suggests that a single mental representation is constructed that contains information from the text as well as from the picture. When either text or pictures are presented, the mental representation contains only the information from one source.

1.3. Empirical evidence for the two-mental-representations and single-mental representation view

So far, there are no studies to our knowledge that have attempted to answer the question of which kind of mental representation is constructed when text and pictures are integrated with each other under multimedia learning conditions; moreover, there are very few studies that investigated text-picture integration in other contexts. Of course, many studies exist in which text and pictures were presented to individuals and performance measures were applied afterwards (e.g., Bransford & Johnson, 1972; Mayer & Anderson, 1991; for an overview, see Mayer, 2009). In these studies, performance has often been interpreted as evidence for the existence of an integrated representation. However, the exact nature of the underlying mental representation has not been explored any further. Other studies have tried to investigate the exact nature of the mental representations more explicitly; however, although the results indicated that text and pictures were integrated with each other, they did not aim at testing and were thus inconclusive regarding the two-mental-representations view versus single-mental-representation view (e.g., Bower & Morrow, 1990;