Picture or Text first? Explaining Sequence Effects When Learning with Pictures and Text

Alexander Eitel* and Katharina Scheiter

Knowledge Media Research Center, Tuebingen, Germany,

{a.eitel, k.scheiter}@iwm-kmrc.de

Author Note

*Correspondence concerning this article should be addressed to Dr. Alexander Eitel, Knowledge Media Research Center, Schleichstrasse 6, 72076 Tuebingen, Germany. Tel: +49 7071 979 335; Fax: +49 7071 979 126; Email: a.eitel@iwm-kmrc.de.

This research was funded by the Pact for Research and Innovation of the Competition Fund of the Leibniz Gemeinschaft.

Abstract

The present article reviews 42 studies investigating the role of sequencing of text and pictures for learning outcomes. Whereas several of the reviewed studies revealed better learning outcomes from presenting the picture before the text rather than after it, other studies demonstrated the opposite effect. Against the backdrop of theories on memory representations, these results are explained by a recency effect: That is, recall of information should be superior for the medium (text or picture) presented second, and thus, in closer temporal proximity to the assessment. As a consequence, the type of knowledge assessed (text-based vs. picture-based) and its congruence with the more recent medium should determine whether better learning results are found when presenting the picture or text first. Against the backdrop of theories on mental model construction, results were explained by a facilitation effect for the medium (text or picture) presented second. As a consequence, the relative complexity of information conveyed by the picture and by the text should determine which medium is better to be processed first, with less complex information being processed first leading to better comprehension. To conclude, the review suggests that it is not so much the sequence of text and pictures per se that affects learning outcomes than these boundary conditions (i.e., type of assessed knowledge, relative complexity of text and picture). Accordingly, the present review seeks to stimulate further research along the boundary conditions to better understand the processes involved when learning with text and pictures.

Keywords: learning with text and pictures; sequential presentation; multimedia; multiple external representations; graphic organizers

Picture or Text first? Explaining Sequence Effects When Learning with Pictures and Text

Introduction

Over the past decades, a wealth of empirical research has demonstrated that students learn better with text and pictures than with text only or picture only (see Anglin, Vaez, & Cunningham, 2004; Carney & Levin, 2002; Fletcher & Tobias, 2005; Levie & Lentz, 1982; Vekiri, 2002 for reviews). This finding is known as the multimedia effect (cf. Mayer, 2009). Multimedia effects have been found not only when text and pictures were presented simultaneously (see Mayer, 2009 for a review), but also when they were presented sequentially (e.g., McCrudden, Schraw, & Lehman, 2009).

If text and pictures are presented in a certain temporal sequence, the question is: Which sequence is better for learning – picture or text first? Existing theories on memory representations (e.g., Kulhavy, Stock, & Kealy, 1993; Schooler, 2002) and mental model construction (e.g., Schnotz, 2002; van Dijk & Kintsch, 1983) allow for explaining both why it is better for learning to process the picture before text and also why it is better for learning to process the picture after text. Accordingly, on an empirical level, some studies obtained better learning outcomes when presenting the picture before rather than after the text (e.g., Robinson, Corliss, Bush, Bera & Tomberlin, 2003; Verdi, Johnson, Stock, Kulhavy, & Whitman, 1997), whereas other studies demonstrated the exact opposite effect (e.g., Huff & Schwan, 2008; Shaw, Nihalani, Mayrath & Robinson, 2012). Therefore, in the present review we hypothesize that it is not the sequence of presenting text and pictures per se that predicts learning outcomes. Rather, it is the functions that text and pictures have for the processes and outcomes of learning that make the difference (e.g., Carney & Levin, 2002; Vekiri, 2002); amongst other influences such as prior knowledge, these functions depend on the sequence in which text and pictures are presented (cf. Ainsworth, 2006). Accordingly, results of the empirical studies reviewed in this article are analyzed with respect to the functions of text and

pictures in a given sequence. From the analysis, two boundary conditions are derived that may determine when it is better for learning to process the picture or text first. These are (1) the type of assessed knowledge, and (2) the relative complexity of information presented within the picture and the text. These boundary conditions should be considered guidelines for further research in this context. Further research would be necessary to empirically validate when and why it is better for learning to process the picture before text, or text before the picture, and thus, to be able to derive more specific instructional recommendations.

To systematically study why processing of a picture is beneficial for processing of text and vice versa, the majority of studies reviewed in the present paper investigated effects of presenting a picture before or after text in a sequential display (39 out of 42 studies). In the three remaining studies, students' eye fixations while processing text and pictures in a concurrent display were used as indicators for the sequence with which verbal and pictorial information was processed, as it is assumed that visual attention on a stimulus reflects immediate cognitive processing (cf. eye-mind hypothesis; Just & Carpenter, 1980). Whereas influential reviews in the research areas of reading (Rayner, 1998, 2009), scene perception (Henderson, 2003) and multimedia learning (van Gog & Scheiter, 2010) lend support to the eye-mind hypothesis, it is notable that there are situations under which it may and may not hold true. According to Hyönä (2010), the eye-mind hypothesis is likely to hold true if the available visual environment is relevant to the task at hand. In the studies reviewed in the present article, students were usually instructed to learn the given information in preparation of a knowledge or comprehension test. Thus, the available visual environment was indeed relevant to the task at hand so that the eye-mind hypothesis was likely to hold true for the studies reviewed in this article.

For the purpose of the review, a broad definition of the terms 'text' and 'picture' was applied. Text refers to any kind of information in a verbal code such as short or long prose, expository text or verbal instructions in a written or spoken format. The defining criterion for

text is that it comprises arbitrary symbols that are associated with the represented objects only by convention, and not by structural similarity (cf. descriptive representation; Schnotz, 2002). Pictures, by contrast, are defined as being associated with the represented object by similarity or common structural properties. Thus, photographs are defined as pictures because they are similar to what they represent (first-order isomorphism; Shah, Freedman, & Vekiri, 2005). In addition, other types of visual displays such as maps, diagrams, graphs, graphic organizers[1], matrices, geographical, or concept maps are defined as pictures even though they do not necessarily share physical similarities to what they represent, and even though parts of their structure are specified by convention (cf. Schnotz, 2002). Common to these displays, however, is that arrangements of objects in space are used to represent structural and/or conceptual features (cf. Hegarty, 2011; Larkin & Simon, 1987). For instance, objects belonging together are presented in close proximity in a diagram. Similarly, a higher bar in a graph represents a higher semantic value associated with it. Accordingly, all kinds of visual displays where space is meaningful are treated as pictures in the present article.

Applying this broad definition of text and picture, the present article reviews empirical studies from the research areas of multimedia learning, learning with graphic organizers, learning with maps and text, and learning with multiple external representations. The goal is to explain apparently contradictory findings for sequence effects when learning with text and pictures, and derives two boundary conditions that may predict under which conditions pictures are better to be processed before versus after the text. To this end, the present review refers to theories on memory representations (e.g., Kulhavy et al., 1993; Schooler, 2002) to explain when it is better for recall performance to present the picture before text, or text before the picture. Moreover, it refers to theories on mental model construction (e.g., Schnotz, 2002; van Dijk & Kintsch, 1983) to explain when it is better for comprehension to process the picture or text first. In the following section, we address how sequencing of text and pictures may affect memory representations, and therefore recall performance.

To identify potentially relevant studies for the present review, the computerized databases for research in psychology (PsycINFO) and education (ERIC) were searched by entering combinations of the keywords “learning with text and pictures”, “multimedia”, “multiple external representations”, “graphic organizers”, “graphic overviews”, “graphic advance organizers”, “graphic post organizers” and the keywords “sequence”, “order”, “presentation order”, “before”, and “after” using both “AND” and “OR” of the Boolean operators (up to January 2014). Moreover, to not miss any relevant study, we screened the articles that were cited in already identified papers, and incorporated the relevant articles.

For studies to be selected for inclusion in our review, they had to meet each of the following five criteria: (1) studies used randomized assignment to groups and quantitative data analysis; (2) text *and* pictures were processed in an identifiable sequence (either because sequence was experimentally varied or processing sequence was identifiable via eye movement data); (3) text and pictures were created by researchers (or instructors) and were not self-generated by learners; (4) learning outcomes were measured (recall, recognition, and/or comprehension of presented information); (5) results of the studies were clearly interpretable (design not confounded by extraneous factors).

In the end, 42 studies located in 26 journal articles, one PhD thesis, one conference proceeding, and one master thesis met the criteria, and hence were included for review (see Tables 1 and 2). Of these 42 studies, 16 studies directly compared learning outcomes from processing the picture before versus after the text (see studies marked with an asterisk in Tables 1 and 2). The remaining 26 studies investigated sequencing effects more indirectly by comparing whether processing the picture before versus after text was better for learning than processing text or picture only.

How Sequencing Affects Recall Performance

In this section, we review empirical studies that were mostly conducted in the context of theories on memory representations. In these studies, text and pictures that are used as to-be-learned materials usually have a high information overlap (in its extreme form containing redundant information). Pictures represent most of the information that is stated in the corresponding text in an organized manner, since it is meaningfully distributed in space (e.g., graphic organizer; Robinson, 1998). The main learning task is to recall and recognize information from text and pictures. As a result, recall and recognition performance are the main learning outcome measures. According to influential theories in this context (e.g., Kulhavy et al., 1993), better recall and recognition result from richer and more connected memory representations. Thus, the more connections can be formed with prior domain knowledge and the to-be-learned materials (i.e., text and pictures), but also between the to-be-learned materials and the learning assessment (recall and recognition test), the better the learning outcomes. Studies that are reviewed in this context yield contradictory results at first sight, with some studies showing better recall performance from presenting the picture before text, and other studies showing better recall performance from presenting the picture after text. In the following, results of these studies, together with their theoretical explanations, are presented. Subsequently, the apparent inconsistencies among the studies are resolved by referring to a recency effect (cf. Baddeley & Hitch, 1993), meaning that it is better for learning outcomes if the type of information that is assessed (text-based vs. picture-based) maps to the information that is provided by the representation (text vs. picture) that is presented last, and therefore, most recently prior to the assessment.

Better Recall from Presenting the Picture before Text

According to research in the context of the bushiness hypothesis (Baggett, 1984) and the model of working memory operations (Kulhavy et al., 1993), recall is fostered by presenting the picture before text. The bushiness hypothesis (cf. Baggett, 1984) rests on the

assumption that prior domain knowledge as well as information extracted from pictures and text are represented as concepts in memory that have a number of possible associations to be formed with other concepts. Processing a picture leads to a visual concept, which allows forming more associations with other concepts compared to a verbal concept, which is inferred from text. Thus, the visual concept is assumed to be 'bushier' than the verbal concept. When learning with pictures and text, the visual and verbal concepts will first be connected to the already existing semantic network. The number of associations that can be formed with the existing semantic network is determined by the learners' level of prior domain knowledge. If the picture (i.e., bushier concept) is processed first during the learning episode, it allows linking more of the subsequent information than if the text is processed first. This increases the likelihood of creating a compound concept containing information from both picture and text, which can foster recall performance, especially when prior knowledge is low, so that the total number of possible associations that can be formed between the existing semantic network and the visual and verbal concepts is highly constrained (i.e., bushiness hypothesis). This hypothesis was tested in an empirical study (Baggett, 1984) in which students with low prior knowledge had to recall the names of pieces of a construction kit from a film that presented the moving pictures either before (for 21sec, 14sec, 7sec), concurrently or after (for 21sec, 14sec, 7sec) the corresponding verbal narration. The sequence of presenting text and picture information was experimentally varied. In line with the bushiness hypothesis, results from both immediate and delayed testing (7 days later) revealed better recall from presenting the pictures before the narration than from presenting the pictures after narration (see Table 1). Best recall performance was achieved in conditions with concurrent narration as well as with the pictures preceding the narration by seven seconds.

The model of working memory operations (Kulhavy et al., 1993; Verdi & Kulhavy, 2002) is based on dual coding theory (DCT; Paivio, 1986). The DCT states that information

from pictures and text are encoded into two separate but connected memory stores (i.e., nonverbal and verbal memory store). Retrieving information from one memory store automatically activates the corresponding information in the other memory store so that it is sufficient to retrieve the information from one of the two stores. In consequence, the better information from the two memory stores is connected, the more easily it can be retrieved. Kulhavy et al. (1993) in their model apply DCT to the processing of maps. According to Kulhavy et al. (1993), maps have a special status in memory. They are represented in memory as intact, holistic units that can be held in working memory as a single chunk (Miller, 1956) even though they may contain considerable information about features embedded within the map framework. Thus, when a map is presented prior to corresponding text, information from the map picture can be held as an intact unit in working memory while subsequently encoding information from text without exceeding the capacity of the cognitive system. This allows for simultaneous encoding of map and text information, leading to connected memory representations and hence better retrieval.

In contrast, if the map is presented after text, it should be much more difficult to connect the information from map and text. Due to the linear format of text, it is assumed that text is represented as numerous unrelated propositions in memory. Thus, keeping all the information from text active in working memory as well as retrieving it from long-term memory into working memory requires a considerable amount of resources. If the corresponding map picture is subsequently presented, connecting information from text and map might fail because keeping text propositions active in working memory while encoding information from the map picture exceeds the capacity of the cognitive system. As a consequence, the ease of retrieving information with text before map should be inferior to the ease of retrieving information with map before text. Empirical studies that experimentally varied whether an image was presented before versus after a text yielded support for the model of working memory operations (see Table 1). First, studies showed that presenting an

image of a map prior to the corresponding text fostered recall performance in low prior knowledge learners compared to presenting the text prior to the map image (Dean & Enemoh, 1983; Verdi et al., 1997). Second, two experiments of Verdi, Kulhavy, Stock, Rittschoff, and Johnson (1996) showed that presenting biology diagrams to middle-school students prior to presenting text led to better recall and labeling performance than presenting the text before the diagrams; thereby extending the model to pictorial representations other than maps.

Similarly, studies conducted in the context of learning with graphic organizers yield support for the claim that presenting the picture before text leads to better recall than presenting the picture after the text (Robinson et al., 2003; Simmons, Griffin, & Kameenui, 1988) or the text only (Alvermann, 1981; Snouffer & Thistlethwaite, 1980). In a study by Simmons et al. (1988), delayed recall of information from a science text was better when students studied a graphic organizer before rather than after the text. Moreover, in one of three experiments conducted by Robinson et al. (2003), students were better able to recall macro-propositions and relational information when the graphic organizer was presented as a complete set before rather than after the text. Robinson et al. (2003) concluded that presenting the graphic organizer as a complete set before the text provided learners with an overarching scaffold onto which relevant details from subsequently read text could be mapped. Similar to the explanations by Baggett (1984) and Kulhavy et al. (1993), this resulted in connected memory representations and hence in better retrieval.

To sum up, research in the context of the bushiness hypothesis (Baggett, 1984) and the model of working memory operations (Kulhavy et al., 1993) suggests that learners are better able to form connections between text and picture representations when the picture is presented prior to presenting text, thereby yielding a richer and better connected memory representation fostering recall.

[Insert Table 1 about here]

Better Recall from Presenting the Picture after Text

In this section, we review studies according to which presenting the picture after text is desirable for recall performance (see Table 2). In the reviewed studies, pictures were graphic organizers (e.g., Robinson, 1998) or visualizations of dynamic scenes (e.g., Huff & Schwan, 2008). Both types of pictures basically display what is stated in the text so that there is a relatively high information overlap. Using text and pictures with such an information overlap, research has shown that presenting the text after the picture is detrimental to recognition performance. This finding is known as the verbal overshadowing effect (e.g., Meissner & Brigham, 2001). One viable explanation for the verbal overshadowing effect is a transfer-inappropriate processing shift (Chin & Schooler, 2008; Dodson, Johnson, & Schooler, 1997; Schooler, 2002). According to a transfer-inappropriate processing shift, subsequent verbalization of an initially presented picture (e.g., a picture of a face) can disrupt a holistic memory representation constructed from the picture, in turn being detrimental to performance in a recognition test. Accordingly, across the past two decades several empirical studies have demonstrated that if text is presented after the picture, memory accuracy is disrupted for recognition of various types of visual stimuli such as map configurations, faces, or cars (see Chin & Schooler, 2008; Meissner & Brigham, 2001; Meissner, Sporer, & Susa, 2008; Schooler, 2002 for reviews).

Moreover, studies have shown that if similar information is given in both text and pictures, presenting text before the picture has desirable effects on the recognition and reproduction of dynamic scenes (Huff & Schwan, 2008, 2012), as well as on the recall and application of concept relations compared to presenting the text only (Kauffman & Kiewra, 2010; Kiewra, Kauffman, Robinson, & Dubois, & Staley, 1999; Robinson, Katayama, Dubois, & Devaney, 1998, Exp. 2; Robinson & Kiewra, 1995; Robinson & Schraw, 1994). Additionally, in the study of Shaw et al. (2012), students were better able to apply knowledge about concepts and their relations when they learned with picture after text compared with

picture before text. It can be concluded that the picture representation was better accessible in the assessment when it was presented after the text rather than before the text, because it was the most recent representation prior to the assessment. This, in turn, fostered recall performance (cf. recency effect; Baddeley & Hitch, 1993). In a similar vein, in two studies of McCrudden et al. (2009), three types of dependent variables were assessed, one of which was recall of the causal sequence (explicitly depicted in picture). Results of the study revealed that presenting the picture after text led to higher outcomes on all three dependent variables than presenting text twice. As expected by a recency effect, recall of the causal sequence (as depicted in picture) profited the most from presenting the picture after text, and thus, from the picture as the most recent representation prior to the assessment (see Table 2).

Boundary Condition: Type of Assessed Knowledge

In the preceding sections some studies showed that it was better for learning outcomes to present the picture before text (e.g., Robinson et al., 2003; Verdi et al., 1997), whereas other studies revealed the exact opposite effect (e.g., Huff & Schwan, 2008; Shaw et al., 2012). These seemingly contradictory findings may be reconciled by referring to a recency effect (cf. Baddeley & Hitch, 1993). On the one hand, given low prior knowledge, inspecting the picture before the text fostered text-based recall, because this sequence of presenting information increased the likelihood of creating connected memory representations so that the text representation was better retrieved when asked to recall facts from text (cf. Baggett, 1984; Kulhavy et al., 1993). On the other hand, inspecting the picture after text fostered recall and recognition of picture information, because the picture as the most recent representation was better accessible during retrieval, thus fostering performance (cf. Baddeley & Hitch, 1993; Schooler, 2002).

Results from a study of Peverly (1981) are in line with this argumentation. In this study, recall of a story was assessed after presenting either picture before text, text before

picture, text twice or picture twice. The only factor that influenced the results was the medium that was presented second (last). Results for recall of the story were consistently better in the two conditions with the text presented last than in the two other conditions, thereby supporting the recency-effect explanation. If mainly pictorial recall or recognition is assessed, then it may be better to present the picture after text so that the picture representation is better accessible in the assessment, fostering performance. Accordingly, a simple recency effect may underlie apparently contradictory findings from studies investigating sequence effects in the context of theories on memory representations.

In conclusion, the boundary condition that may determine whether it is better to present text or pictures first, and therefore reconcile findings concerning sequence effects, is the type of assessed knowledge. When recall is mainly text-based, then it should be better for learning outcomes to present the picture before text. When recall is mainly picture-based, presenting the picture after text should foster learning outcomes. How sequence effects may be explained in the context of theories on mental model construction is addressed in the following section.

How Sequencing Affects Comprehension

In this section, we review empirical studies that were mostly conducted in the context of theories on mental model construction. In many of these studies, text and pictures are used to explain processes involved in scientific phenomena (e.g., how cell reproduction works; Stalbovs, Eitel, & Scheiter, 2013). The main learning task is to understand the scientific phenomena. As a result, comprehension of such a phenomenon is the main learning outcome measure, which is usually measured by requiring learners to draw inferences based on the presented information. It is assumed that text and pictures both contribute to constructing and updating a mental model that reflects comprehension (e.g., van Dijk & Kintsch, 1983). Depending on the level of prior knowledge and on the sequence in which text and pictures are

processed, text and pictures have different functions in the process of mental model construction, and thus, in the process of constructing comprehension. Accordingly, empirical studies found better comprehension from processing a picture before text as well as from processing text before a picture. Results of those studies will be presented and explained in the following. Subsequently, it is suggested that the relative complexity[2] of information presented in text and picture may determine whether it is better for comprehension to process the text before the picture or the picture before the text.

Better Comprehension from Processing the Picture before Text

When reading to understand text, according to the construction-integration model (van Dijk & Kintsch, 1983) a reader first constructs a mental representation of the text surface structure, from which both a propositional representation of the semantic content (i.e., text base) as well as a mental model of the specific situation described in the text are generated. The text base is constructed based solely on semantic information explicitly stated in the text. The text base alone usually yields an impoverished and often even incoherent network. To achieve better comprehension, relations that are only implicit in a text must be inferred to yield a coherent mental structure (Glenberg & Langston, 1992). Thus, understanding a text often requires interpreting the text by integrating text propositions with prior knowledge, mentally created images, or information extracted from a previously inspected picture (e.g., Bransford & Johnson, 1972).

However, especially readers with low prior knowledge sometimes fail to construct a coherent mental model of the situation described in a text (cf. Bransford & Johnson, 1972). They construct a mental model that inadequately reflects the contents or situations described in a text, thereby hampering comprehension (Schnotz & Bannert, 2003; Schnotz & Kürschner, 2008). By contrast, if prior knowledge is high (cf. McNamara, Kintsch, Songer, & Kintsch, 1996), or if a picture is presented prior to reading the corresponding text, the process

of constructing an adequate mental model from text, and hence comprehension, can be facilitated. Unlike text, pictures are related to their represented referents via structural similarity or commonality (cf. Hegarty, 2011; Schnotz, 2002) so that spatial relations expressed among the objects in a picture can be mapped onto the corresponding semantic relations to provide the structure of the mental model (analogical structure mapping; Schnotz & Bannert, 2003). This means that information about the structural relations among the objects in a picture is preserved within the mental model (cf. Johnson-Laird, 1980). As a consequence, a mental model can be directly constructed from the picture without requiring much interpretation or inference of additional information (Glenberg & Langston, 1992; Larkin & Simon, 1987; Hegarty & Just, 1993). The picture is considered to be one possible expression of a mental model (Gyselinck & Tardieu, 1999; Gyselinck, Jamet, & Dubois, 2008).

Thus, processing of a picture may initially provide learners with the structure of a mental model so that part of the mental model construction process may already be completed based on the picture. When processing subsequent text, corresponding steps of mental model construction are not needed anymore. Thus, instead of having to construct a mental model from scratch, initial picture inspection may provide learners with a mental scaffold facilitating subsequent processes of mental model construction from text (cf. Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; Gyselinck et al., 2008; Schnotz & Bannert, 2003). Accordingly, in the studies by Eitel, Scheiter, Schüler, Nyström et al. (2013) as well as by Eitel, Scheiter, and Schüler (2013), presenting a causal system picture to low prior knowledge learners before presenting the corresponding text led to better comprehension and faster reading of text about the system's spatial structure compared with presenting just text. These effects held true even if the initial picture presentation was very short (i.e., 600ms or 2sec), suggesting that providing low prior knowledge learners with the global structure of an adequate mental model (i.e., a mental scaffold) can have beneficial effects on subsequent

mental model construction, and thus on comprehension (see Table 1). Further evidence for this assumption comes from a recent study by Stalbovs et al. (2013), which shows that initially attending to the picture instead of attending to the text was related to more successful learning with multimedia about the biological processes of mitosis and meiosis. In a similar vein, Salmerón, Baccino, Cañas, Madrid, and Fajardo (2009) showed that reading a graphical overview at the beginning of a difficult hypertext presentation was related to improvements in comprehension (especially when prior knowledge was low), whereas reading the overview at the end of an easy hypertext was related to a decrease in hypertext comprehension. The authors concluded that initially processing the overview increased the salience of the hypertext structure, thereby supporting low prior knowledge learners in generating inferences based on subsequent text.

Moreover, due to the specific nature of pictures (Stenning & Oberlander, 1995), the mental scaffold provided by the picture may constrain the range of (erroneous) interpretations or inferences that are made based on the text (cf. Ainsworth, 2006; Scaife & Rogers, 1996). In particular, pictures can assist in the process of constructing a mental model from text because they can make relations explicit that are only implicitly conveyed by the text (cf. Glenberg & Langston, 1992; Gyselinck & Tardieu, 1999; McCrudden, Magliano, & Schraw, 2011; Zwaan & Radvansky, 1998). Thus, pictures may give a specific example on how to interpret text (cf. interpretation function; Levin, Anglin, & Carney, 1987). In the case of a well-designed picture, this can make text more coherent and comprehensible, thus fostering understanding (Carney & Levin, 2002; Gyselinck et al., 2008). Accordingly, in a study by Bransford and Johnson (1972) comprehension of a text passage was improved when a picture about the situation described in the passage was presented prior to the text (see Borges & Robins, 1980 for a replication). Comprehension was improved compared with presenting just text and compared with presenting the picture after text. Moreover, presenting the (coherent) picture before text was also better than presenting the picture before text when the picture contained

the same objects but in a rearranged manner (partial context). Bransford and Johnson (1972) concluded that the appropriate context given by the (coherent) picture before text led to better comprehension; for the context to be helpful, it was required that the relations between the objects described in the text were provided by the initial picture – understanding the relations within the context was a prerequisite for understanding the events suggested by the passage. Similarly, McCrudden et al. (2011, Exp. 1) showed that presenting a causal diagram prior to presenting text led to better learning outcomes for sentences that semantically overlapped with the diagram and shorter reading times than when learning with just text. The authors concluded that diagrams helped by making relations explicit, thus facilitating subsequent processing of text.

According to Schnotz (2005) presenting the picture after text may even provide learners with a disadvantage that is absent when pictures are processed prior to text. According to Schnotz a text never describes a subject matter with enough detail to fit just one single picture or one mental model. Thus, a mental model constructed from just text will always differ in some respects from the picture that illustrates the subject matter. If such a text was presented prior to the corresponding picture, the picture would likely interfere with the mental model initially constructed from text, thus being detrimental to comprehension. In contrast, if the picture was presented before the text, subsequent mental model construction would be based on the specific mental model initially constructed from the picture, thus fostering comprehension. This assumed superiority of presenting pictures before rather than after the text is called the picture-text sequencing effect (Schnotz, 2005). Accordingly, in an empirical study in which students learned with text and pictures about the principle of plate tectonics, Ullrich (2011) showed that presenting the picture before text in a sequential format led to better recall and comprehension than presenting the text before the picture (see Table 1).

To sum up, initially processing pictures can facilitate processing of text by constraining interpretation (Ainsworth, 2006), and thus by resolving ambiguity that is usually present in text. Moreover, information extracted from the picture can act as a scaffold to facilitate the process of constructing an adequate mental model, which in turn fosters comprehension, especially for learners low in prior knowledge (e.g., Eitel, Scheiter, Schüler, Nyström et al., 2013; McCrudden et al., 2011).

Better Comprehension from Processing the Picture after Text

Similar to processing text when processing pictures with the goal of understanding their displayed contents, learners are assumed to construct a mental model (cf. van Dijk & Kintsch, 1983). According to influential models of learning with text and pictures such as the cognitive theory of multimedia learning (Mayer, 2009) or the integrative model of text and picture comprehension (Schnotz, 2002), constructing a mental model from a picture roughly involves two processing steps. First, relevant information from pictures has to be perceived or selected from the instruction. According to Schnotz (2002), this process takes place in a largely automated manner by making use of perceptual processes and visual routines. The learner creates a perceptual representation of the visuo-spatial relations depicted in the picture. In a second step, visuo-spatial relations from the perceptual representation are then mapped onto semantic relations to provide the structure of the mental model (analogical structure mapping; Schnotz & Bannert, 2003). According to Mayer (2009), selected images are organized into a pictorial mental model by establishing connections between parts of the picture.

When learning with complex pictures, however, *selecting* the relevant information that is later used for mental model construction may be difficult, which in turn may impair comprehension. In a complex graph, such as a complex weather map in the domain of meteorology, task-relevant information may need to be selected from a much larger amount of

displayed information (Canham and Hegarty, 2010). In such a complex graph it may be hard for students to distinguish between which information is relevant and which information is irrelevant with regard to solving the current task, especially when prior knowledge levels of learners are low. With increasing prior knowledge or expertise, however, students learn to separate task-relevant from task-irrelevant information so that they select only the relevant information and ignore the irrelevant information (cf. information reduction hypothesis; Haider & Frensch, 1996). Studies using materials from meteorology (Canham & Hegarty, 2010; Lowe, 1993, 1994, 1996, 2004), medicine (Lesgold et al., 1988), art (Antes & Kristjanson, 1991), chess playing (Charness, Reingold, Pomplun, & Stampe, 2001) or biology (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010) provide evidence for the information reduction hypothesis, showing that more expert students focus more on elements that are thematically relevant than novice students do. Accordingly, a higher level of prior knowledge or expertise can lead to selecting more relevant information from a complex picture, which in turn can be helpful for comprehension (e.g., Canham & Hegarty, 2010).

In other words, high prior knowledge can constrain the process of selecting information from a complex picture, in turn being helpful to mental model construction. One way to increase prior knowledge levels of students before they learn with complex pictures is to initially provide them with domain knowledge given in a text as was done in two experiments reported in Canham and Hegarty (2010). In these studies, novice students were either taught or not taught the principles of meteorology using mainly text prior to processing complex weather maps (text-picture sequential format). Eye movements as well as the ability to draw inferences from the weather maps were compared between students who were taught the principles of meteorology initially (i.e., high prior knowledge students) and students who were not taught the principles of meteorology initially (i.e., low prior knowledge students). Results revealed that high prior knowledge students attended more to task-relevant information in the maps than low prior knowledge students did, which resulted in superior

performance in inference generation from the weather map (see Table 2). These results suggest that task-relevant knowledge acquired from initially presented verbal instructions effectively guided attention to relevant parts in the picture, in turn fostering inference making (comprehension). This suggests that the text guided and constrained the information selection process from the complex picture, which in turn fostered comprehension.

The idea of text-guided processing of pictures has received empirical support in research on learning with text and pictures and in related domains. When presenting text and pictures concurrently, the text was used as a guide on how to process the concurrently presented picture (Folker, Ritter, & Sichelschmidt, 2005; Hegarty & Just, 1993; Ozcelik, Arslan-Ari, & Cagiltay, 2010; Rummer, Schweppe, Fürstenberg, Scheiter, & Zindler, 2011; Schmidt-Weigand, Kohnert, & Glowalla, 2010a, 2010b; Schwonke, Berthold, & Renkl, 2009; Van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009). Thus, text guidance may be helpful for comprehension not only when it provides (additional) content information, but also when it guides attention to the relevant parts in the picture without providing further content information. Such effects have been found in the signaling literature. Here, several types of cues guided attention towards relevant parts in complex static and dynamic learning materials without giving additional content information (e.g., Bétrancourt, 2005; Canham & Hegarty, 2010; De Koning et al., 2009; Hegarty, Kriz, & Cate, 2003; Jarodzka et al., 2010; Mautone & Mayer, 2001; Ozcelik et al., 2010; Scheiter & Eitel, 2010; Van Gog, Jarodzka, Scheiter, & Gerjets, 2009).

To conclude, given low prior knowledge, pictures may foster comprehension when processed after text, because information extracted from initially processed text can act as a guide to facilitate the selection of relevant information subsequently presented in the picture (Canham & Hegarty, 2010; Hegarty & Just, 1993).

[Insert Table 2 about here]

Boundary Condition: Relative Complexity

As shown by studies in the previous sections, inspecting the picture both before and after the text can foster comprehension. On the one hand, studies showed that processing the picture before text helped to constrain interpretation of text that was ambiguous and hard to understand without sufficient background knowledge or context, thereby fostering comprehension via facilitated mental model construction (e.g., Bransford & Johnson, 1972; Glenberg & Langston, 1992; Schnotz, 2005). On the other hand, other studies showed that initially processed text guided attention towards the relevant parts of a subsequently presented complex picture, thereby fostering comprehension (e.g., Canham & Hegarty, 2010; Hegarty & Just, 1993; Lowe, 2004). One may conclude that it is helpful to learning if the medium that contains less complex information is presented first. As a result, information presented in the first medium is more likely to be understood even for low prior knowledge students, and thus, it can guide or facilitate processing of the more complex information presented in the other medium. Accordingly, the boundary condition that may determine whether it is better for comprehension to process the picture or text first is the relative complexity of picture and text.

This argumentation is in line with Ainsworth (2006), stating that it is reasonable to start an instruction by presenting the least complex representations to the learner. Moreover, this argumentation is in line with assumptions made by the elaboration theory of instruction (Reigeluth, Merrill, Wilson, & Spiller, 1980). According to this theory, an instruction should be presented in a way that the less detailed and less complex information should be presented first, and thus prior to presenting more detailed and complex information. In analogy to a zoom lens, the theory prescribes that an instruction should begin with a wide-angle view of the subject matter, which shows the major relationships among those parts but which still lacks in details. Afterwards, the subject matter should be divided into the subparts (“zooming in”) so that students can elaborate on each subpart. This zooming in should be continued until the desired level of detail is reached. This type of sequencing an instruction in an easy-to-

complex manner has received much empirical support (e.g., Ainsworth, Wood, & O'Malley, 1998; Weidenmann, Paechter, & Hartmannsgruber, 1999). In the context of learning with text and pictures, one would conclude that the medium containing less complex information, whether text or the picture, should be presented first. As such, it can facilitate processing of the medium presented second (e.g., via constraining interpretation or attention guidance; Ainsworth, 2006; Hegarty & Just, 1993), and thus foster comprehension.

Further Research along Boundary Conditions

To sum up, we reviewed empirical studies that were conducted in the context of theories on memory representations (e.g., Kulhavy et al., 1993) and on mental model construction (e.g., van Dijk & Kintsch, 1983). As the present review suggests, a recency effect may explain apparently contradictory findings from studies investigating sequence effects in the context of theories on memory representations. In conclusion, the type of assessed knowledge (text-based vs. picture-based recall) is assumed to moderate whether it is better for learning to present the picture before or after the text. Whereas a picture-before-text sequence should lead to better recall in a text-based assessment, a picture-after-text sequence should lead to better picture-based recall and recognition. The studies reviewed in this article seem to support this hypothesis (see previous sections). However, in the context of theories on memory representations, most studies that directly compared presenting the picture before versus after the text used an assessment that was based on information from both text and picture. Fewer studies used a merely text-based assessment and, to our knowledge, there are so far no studies that directly compare presenting the picture before versus after the text and use a merely picture-based assessment (see Table 3). Such research, however, would be crucial to empirically validate the recency-effect explanation of sequencing effects as formulated within the present article. Further research should therefore systematically

manipulate the sequence of presenting text and pictures together with the type of assessed knowledge (text-based vs. picture-based).

In addition, the present review suggests that the relative complexity of the picture and text may explain findings of better comprehension from studies conducted in the context of theories on mental model construction. The reviewed studies seem to support the hypothesis that it is helpful for comprehension if the medium that contains the less complex information (text or picture) is presented first, and thus may guide or facilitate processing of the more complex information presented in the second medium (text or picture). However, so far there exist only few studies that directly investigate this (see Table 3). Accordingly, further studies that systematically investigate the effects of text-picture versus picture-text sequences in combination with the relative complexity of text and pictures on comprehension outcomes are needed. Results of such studies could provide additional empirical support in favor of our hypothesis that an easy-to-complex sequencing of multimedia instructions could indeed explain effects of better comprehension, regardless of whether the picture or text would be presented first. Hence, this research would contribute to our knowledge about the interplay between the dimensions of sequencing and complexity in the process of mental model construction.

In conclusion, the ultimate goal of the present review was to generate informed hypotheses based on the given research evidence about how to explain sequencing effects when learning with pictures and text. The present review seeks to stimulate further research that more systematically tests for the validity of the proposed hypotheses (boundary conditions) to better understand the processes involved when learning with pictures and text.

[Insert Table 3 about here]

In the present review, we made use of two distinct explanations for sequence effects when learning with text and pictures (i.e., recency effects; facilitated processing of medium

presented second), but these explanations may not be specific to the situation of learning with text and pictures. For instance, since the 1960s recency effects have been well-established in memory research, where they were mostly studied using unrelated word lists (e.g., Murdock Jr., 1962). This suggests that recency effects are not bound to the situation of learning with text and pictures. Moreover, the idea of an easy-to-complex sequencing of representations with the intention to facilitate comprehension (as suggested in the present review) may also not be specific to learning with text and pictures only. For instance, in a mathematical learning environment designed for primary school children (COPPERS; Ainsworth et al., 1998), coin problems were presented to children via increasingly abstract representations: First as pictures, then as a mixture of text and pictures, then as text only, and then as algebra. One may assume that this sequence was better for learning because initially acquired comprehension of the more concrete or easy representation (i.e., realistic picture) facilitated processing and comprehension of the more abstract and complex representation presented later in the sequence (i.e., algebra). However, to our knowledge there is not much empirical research investigating the effectiveness of easy-to-complex sequencing compared to other types of sequencing of representations. Hence, it remains to be tested in empirical studies whether sequence effects can generally be explained by facilitated processing of more complex representations due to the initial processing of easier representations.

Regardless of their generalizability, in the present review explanations of sequence effects, namely recency effects and facilitated processing of the medium presented second, were assumed to be independent. This makes sense, considering that for recency effects to apply, the congruency between the format of the most recent representation at learning (text or picture), and the representation format of the assessment (text-based or picture-based recall), is important. By contrast, according to theories on comprehension and mental model construction, the format of the assessment is not important. Comprehension is assumed to be a modality- or media-unspecific construct such that better comprehension would be equally

applicable to text-based and picture-based assessments (cf. Gernsbacher, Varner, & Faust, 1990). Accordingly, studies conducted in the two research contexts (memory vs. comprehension) that investigated mainly recall or mainly comprehension were treated separately in this review.

However, the learning outcome measures of recall and comprehension may not be entirely independent of each other. On the one hand, to demonstrate comprehension of a subject matter in a subsequent assessment, one has to recall what one had understood initially. On the other hand, correctly understanding a subject matter often requires processing it on a deeper level to be able to draw the required inferences, and deeper processing is known to facilitate recall in addition to facilitating comprehension (cf. Craik & Lockhart, 1972; Salomon, 1984). Thus, to test whether recall and comprehension outcomes, and therefore, whether their two separate theoretical explanations (recency effects; facilitated processing of medium presented second) are indeed independent of each other, future empirical research should investigate whether systematically manipulating the relative complexity of text and pictures may interact with the systematic manipulation of the assessment type (recall vs. comprehension) when studying sequence effects in learning from text and pictures. Such research should take care that the learning outcome measures are valid and reliable in assessing the constructs of recall and comprehension.

Another interesting direction for future research would be to continue analyzing processing data when studying effects of the sequence of presenting text and pictures. Presenting text and pictures in a sequential manner has a large advantage compared to presenting text and pictures simultaneously; that is, the former allows studying in isolation how the medium presented first (picture or text) affects processing and learning from the medium presented second (text or picture). This can provide valuable information, especially when process data is analyzed. For instance, by analyzing the eye movements of students learning with a sequential text-before-picture presentation, Canham and Hegarty (2010) were

able to provide empirical data in favor of the claim that processing of a text prior to inspecting a complex picture can be helpful to comprehension, because information extracted from the text guides attention towards the corresponding relevant information in the picture. Similarly, other types of processing data such as think-aloud protocols or self-explanations have been shown to provide valuable information regarding processes taking place when learning with text and pictures (e.g., Ainsworth & Loizou, 2003; Butcher, 2006; Chi, 2000). Hence, further research may continue making use of such data to study how processing of text may interact with the processing of pictures. This may provide further information about processes that underlie successful learning with text and pictures, thereby providing a basis from which instructional recommendations may be derived in the future.

The Influence of Learner Characteristics

Future research regarding this topic should also focus more on the influence of certain learner characteristics such as prior knowledge, reading abilities, or visuo-spatial abilities, since they might strongly influence effects of the learning instruction. For instance, visuo-spatial abilities might play a role because presenting the picture before text might reduce the degree of required visuo-spatial reasoning based on the text. In the studies of Eitel, Scheiter, and Schüler (2013) and of Eitel, Scheiter, Schüler, Nyström et al. (2013), presenting a picture of a pulley system before text fostered comprehension and sped up processing of subsequent text about the system's spatial structure (compared to presenting text only). It was concluded that part of the required mental model construction was already completed based on the initial picture inspection, which facilitated subsequent visuo-spatial reasoning processes based on the text, thereby speeding up the reading process and fostering comprehension. One might conclude that this facilitating function of the picture would be especially helpful for learners with low visuo-spatial abilities. If, in contrast, the picture is presented after text, then students would first need to construct a mental model based on text only, which would require a higher

degree of visuo-spatial reasoning that could be detrimental especially for learners low in visuo-spatial abilities, whereas learners high in visuo-spatial abilities might be able to compensate for the missing picture in the initial position.

In a similar vein, a study of Dean and Enemoh (1983) has shown that presenting the picture before text could compensate for low prior knowledge levels. When the picture was presented before text in their study, students low in prior knowledge scored equally high on a free recall test as students high in prior knowledge, and higher than when low prior knowledge students received the picture after text. Referring to theories on memory representations (e.g., Kulhavy et al., 1993), one may explain these findings by assuming that either prior knowledge or the picture in the primary position provided learners with an organized mental structure that allowed for connecting and integrating subsequent text information, hence fostering retrieval. Thus, it appears that especially students with low prior knowledge may profit from presenting the picture before text, while students high in prior knowledge may not necessarily need this kind of help. It is conceivable that high prior knowledge students might even benefit from having the more demanding task of first trying to understand the text on their own, without having the help of supporting pictures. So far, however, systematic research concerning this potential moderating role of prior knowledge when presenting text and pictures in different sequences is missing. Therefore, an important avenue for further research is to study the role of relevant learner characteristics (e.g., prior knowledge, reading abilities, visuo-spatial abilities) on learning with different text-picture sequences.

The Influence of Segment Size and Pacing

Two other relevant factors with respect to sequencing effects when learning with pictures and text are the size of the segments and the pacing of the sequence. Several of the studies reviewed in this article investigated effects of an instructor-paced and coarse-grained

sequence of presenting text and pictures (25 in total); that is, they addressed the situation of presenting the whole picture once before or after presenting the whole text (e.g., Robinson et al., 2003; Ullrich, 2011; Verdi et al., 1997). Some other studies investigated effects of multiple cycles of text-picture processing (10 in total), in which only part of the information from text and picture was given within each cycle (e.g., Baggett, 1984; Robinson et al., 1998; Shaw et al., 2012). From the point of view of the temporal contiguity principle, presenting text and pictures in close temporal proximity is generally seen as more effective for learning than presenting them in a temporally discontiguous manner (see Ginns, 2006; Mayer, 2009 for overviews). Accordingly, it could be assumed that performance decreases along a continuum from a simultaneous presentation to a fine-grained sequential presentation to a coarse-grained sequential presentation (see also Mayer & Anderson, 1991, 1992; Mayer, Moreno, Boire, & Vagge, 1999; Mayer & Sims, 1994), especially when learning materials are complex (Ginns, 2006) and when they are presented in a short and system-paced manner (cf. segmenting principle; Mayer, 2009). While we do not doubt this, the present review nevertheless shows that even instructor-paced and coarse-grained sequential presentations of text and pictures produced better learning outcomes than presenting text only or picture only (e.g., McCrudden et al., 2011). Whether segment size and pacing also moderate effects of presenting the picture before versus after text in a sequential display remains to be subject for further empirical research.

Summary and Conclusions

In the present article, studies were reviewed that showed better learning outcomes from presenting the picture before text as well as from presenting text before the picture. At first sight, the reviewed studies revealed a mixed pattern of results regarding whether it is better for learning to process a picture or text first. While in some studies presenting the picture before text was better for learning outcomes (e.g., Dean & Enemoh, 1983; Robinson et

al., 2003), other studies revealed the exact opposite effect (e.g., Huff & Schwan, 2008; Shaw et al., 2012). Against the backdrop of theories on memory representations and mental model construction, in the present article we hypothesized that two boundary conditions, namely (1) the type of assessed knowledge, and (2) the relative complexity of information conveyed by the picture and by text, would determine whether it is better for learning outcomes to process the picture or text first. Whereas the reviewed studies tended to support our hypotheses, the present review also shows that systematic research still has to be done to provide sufficient empirical evidence in favor of our claims (e.g., research using picture-based assessments in the context of sequencing effects).

Accordingly, with this review we want to give guidelines for further research, which is research that is conducted along our hypothesized boundary conditions. Such research could provide evidence for our hypotheses regarding the cognitive processes that may underlie effects of the sequence of presenting text and pictures. Understanding which cognitive processes are responsible for a certain sequential presentation to be better for learning might provide valuable information about which processes ought to be stimulated to foster the learning success in the future. Thus, in the long run such information may provide the theoretical basis from which more specific instructional recommendations could be derived about when and how to process pictures and text to foster the learning success.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

References marked with an asterisk comprise studies included in the review.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183–198. doi:10.1016/j.learninstruc.2006.03.001

Ainsworth, S. E., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*, 669-681. doi:10.1016/S0364-0213(03)00033-8

Ainsworth, S., Wood, D., & O'Malley, C. (1998). There is more than one way to solve a problem: Evaluating a learning environment that supports the development of children's multiplication skills. *Learning and Instruction, 8*, 141-157. doi:10.1016/S0959-4752(97)00013-3

*Alvermann, D. E. (1981). The compensatory effect of graphic organizers on descriptive text. *The Journal of Educational Research, 75*, 44-48.

Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 865-916). Mahwah, NJ: Lawrence Erlbaum.

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating artists from nonartists by their eye-fixation patterns. *Perceptual and Motor Skills, 73*, 893-894. doi:10.2466/pms.1991.73.3.893

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition, 21*, 146-155. doi:10.3758/BF03202726

*Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology, 76*, 408. doi: 10.1037/0022-0663.76.3.408

Bétrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). Cambridge, MA: Cambridge University Press.

*Borges, M. A., & Robins, S. L. (1980). Contextual and motivational cue effects on the comprehension and recall of prose. *Psychological Reports, 47*, 263-268. doi:10.2466/pr0.1980.47.1.263

*Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726. doi:10.1016/S0022-5371(72)80006-9

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*, 182-197. doi:10.1037/0022-0663.98.1.182

*Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155-166. doi:10.1016/j.learninstruc.2009.02.014

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*, 5-26. doi:10.1023/A:1013176309260

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory and Cognition, 29*, 1146-1152. doi:10.3758/BF03206384

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20*, 396-413. doi:10.1080/09541440701728623

Craik, F. I. M., & Lockhart, S. E. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. doi:10.1016/S0022-5371(72)80001-X

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*, 113–140.

*Dean, R. S., & Enemoh, P. A. C. (1983). Pictorial organization in prose learning. *Contemporary Educational Psychology, 8*, 20-27. doi:10.1016/0361-476X(83)90031-0

*Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization in prose learning. *Journal of Educational Psychology, 73*, 57-64. doi:10.1037/0022-0663.73.1.57

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition, 25*, 129-139. doi:10.3758/BF03201107

*Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology, 27*, 451-461. doi:10.1002/acp.2922

*Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction, 28*, 48-63. doi:10.1016/j.learninstruc.2013.05.002

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

Folker, S., Ritter, H., & Sichelschmidt, L. (2005). Processing and integrating multimodal material: The influence of color-coding. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 690–695). Mahwah, NJ: Erlbaum.

Gernsbacher, M. A., Varner, K. R., & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 430-445. doi:10.1037/0278-7393.16.3.430

Ginns, P. (2006). Integrating information: a meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16,* 511-525. doi:10.1016/j.learninstruc.2006.10.001.

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. doi:10.1016/0749-596X(92)90008-L

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology, 22,* 353-374. doi:10.1002/acp.1411

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom, and why? In H. van Oostendorp, & S. Goldman (Eds.), *The construction of mental representations during reading* (pp. 195-218). Mahwah, NJ: Lawrence Erlbaum.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30,* 304-337. doi:10.1006/cogp.1996.0009

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3,* 446-474. doi: 10.1111/j.1756-8765.2011.01150.x

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717-742. doi:10.1006/jmla.1993.1036

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21,* 325-360. doi:10.1207/s1532690xci2104_1

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498-504. doi:10.1016/j.tics.2003.09.006

*Huff, M., & Schwan, S. (2008). Verbalizing events: Overshadowing or facilitation?. *Memory and Cognition, 36*, 392-402. doi:10.3758/MC.36.2.392

*Huff, M., & Schwan, S. (2012). The verbal facilitation effect in learning to tie nautical knots. *Learning and Instruction, 22*, 376-385. doi:10.1016/j.learninstruc.2012.03.001

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172-176. doi:10.1016/j.learninstruc.2009.02.013

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*, 146-154. doi:10.1016/j.learninstruc.2009.02.019

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science, 4*, 71-115. doi:10.1016/S0364-0213(81)80005-5

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354. doi:10.1037//0033-295X.87.4.329

*Kauffman, D. F., & Kiewra, K. A. (2010). What makes a matrix so effective? An empirical test of the relative benefits of signaling, extraction, and localization. *Instructional Science, 38*, 679-705. doi:10.1007/s11251-009-9095-8

*Kiewra, K. A., Kauffman, D. F., Robinson, D. H., Dubois, N. F., & Staley, R. K. (1999). Supplementing floundering text with adjunct displays. *Instructional Science, 27*, 373-401. doi:10.1023/A:1003270723360

Kulhavy, R. W., Stock, W. A., & Kealy, W. A. (1993). How geographic maps increase recall of instructional text. *Educational Technology Research and Development, 41*, 47-62. doi:10.1007/BF02297511

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99. doi:10.1016/S0364-0213(87)80026-5

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ, England: Lawrence Erlbaum.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology, 30,* 195-232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: Vol. 1, Basic research* (pp. 51-85). New York: Springer-Verlag.

Lowe, R. K. (1993). Constructing a mental representation from an abstract technical diagram. *Learning and Instruction, 3,* 157-179. doi:10.1016/0959-4752(93)90002-H

Lowe, R. K. (1994). Selectivity in diagrams: Reading beyond the lines. *Educational Psychology, 14*, 467-491. doi:10.1080/0144341940140408

Lowe, R. K. (1996). Background knowledge and the construction of a situational representation from a diagram. *European Journal of Psychology of Education, 11*, 377-397. doi:10.1007/BF03173279

Lowe, R. K. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction, 14*, 257-274. doi:10.1016/j.learninstruc.2004.06.003

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93,* 377-389. doi:10.1037/0022-0663.93.2.377

Mayer, R. E. (2009). *Multimedia learning. 2nd edition*. Cambridge: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490. doi:10.1037/0022-0663.83.4.484

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452. doi:10.1037/0022-0663.84.4.444

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638-643. doi:10.1037/0022-0663.91.4.638

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. doi:10.1037/0022-0663.86.3.389

*McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011). The effect of diagrams on online reading processes and memory. *Discourse Processes, 48*, 69-92. doi:10.1080/01638531003694561

*McCrudden, M. T., Schraw, G., & Lehman, S. (2009). The use of adjunct displays to facilitate comprehension of causal relationships in expository text. *Instructional Science, 37*, 65-86. doi:10.1007/s11251-007-9036-3

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43. doi:10.1207/s1532690xci1401_1

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*, 603-616. doi:10.1002/acp.728

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20,* 414-455. doi:10.1080/09541440701728581

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63,* 81-97. doi:10.1037/h0043158

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482-488. doi:10.1037/h0045106

*O'Keefe, E. J., & Solman, R. T. (1987). The influence of illustrations on children's comprehension of written stories. *Journal of Literacy Research, 19,* 353-377. doi:10.1080/10862968709547611

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26,* 110-117. doi:10.1016/j.chb.2009.09.001

Paivio, A. (1986). *Mental representations: a dual coding approach.* Oxford, UK: Oxford University Press.

*Peverly, S. (1981). *The effects of diagrams-before text vs. diagrams-after text in the processing of novel text information.* Retrieved from ERIC database. (ED212995)

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62,* 1457-1506. doi:10.1080/17470210902816461

Reigeluth, C. M., Merrill, M. D., Wilson, B. G., & Spiller, R. T. (1980). The elaboration theory of instruction: A model for sequencing and synthesizing instruction. *Instructional Science, 9,* 195-219. doi:10.1007/BF00177327

Robinson, D. H. (1998). Graphic organizers as aids to text learning. *Reading Research and Instruction, 37*, 85-105. doi:10.1080/19388079809558257

*Robinson, D. H., Corliss, S. B., Bush, A. M., Bera, S. J., & Tomberlin, T. (2003). Optimal presentation of graphic organizers and text: A case for large bites?. *Educational Technology Research and Development, 51*, 25-41. doi:10.1007/BF02504542

*Robinson, D. H., Katayama, A. D., Dubois, N. F., & Devaney, T. (1998). Interactive effects of graphic organizers and delayed review on concept application. *The Journal of Experimental Education, 67*, 17-31. doi: 10.1080/00220979809598342

*Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology, 87*, 455-467. doi:10.1037/0022-0663.87.3.455

*Robinson, D. H., & Schraw, G. (1994). Computational efficiency through visual argument: Do graphic organizers communicate relations in text too effectively? *Contemporary Educational Psychology, 19*, 399-415. doi:10.1006/ceps.1994.1029

Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied, 17*, 159-173. doi:10.1037/a0023588

*Salmerón, L., Baccino, T., Cañas, J. J., Madrid, R. I., & Fajardo, I. (2009). Do graphical overviews facilitate or hinder comprehension in hypertext?. *Computers and Education, 53*, 1308-1319. doi:10.1016/j.compedu.2009.06.013

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*, 647-658. doi:10.1037/0022-0663.76.4.647

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*, 185-213. doi:10.1006/ijhc.1996.0048

Scheiter, K., & Eitel, A. (2010). The effects of signals on learning from text and diagrams: how looking at diagrams earlier and more frequently improves understanding. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference – 6th International Conference, Diagrams 2010* (LNAI 6170, pp. 264-270). Heidelberg: Springer.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction, 20,* 100-110. doi:10.1016/j.learninstruc.2009.02.011

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24,* 226–237. doi:10.1002/acp.1554

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14,* 101-120. doi:10.1023/A:1013136727916

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). Cambridge, MA: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13,* 141-156. doi:10.1016/S0959-4752(02)00017-8

Schnotz, W., & Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science, 36,* 175-190. doi:10.1007/s11251-007-9029-2

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16,* 989-997. doi:10.1002/acp.930

Schwonke, R. Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*, 1227-1243. doi:10.1037/a0013247

Shah, P., Freedman, E. G., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 426-476). New York, NY: Cambridge University Press.

*Shaw, S., Nihalani, P., Mayrath, M., & Robinson, D. H. (2012). Graphic organizers or graphic overviews? Presentation order effects with computer-based text. *Educational Technology Research and Development, 60*, 807-820. doi:10.1007/s11423-012-9257-2

*Simmons, D. C., Griffin, C. C., & Kameenui, E. J. (1988). Effects of teacher-constructed pre-and post-graphic organizer instruction on sixth-grade science students' comprehension and recall. *The Journal of Educational Research, 82*, 15-21.

*Snouffer, N. K., & Thistlethwaite, L. L. (1980). The effects of the structured overview and vocabulary pre-teaching upon comprehension levels of college freshmen reading physical science and history materials. *Journal of the Association for the Study of Perception, 15*, 11-16.

*Stalbovs, K., Eitel, A., & Scheiter, K. (2013, August). *Which cognitive processes predict successful learning with multimedia?* Paper presented at the Fifteenth Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Munich, Germany.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19*, 97-140. doi:10.1016/0364-0213(95)90005-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296. doi:10.1023/A:1022193728205

*Ullrich, M. (2011). *Einflüsse der Verarbeitungsreihenfolge auf den Wissenserwerb mit Texten und Bildern* [Influences of processing order on knowledge acquisition from text and images] (Doctoral dissertation, University of Koblenz-Landau). Retrieved from http://kola.opus.hbz-nrw.de/volltexte/2011/631/

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

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25,* 785-791. doi:10.1016/j.chb.2009.02.007

Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25,* 325-331. doi:10.1016/j.chb.2008.12.021

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20,* 95-99. doi:10.1016/j.learninstruc.2009.02.009

Vekiri, I. (2002). What Is the Value of Graphical Displays in Learning? *Educational Psychology Review, 14,* 261-312. doi:10.1023/A:1016064429161

*Verdi, M. P., Johnson, J. T., Stock, W. A., Kulhavy, R. W., & Whitman-Ahern, P. (1997). Organized spatial displays and texts: Effects of presentation order and display type on learning outcomes. *The Journal of Experimental Education, 65,* 303-317.

Verdi, M. P., & Kulhavy, R. W. (2002). Learning with maps and texts: An overview. *Educational Psychology Review, 14,* 27-46. doi:10.1023/A:1013128426099

*Verdi, M. P., Kulhavy, R. W., Stock, W. A., Rittschof, K. A., & Johnson, J. T. (1996). Text learning using scientific diagrams: Implications for classroom use. *Contemporary Educational Psychology, 21*, 487-499. doi:10.1006/ceps.1996.0033

Weidenmann, B., Paechter, M., & Hartmannsgruber, K. (1999). Structuring and sequencing of complex text-picture combinations. *European Journal of Psychology of Education, 14*, 185-202. doi:10.1007/BF03172965

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185. doi:10.1037/0033-2909.123.2.162

Footnotes

[1]Graphic organizers are spatial adjuncts to text that use visual information such as lines and drawings to communicate important text information. The visual information is meaningfully distributed over space.

[2]Complexity is defined as a combination of the two dimensions *element interactivity* (number of elements that have to be held in working memory simultaneously) and *incoherence* (number of inferences that have to be drawn based on the presented elements and prior knowledge).

Table 1

Reviewed studies showing beneficial effects from processing pictures *before* text

Study	Prior knowledge	Instructional topic	Complexity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Alvermann (1981)	N/A	Biology: body dehydration	Moderate	Graphic organizer	Pic before text vs. text only (coarse sequence)	Instructor	Text-pic-based recall (immediate & delayed)	Pic before text > text only	d = 0.66 (immediate recall); d = 0.95 (delayed recall)
Baggett (1984)*	Low	Technics: assembly kit	Moderate	Film sequence	Pic presented for 0s, 7s, 14s, or 21s before vs. after text (interleaved sequence)	Instructor	Text-pic-based recall (immediate & delayed)	Pic before text > pic after text	d = 0.14 (immediate recall); d = 0.58 (delayed recall)
Borges & Robins (1980)	Low	Romance	High	Line drawing	30s pic before 30s text vs. 30s text only (coarse sequence)	Instructor	Text-pic-based recall & comprehension	Pic before text > text only	d = 1.67 (recall); d = 1.70 (comprehension)
Bransford & Johnson (1972, Exp. 1)*	Low	Romance	High	Line drawing	30s pic before 30s vs. 30s text before 30s pic (coarse sequence)	Instructor	Text-pic-based recall & comprehension	Pic before text > pic after text	d = 1.99 (recall); d = 2.14 (comprehension)
Dean & Enemoh (1983)*	Low vs. high	Geology: meander formation	Moderate	Photograph	5min pic before 5min text vs. 5min text before 5min pic (coarse sequence)	Instructor	Text-based recall	Pic before text > pic after text	d = 1.62 (low prior knowledge); d = 0.51 (high prior knowledge)
Dean & Kulhavy (1981, Exp. 2)	Low	Narrative: fictitious tribe	Moderate	Schematic map image	6min pic before 15min text vs. 15min text only (coarse sequence)	Instructor	Text-pic-based recall	Pic before text > text only	d = 2.03

Study	Prior know ledge	Instructio-nal topic	Comple-xity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Eitel, Scheiter & Schüler (2013)	Low	Mechanics: pulley system	Moderate	Scienti-fic dia-gram	Pic before text vs. text only (coarse sequence)	Self	Pic- & text-based recall, compre-hension	Pic before text > text only	d = 0.54 (pic-based recall); d = 0.29 (text-based recall); d = 1.33 (comprehension)
Eitel, Scheiter, Schüler, Nyström & Holmqvist (2013, Exp. 2)	Low	Mechanics: pulley system	Moderate	Scienti-fic dia-gram	Self-paced pic before 105s text vs. 105s text only (coarse sequence)	Self & instructor	Pic- & text-based recall, compre-hension	Pic before text > text only	d = 1.44 (pic-based recall); d = 0.84 (text-based recall); d = 1.12 (comprehension)
McCrudden, Magliano & Schraw (2011, Exp. 1)	Low	Medicine: kidney stones development in space	Moderate	Causal diagram	4min pic before 4min text vs. 4min text only (coarse sequence)	Instructor	Text- & pic-based recall	Pic before text > text only	d = 1.28 (text-based recall); d = -0.31 (pic-based recall)
McCrudden, Magliano & Schraw (2011, Exp. 2)	Low	Medicine: kidney stones development in space	Moderate	Causal diagram	4min pic before 4min text vs. 4min text twice (coarse sequence)	Instructor	Text- & pic-based recall	Pic before text > text twice	d = 0.68 (text-based recall); d = -0.89 (pic-based recall)
Peverly (1981)*	Low to mode rate	Social & natural sciences	Moderate	Line drawing s	2min pic after 2min text vs. 2min pic before 2min text (interleaved sequence)	Instructor	Text-based recall	Text presented last > Pic presented last	N/A
Robinson, Corliss, Bush, Bera & Tomberlin (2003, Exp. 3)*	Mode -rate	Psychology: abnormal behavior	Moderate	Anima-ted graphic orga-nizer	Pic before text vs. text before picture (coarse sequence)	Self	Text-pic-based recall & transfer	Pic before text > pic after text	d = 0.78 (verbal recall); d = -0.06 (transfer)

Study	Prior know ledge	Instructio-nal topic	Comple-xity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Salmerón, Baccino, Cañas, Madrid & Fajardo (2009, Exp.1)	Low to high	History: Renaissance	Moderate	Graphic orga-nizer	Text-pic sequence identified via eye movements (interleaved sequence)	Self	Text-based recall & comprehension	Pic after text related to worse comprehension	N/A
Salmerón, Baccino, Cañas, Madrid & Fajardo (2009, Exp.2)	N/A	Biology: biodiversity	Moderate	Graphic orga-nizer	Text-pic sequence identified via eye movements (interleaved sequence)	Self	Text-based recall & comprehension	Pic before text related to better recall	N/A
Simmons, Griffin & Kameenui (1988)*	Low to mode-rate	Physics & chemistry: how matter moves	Moderate	Graphic orga-nizer	Pic before vs. after text (interleaved sequence)	Instructor	Text-based recall (immediate & delayed)	Pic before text > pic after text	d = 0.70 (immediate recall); d = 1.19 (delayed recall)
Snouffer & Thistlethwaite (1980)	N/A	History & physical science	Moderate	Graphic orga-nizer	10min pic, then 15min text vs. 15min text only (coarse sequence)	Instructor	Text-pic-based recall & comprehension	Pic before text > text only	N/A
Stalbovs, Eitel & Scheiter (2013)	Mode-rate	Biology: cell repro-duction	Moderate	Scienti-fic dia-gram	Text-pic sequence identified via eye movements (interleaved sequence)	Self	Text-pic-based recall & comprehension	Early attention to pic related to better recall & comprehension	N/A
Ullrich (2011, Exp. 1)*	Mode-rate	Geology: plate tectonics & volcano activity	Moderate	Scienti-fic dia-gram	Pic before text vs. pic after text (interleaved sequence)	Self	Text-pic-based recall & recognition, comprehension	Pic before text > pic after text	d = 0.46 (recall); d = 0.02 (recognition); d = 0.60 (comprehension)

Study	Prior know ledge	Instructio-nal topic	Comple-xity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Verdi, Johnson, Stock, Kulhavy & Whitman (1997, Exp. 1)*	Low	Narrative: fictional events	Moderate	Schema-tic map image	4.5min pic before 4.5min text vs. 4.5min text before 4.5min pic (coarse sequence)	Instructor	Text- & text-pic-based recall, text-pic integration	Pic before text > pic after text	d = 0.55 (text-based recall); d = 0.58 (text-pic-based recall); d = 0.61 (text-pic integration)
Verdi, Johnson, Stock, Kulhavy & Whitman (1997, Exp. 2)*	Low	Biology: animal cell	Moderate	Scienti-fic dia-gram	4.5min pic before 4.5min text vs. 4.5min text before 4.5min pic (coarse sequence)	Instructor	Text-pic-based recall & text-pic integration	Pic before text > pic after text	d = 0.46 (text-pic-based recall); d = 0.33 (text-picture integration)
Verdi, Kulhavy, Stock, Rittschof & Johnson (1996, Exp. 1)*	N/A	Biology: animal cell	Moderate	Scienti-fic dia-gram	4.5min pic before 4.5min text vs. 4.5min text before 4.5min pic (coarse sequence)	Instructor	Text-pic-based recall & text-pic integration	Pic before text > pic after text	d = 0.56 (text-pic-based recall); d = 0.46 (text-pic integration)
Verdi, Kulhavy, Stock, Rittschof & Johnson (1996, Exp. 2)*	N/A	Biology: flower parts	Moderate	Scienti-fic dia-gram	4.5min pic before 4.5min text vs. 4.5min text before 4.5min pic (coarse sequence)	Instructor	Text-pic-based recall & text-pic integration	Pic before text > pic after text	d = 0.61 (text-pic-based recall); d = 0.46 (text-pic integration)

Notes: Studies that are marked with an asterisk contain a direct comparison of picture before text vs. picture after text. “Picture” is abbreviated by “Pic” within this table.

[a] Complexity of materials means that the materials used in the reviewed studies were roughly categorized into low, moderately, and highly complex depending on their *element interactivity* (number of elements that have to be held in working memory simultaneously; Sweller, van Merriënboer, & Paas, 1998) and *incoherence* (number of inferences that have to be drawn based on the presented elements and prior knowledge).

[b] Regarding the type of sequence, it is to note that “coarse sequence” means that all picture information is presented before or after all text information; “interleaved sequence” means that there is more than one text-picture cycle, even if the different text-picture cycles are not related to each other content-wise.

[c] “Text-pic-based recall” means that the assessment was based on information from both text and picture. Thus, recall was both text- and picture-based.

[d] Main findings and effect sizes (Cohen’s d) are reported only regarding effects of the sequence of presenting text and pictures: d = 0.2 (small effect), d = 0.5 (medium effect), d = 0.8 (large effect). They were estimated based on the data reported in the studies.

Table 2

Reviewed studies showing beneficial effects from processing pictures *after* text

Study	Prior know ledge	Instructional topic	Complexity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Canham & Hegarty (2010, Exp. 1)*	Low	Meteoro-logical principles	Moderate to high	Weather maps	Pic after text vs. pic before text (interleaved sequence)	Self	Comprehension	Pic after text > pic before text	d = 2.17
Canham & Hegarty (2010, Exp. 2)*	Low	Meteoro-logical principles	Moderate to high	Weather maps	Pictures after text vs. pictures before text (interleaved sequence)	Self	Comprehension	Pic after text > pic before text	d = 2.24
Huff & Schwan (2008, Exp. 1)	Low	Object tracking	Low	Short video clips	8s pic before text vs. 8s pic only (interleaved sequence)	Instructor	Pic-based recognition	Pic only > pic before text	N/A
Huff & Schwan (2008, Exp. 2)	Low	Object tracking	Low	Short video clips	8s pic before text vs. 8s pic only (interleaved sequence)	Instructor	Pic-based recognition	Pic after text > pic only	N/A
Huff & Schwan (2012)	Low	Nautical knots	Moderate to high	Short film clips	13-25s pic after 16-31s text vs. 13-25s pic only (interleaved sequence)	Self & instructor	Performance in tying knots (number of attempts & memory for correct attempts)	Pic after text > pic only	d = -1.17 (number of attempts); d = 0.68 (memory for correct attempts)
Kauffman & Kiewra (2010, Exp. 1)	Low	Biology: wildcats	Moderate	Graphic organizer (matrix)	15min pic after 13min text vs. 15min text after 13min text (coarse sequence)	Instructor	Text-pic-based recall of facts & relations	Pic after text > text only	d = 1.82 (facts); d = 2.45 (relations)

Study	Prior know ledge	Instructio-nal topic	Complexity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Kiewra, Kauffman, Robinson, Dubois & Staley (1999, Exp. 2)	N/A	Biology: social groupings of fish	Moderate	Graphic organizer (matrix)	5, 10 or 15min pic after text vs. 5, 10 or 15min text after text (coarse sequence)	Instructor	Text-pic-based recall of relations	Pic after text > text only	N/A
Kiewra, Kauffman, Robinson, Dubois & Staley (1999, Exp. 3)	N/A	Biology: wildcats	Moderate	Graphic organizer (matrix)	10 or 20min pic after 13min text vs. 10 or 20min text after 13min text (coarse sequence)	Instructor	Text-pic-based recall of relations	Pic after text > text only	N/A
McCrudden, Schraw & Lehman (2009, Exp. 1)	Low	Physics: how airplanes achieve lift	Moderate	Causal diagram	4min pic after 4min text vs. 4min text twice (coarse sequence)	Instructor	Recall of causal sequence; transfer; comprehension	Pic after text > text only	d = 1.13 (recall of causal sequence); d = 0.59 (transfer); d = 0.56 (compre-hension)
McCrudden, Schraw & Lehman (2009, Exp. 2)	Very low	Medicine: how kidney develop in space	Moderate	Causal diagram	4min pic after 4min text vs. 4min text twice (coarse sequence)	Instructor	Recall of causal sequence; transfer; comprehension	Pic after text > text only	d = 1.62 (recall of causal sequence); d = 0.71 (transfer); d = 0.76 (compre-hension)
O'Keefe & Solman (1987, Exp. 1)*	Low	Story about "Flat Stanley"	Moderate	Photo-graphs	Pic after text vs. pic before text	Self	Text-pic-based recall	Pic after text = pic before text	d = 0.07
O'Keefe & Solman (1987, Exp. 2)*	Low	Story about "Flat Stanley"	High	Photo-graphs	pictures after text vs. pictures before text	Self	Text- & text-pic-based recall	pictures after text = pictures before text	d = 0.15 (text-based); d = -0.28 (text-pic-based)

Study	Prior know ledge	Instructio-nal topic	Complexity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Robinson, Katayama, Dubois & Devaney (1998, Exp. 1)	N/A	Psychology: abnormal behavior	Moderate	Graphic orga-nizers	Pic after text vs. text only (50min in total; interleaved sequence)	Instructor	Transfer (applications test)	Pic after text = text only	d = -0.11
Robinson, Katayama, Dubois & Devaney (1998, Exp. 2)	N/A	Psychology: sleep disorders	Moderate	Graphic orga-nizers	Pic after text vs. text only (50min in total; interleaved sequence)	Instructor	Transfer (applications test)	Pic after text > text only	d = 0.51
Robinson & Kiewra (1995, Exp. 1)	N/A	Psychology: abnormal behavior	Moderate	Graphic orga-nizers	15min pic after 45min text vs. 15min text after 45min text (coarse sequence)	Instructor	Text-based & text-pic-based recall of facts, recall of relations & transfer	Pic after text > text only (for relations)	d = -0.71 (text-based recall); d = -0.04 (text-pic-based recall); d = 0.39 (relations); d = -0.20 (transfer)
Robinson & Kiewra (1995, Exp. 2)	N/A	Psychology: abnormal behavior	Moderate	Graphic orga-nizers	15+15min pics after 45min text vs. 15+15min text after 45min text (coarse sequence)	Instructor	Text-based & text-pic-based recall of facts, recall of relations & transfer	Pic after text > text only	d = 0.47 (text-based recall); d = 1.32 (text-pic-based recall); d = 1.60 (relations); d = 0.93 (transfer)
Robinson & Schraw (1994, Exp. 1)	N/A	Biology: social groupings of fish	Moderate	Graphic orga-nizers	5min pic after 3min text vs. 5min text after 3min text (coarse sequence)	Instructor	Text-pic-based recognition of relations	Pic after text > text only	d = 1.23
Robinson & Schraw (1994, Exp. 2)	N/A	Biology: social groupings of fish	Moderate	Graphic orga-nizers	1min pic after 3min text vs. 1min text after 3min text (coarse sequence)	Instructor	Text-pic-based recognition of relations	Pic after text > text only	d = 1.46

Study	Prior know ledge	Instructio-nal topic	Complexity of materials[a]	Type of picture	Type and length of sequence[b]	Pacing	Type of assessed knowledge[c]	Main findings	Standardized mean difference (*d*)[d]
Robinson & Schraw (1994, Exp. 3)	N/A	Biology: social groupings of fish	Moderate	Graphic orga-nizers	5min pic after 3min text vs. 5min text after 3min text (coarse sequence)	Instructor	Text-pic-based recognition of relations (after delay of 25min)	Pic after text = text only	d = 0.41
Shaw, Nihalani, Mayrath & Robinson (2012)*	Low	Psychology: sleep disorders	Moderate	Graphic orga-nizers	1.5min pic after 1.5-2min text vs. 1.5min pic before 1.5-2min text (interleaved sequence)	Instructor	Text-pic-based recall & transfer	Pic after text > pic before text	d = 0.11 (recall); d = 0.30 (transfer)

Notes: Studies that are marked with an asterisk contain a direct comparison of picture before text vs. picture after text. "Picture" is abbreviated by "Pic" within this table.

[a] Complexity of materials means that the materials used in the reviewed studies were roughly categorized into low, moderately, and highly complex depending on their *element interactivity* (number of elements that have to be held in working memory simultaneously, Sweller et al., 1998) and *incoherence* (number of inferences that have to be drawn based on the presented elements and prior knowledge).

[b] Regarding the type of sequence, it is to note that "coarse sequence" means that all picture information is presented before or after all text information; "interleaved sequence" means that there is more than one text-picture cycle, even if the different text-picture cycles are not related to each other content-wise.

[c] "Text-pic-based recall" means that the assessment was based on information from both text and picture. Thus, recall was both text- and picture-based.

[d] Main findings and effect sizes (Cohen's d) are reported only regarding effects of the sequence of presenting text and pictures: d = 0.2 (small effect), d = 0.5 (medium effect), d = 0.8 (large effect). They were estimated based on the data reported in the studies.

Table 3

Reviewed studies that directly compare learning outcomes from presenting pictures *before* versus *after* the text as a function of the proposed boundary conditions.

Type of assessed knowledge											
Text-based recall				**Text-pic-based recall**[a]				**Comprehension**			
Study	Relative complexity of materials	Main findings	Mean effect size (d)	Study	Relative complexity of materials	Main findings	Mean effect size (d)	Study	Relative complexity of materials	Main findings	Mean effect size (d)
Dean & Enemoh (1983)	Pic = text	Pic-before > pic-after	d = 1.07	Baggett (1984)	Pic = text	Pic-before > pic-after	d = 0.36	Bransford & Johnson (1972, Exp. 1)	Pic < text	Pic-before > pic-after	d = 2.14
Peverly (1981)	Pic = text	Pic-before > pic-after	N/A	O'Keefe & Solman (1987, Exp. 1)	Pic = text	Pic-before < pic-after	d = -0.07	Ullrich (2011, Exp. 1)	Pic < text	Pic-before > pic-after	d = 0.60
O'Keefe & Solman (1987, Exp. 2)	Pic = text	Pic-before < pic-after	d = -0.15	O'Keefe & Solman (1987, Exp. 2)	Pic = text	Pic-before > pic-after	d = 0.28	**Mean**	**Pic < text**	**Pic-before > pic-after**	**d = 1.37**
Simmons et al. (1988)	Pic = text	Pic-before > pic-after	d = 0.95	Robinson et al. (2003, Exp. 3)	Pic = text	Pic-before > pic-after	d = 0.78				
Verdi et al. (1997, Exp. 1)	Pic = text	Pic-before > pic-after	d = 0.55	Shaw et al. (2012)	Pic = text	Pic-before < pic-after	d = -0.11				
				Verdi et al. (1997, Exp. 1)	Pic = text	Pic-before > pic-after	d = 0.58				
				Verdi et al. (1997, Exp.	Pic = text	Pic-before > pic-after	d = 0.40				

				2)								
				Verdi et al. (1996, Exp. 1)	Pic = text	Pic-before > pic-after	d = 0.51	Canham & Hegarty (2010, Exp. 1)	Pic > text	Pic-before < pic-after	d = 2.17
				Verdi et al. (1996, Exp. 2)	Pic = text	Pic-before > pic-after	d = 0.54	Canham & Hegarty (2010, Exp. 2)	Pic > text	Pic-before < pic-after	d = 2.24
Mean	**Pic = text**	**Pic-before > pic-after**	**d = 0.61**	**Mean**	**Pic = text**	**Pic-before > pic-after**	**d = 0.36**	**Mean**	**Pic > text**	**Pic-before < pic-after**	**d = 2.21**

Note: "Picture" is abbreviated by "Pic" within this table.

[a] "Text-pic-based recall" means that the assessment was based on information from both text and picture. Thus, recall was both text- and picture-based.