4 Constraints on Learning from Expository Science Texts

Jennifer Wiley and Christopher A. Sanchez

A great deal of instruction in science occurs through reading. Even in science classes where hands-on activities, experiments, or problem-based learning approaches are becoming more popular, background information critical for integrating the activity experiences into disciplinary understanding is often presented via text. Although the genre of most instructional text, especially in science, is expository, elementary reading experiences are generally in the narrative genre, with students usually honing their basic reading skills on story-like texts. However, some time around middle school, students start to receive expository texts and are expected to learn from them (Sweet & Snow, 2003). Not surprisingly, students often have difficulty with this task. Even undergraduate students can be remarkably poor at comprehending the information presented in expository text, and have been shown to be especially inept at judging their own level of comprehension, reflecting an inability to engage in accurate monitoring (Wiley, Griffin, & Thiede, 2005). The downstream effects of poor monitoring accuracy are that readers who are studying on their own will fail to engage in re-reading of poorly learned materials, and will fail to comprehend the texts they read.

The purpose of this chapter is to explore two major constraints on learning from expository texts. The first constraint that will be discussed is that the nature of understanding or comprehending the content of an expository text is somewhat different than comprehending a narrative. In short, students need to be instructed in what it means to understand an expository text so that they can direct their attention toward the right level of processing as they read. A second constraint that will be explored builds on another essential difference between narrative and expository science texts: often, scientific texts describe how and why processes or phenomena occur. In these cases, the fundamental understanding of the content is essentially the mental representation of dynamic relations over space and time. Using an individual differences approach, the later part of the chapter examines the role that spatial abilities and working memory capacity play in comprehension and the development of such mental representations. Finally, experimental interventions that address these constraints and improve learning outcomes are discussed.

The Situation Model and Comprehension of Expository Text

The approach to expository text comprehension taken here is called a *situation model approach*, based on the work of Walter Kintsch, who delineated multiple levels of representation as one processes a text (Kintsch, 1998). The first of these levels is the *surface* level, and this level represents the exact words that are encountered. A second level, the *textbase*, represents a refinement of the actual phrasing into gist-level or macro-structural propositions. Again, this representation largely represents the information that was present in the text, with perhaps some summarization, simplification, or generalization. The final level of representation is called the *situation model*, and it is at this level that connections between the ideas in the text and ideas in prior knowledge are created. This level can also be seen as a mental model of the phenomena. This is the level of representation that captures what the text "meant" on a deep level. It includes a causal model of the phenomena, and involves the generation of inferences or connections across concepts to explain how or why phenomena happen. In other words, this representation is what we define as real comprehension of a text (Kintsch, 1998; Wiley & Myers, 2003). This assumption is based on work in the text processing literature showing how the generation and representation of text in terms of a causal model predicts comprehension measures such as reading times, recall, recognition, and inference generation (Graesser & Bertus, 1998; Trabasso & van den Broek, 1985).

A critical difference between narrative text and expository text is in the closeness of representations between the multiple levels. In narrative text, memory for the narrative (what happened to whom, and when) is largely similar to "understanding" the text. Thus, when students read narratives, they can have a reading goal of remembering what happened, and it will serve them well for most tests about the text. This could either be because representations are so similar, as suggested above, or, alternatively, because one level can be easily constructed from the other owing to the highly accessible prior knowledge about people's intentions and motivations that are the causal links required for most narrative texts. On the other hand, the discrepancy between the *textbase* and the *situation model* is much greater for expository texts. Knowing what sentences appeared in the text is not the same as understanding the process or phenomena they describe. The *situation model*, or mental model of the phenomena, is much further removed from memory for exact sentences in expository text. It requires the construction of a causal model during the reading of the text and cannot be easily constructed afterward from memory. If future comprehension tests will tap inference-level understanding, readers need to attend to their ability to make connections and not just memory (neither fluency nor retrieval cues) for text in order to judge their understanding accurately (Griffin, Wiley, & Thiede, 2008).

Constraints Due to Failure to Understand the Nature of Comprehension

The above suggests that one major constraint in the comprehension of expository text is that readers may generally fail to recognize that comprehension requires engaging in explanatory processing. In reading expository text, the reader's goal is to try to understand how and why a phenomenon occurs, and not just to try to remember the isolated ideas that were read (Wiley et al., 2005). One recent study gives direct support for this proposition. In Thiede, Griffin, Wiley, and Anderson (2010), typical college readers and at-risk college readers were given several expository texts to learn from and were asked to judge how well they would do on a test of comprehension for each text after reading. Then, they also took inference-based comprehension tests on each text. From these measures, intra-individual correlations were computed which represented how well students could predict their own actual performance on comprehension tests. The overall correlation was only .21 for typical college readers, (similar to typical levels of accuracy around .27; Thiede, Griffin, Wiley, & Redford, 2009), but the average intra-individual correlation for at-risk readers was only .08.

In addition, we asked these readers to self-report *how* they judged their level of comprehension at the end of the study. These open-ended responses were coded into four categories. The first set included reports of features of the text such as using the difficulty of the vocabulary in the texts or their length (surface cues). The second set included references to characteristics of the reader such as prior knowledge or interest about the topic (reader cues). The third set consisted of comments related to the ability to remember or recall the texts (memory cues). The final set included comments related to the ability to explain the phenomena (comprehension cues). The striking finding was that only 5 percent of readers in a college sample (and no at-risk readers) spontaneously used comprehension cues as a basis for making comprehension judgments. Most readers used surface or memory-based cues. Further, the nature of the cues used made a striking difference in metacomprehension accuracy. Readers who used only surface, reader, or memory cues had intra-individual correlations of around .15, while readers who used only comprehension cues had correlations at a level of .71. This evidence supports the two key assumptions stated above. First, readers of expository texts tend to default to monitoring their comprehension using cues based in the *surface* or *textbase* levels of representation, and this leads to poor ability to judge actual comprehension when tests tap inferences or causal connections among ideas. Second, when readers *do* use their ability to explain a text as the basis for their comprehension judgments, then students do remarkably well at judging their own understanding, presumably because such cues are linked to the quality of the *situation model.*

Finally, we note that at-risk students [college students who were enrolled in remedial reading courses because they had low American College Test (ACT) verbal scores] were even less likely than the typical readers to consider *situation model*-based cues while judging comprehension. In fact, none of the remedial readers did so spontaneously. Given that at-risk readers display similarities to younger readers, the need to support a better understanding of what it means to understand an expository text may be even more acute in these populations.

Improving Understanding of Expository Text Comprehension

Based on these results, it seems critical to explore how to get readers to have a better understanding of what it means to comprehend expository tests. A number of lines of research have been pursued to address this question using several tasks and conditions that direct a reader's attention to the *situation model.* When combined with comprehension tests that tap the creation of a causal model of the text, these interventions have been found to improve predictive accuracy on expository texts. The results of these studies are listed in Table 4.1, along with an average for no-intervention control conditions that replicates the typical finding in the literature that, without support, intra-individual correlations hover at .27.

Table 4.1 Interventions and Metacomprehension Accuracy

Intervention Condition	*Metacomprehension Accuracy (γ)*
No Intervention	
Anderson & Thiede (2008)	.20
Griffin, Wiley, & Thiede (2008) (two conditions)	.25
Thiede & Anderson (2003) (two conditions)	.28
Thiede, Anderson, & Therriault (2003)	.38
Thiede, Dunlosky, Griffin, & Wiley (2005) (seven conditions)	.28
Thiede, Griffin, Wiley, & Anderson (2010)	.21
Average across all control conditions	.27
Delayed Summary Tasks	
Thiede & Anderson (2003) (two conditions)	.61
Thiede, Griffin, Wiley, & Anderson (2010)	.64
Thiede, Griffin, Wiley, & Anderson (2010)*	.48
Anderson & Thiede (2008)	.64
Delayed Keyword Tasks	
Thiede, Anderson, & Therriault (2003)	.70
Thiede, Dunlosky, Griffin, & Wiley (2005) (five conditions)	.55
Self-explanation Tasks	
Griffin, Wiley, & Thiede (2008)	.67
Concept Mapping Tasks	
Thiede, Griffin, Wiley, & Anderson (2010)*	.67

*At risk sample.

The driving theme behind these experiments has been to get readers to have access to the right level of cues when they consider their level of comprehension. The first set of interventions showed that instituting a delay between reading texts and engaging in a generation task that prompted readers to generate keywords (delayed keyword) or summaries (delayed summary) before judging improved readers' predictions of their own level of comprehension. This result was attributed to the rapid decay that occurs in surface memory for text and the more robust nature of the *situation model* (Kintsch, 1998). Thus, when readers were asked to generate keywords or summaries after a delay, this gave them access to cues about the quality of their *situation model* for each text, which in turn improved the quality of their comprehension judgments. These effects have now been replicated across several studies (see Table 4.1).

The next wave of studies directed readers to attend to the *situation model* in a somewhat more direct fashion, by having them engage in specific reading behaviors that highlight the connections that need to be made as the text is read (i.e., self-explanation and concept mapping). Studies in this vein have used explicit reading instructions that make the purpose of reading clear, but also require the reader to engage in additional constructive activities during reading. A popular intervention to improve learning from expository texts is self-explanation (Chi, 2000; McNamara, 2004). Related approaches include asking students to write causal arguments (Wiley & Voss, 1999) or to engage in how-and-why question-asking behaviors during reading (Graesser & Bertus, 1998). Because of the emphasis of all of these interventions in asking students to engage in causal or explanatory reasoning about the phenomena as they read, these approaches should also improve students' awareness of the quality of their *situation models* and thus their comprehension monitoring ability. To test this, Griffin et al. (2008) gave students clear instructions that the tests they would be taking would be based on connections and inferences about the content of the texts. In addition, some students were prompted to self-explain during a second reading of the texts. The instructions used for this explanation instruction were based on Chi (2000), and prompted students to attempt to connect ideas and think about how and why questions. Under these conditions, students' accuracy at judging their own comprehension improved, with gamma correlations averaging at .67.

Another similar approach, particularly appropriate for at-risk or younger readers, is concept mapping. A concept map is a graphic representation of the underlying structure of the meaning of a text. Constructing concept maps can be an effective organizational strategy, which may help readers formulate the connections among concepts in a text (Weinstein & Mayer, 1986). As noted above, some previous results suggested that metacomprehension accuracy for many at-risk readers is compromised by the use of inappropriate cues based on surface features of a text (Thiede et al., 2010). Thus, supporting the use of diagnostic monitoring may be especially critical for at-risk readers. Further, because self-explanation is a process that may add extra demands on a reader, concept mapping was chosen as an alternate intervention as it has been suggested that such an approach may be particularly helpful and appropriate for low-ability readers (Nesbit & Adesope, 2006; Stensvold & Wilson, 1990). Constructing concept maps is a generative activity that shares many similarities with argumentation and explanation tasks, but because it employs the construction of external, visual representations while readers have access to the texts, it may place fewer demands on the reader than other explanation tasks. Instructing at-risk readers to construct a concept map of a text during reading should help them identify important connections and construct a *situation model* for a text. In turn, it should also increase the salience of that *situation model*-level representation, which they may then use to better monitor their comprehension of a text. Consistent with these predictions, when at-risk readers generated concept maps before judging their comprehension, gammas reached a remarkable .67.

Conclusions about Metacomprehension Constraints

Taken together, the above results are consistent with the main conjecture that readers do not really understand what it means to comprehend expository text. The studies reported above show that readers default to *surface* or *textbase* cues such as memory for the text when they attempt to monitor their understanding. Unfortunately, these cues will not be diagnostic when the goal for reading is to understand how or why phenomena occur. Additionally, monitoring comprehension on this level will result in ineffective studying behaviors and in poor learning from expository science texts.

Thus, one main obstacle that ultimately impedes comprehension and learning from expository text is a failure to understand the explanatory nature of the processing that is required to develop coherent, causal mental models of scientific phenomena. To this point, this chapter has focused primarily on how such a misconception of comprehension relates to poor monitoring accuracy, and has described several interventions that have improved accuracy specifically by putting readers in contexts that support better access to *situation model* cues for comprehension.

The ultimate goal, of course, is improved comprehension, not just improved metacomprehension for expository text. The rationale for the emphasis on metacomprehension is that exploring the conditions that support accurate comprehension monitoring should ultimately lead to better comprehension by supporting more effective study behaviors. Although there are few studies that have directly tested these downstream effects, there are some promising results already in the literature. For example, Thiede, Anderson, and Therriault (2003) gave students an opportunity to restudy whichever texts they wished after a first set of comprehension tests and before a final set. Students assigned to the delayed keyword generation condition, who on average had more accurate monitoring, also made more effective studying decisions. They chose to restudy particularly those texts that they performed poorest on, and were less likely to restudy texts that they did well on. The average first test performance for the texts they chose to restudy was .27, whereas the average first test performance for the non-selected texts was .78. This difference reveals a clear ability to discriminate the texts that were understood from those that were not. In a no-keyword generation, no-delay condition, however, students showed less of a difference in first test performance on restudied (.44) versus not selected texts (.55). Importantly, final test performance was related to differential restudy choices. Students in the delayed keyword condition improved more on final tests, indicating that the intervention, which supported better monitoring accuracy, ultimately led to better comprehension from expository text as well. Further, for other interventions such as concept mapping and self-explanation, positive effects on both metacomprehension and comprehension may be seen more immediately (Chi, 2000; McNamara, 2004; Thiede et al., 2010).

Individual Differences as Constraints on the Construction of Mental Models

Several studies have shown that readers routinely fail to spontaneously generate causal inferences or a coherent causal model while reading scientific texts. The second half of this chapter takes an individual differences approach to explore two possible sources of these failures: working memory capacity (WMC) and spatial ability.

Because the topics of such texts are usually less familiar to readers, and generating inferences requires keeping multiple ideas active at once, expository text comprehension is thought to place a load on working memory (Linderholm & van den Broek, 2002). This may be especially true for low-knowledge readers (McNamara, 2004), although all readers may need explicit support or prompting to compute inferences from expository text (cf. Singer & O'Connell, 2003). For example, Wiley and Myers (2003) have shown that causal inferences may be generated from expository text only when all necessary information is available and adjacent in the text. Putting even a sentence between a key premise and conclusion can prevent readers from generating causal inferences. Findings such as this suggest that individual differences in WMC may play an important role in who develops coherent *situation models* while reading expository texts. Such effects are largely consistent with a resource allocation view of WMC, such that readers with a higher capacity may be able to engage in more complex comprehension processes (Griffin et al., 2008).

A second way in which WMC can relate to comprehension is more consistent with a controlled attention view (Kane, Bleckley, Conway, & Engle, 2001), which posits that some individuals are better able to focus their attention than others. This approach instead suggests that high-WMC readers may be better able to focus on relevant information as they read. A demonstration consistent with this prediction is provided by a set of studies done by Sanchez and Wiley (2006). These studies demonstrated that when participants read expository text that was illustrated with irrelevant images (e.g., "seductive details") (Harp & Mayer, 1997)—images that are tangentially related to the topic, but not conceptually related to the deeper relationships contained within the text—low-WMC readers were more likely to be distracted by these irrelevant details, as seen in eye-tracking measures. This distraction, in turn, led to poorer performance on comprehension measures for low-WMC individuals. High-WMC readers were better able to focus on the relevant information, which led to the construction of higher-quality situation models and better comprehension. The results of this study suggest that, especially in cases where there is a need to ignore irrelevant information, WMC will be predictive of learning and individual differences in WMC may be seen as a constraint on expository text comprehension.

Another class of individual differences that may be particularly important for learning from expository science text is spatial ability. It is widely assumed that learning in the physical sciences requires dealing with and understanding primarily spatial phenomena (Gobert, 2000; Kozma & Russell, 2005). To the extent that comprehension of scientific phenomena requires the construction of a mental model of a process or system, then the ability to visualize or simulate the operation of that phenomenon, and integrate important temporal and spatial interactions, may be critical to the formation of a "runnable" mental model that can be used to recapitulate the physical process (Gentner & Stevens, 1983; Hegarty & Steinhoff, 1997). For example, for learners to successfully understand the geological phenomena of plate tectonics, they must identify not only the relevant conceptual units that are spatial in nature (e.g., plates, magma, plate boundaries), but also how these units interact and change over time, which must then be represented within their own internal, runnable mental model. How closely the internal representation matches the actual physical phenomenon provides a rough proxy of how well the material has been understood. Obviously, the more accurately the learner can mentally construct a representation of the phenomenon, the better the learner understands, or is capable of demonstrating understanding of the topic. Because of the inherently dynamic, spatial nature of the representation for many scientific processes, spatial abilities represent one potential constraint on comprehension.

The most commonly used measures of spatial ability are paper-and-pencil tasks taken directly from the French Reference Kit (French, Ekstrom, & Price, 1963). These tasks fall into two main classes: tasks that tap the ability to rotate objects in space (e.g., block and figure rotation tasks) and tasks that tap the ability to visualize manipulations to objects (e.g., paper folding) or the ability to reconceptualize an existing spatial representation into a revised new whole (Carroll, 1993; Pellegrino & Hunt, 1991). These two subtypes of object manipulation ability are usually highly correlated and difficult to differentiate (Stumpf & Eliot, 1995). Individual differences on measures derived from object manipulation tasks have been shown to predict performance on tasks that explicitly require visuospatial information processing, such as mechanistic reasoning tasks using drawings of physical objects such as gears or pulleys (Hegarty & Steinhoff, 1997). Individual differences in these abilities have also been found to predict the comprehension of narrative text where readers follow the actions of a character in physical space (Fincher-Kiefer & D'Agostino, 2004). Although many have posited that visuospatial abilities are key to science understanding, there is no direct evidence of a relationship with comprehension from expository text.

Similarly, some of the greatest discoveries in science, such as the double helix, the benzene ring, and theories of plate tectonics, have been attributed to the ability of great scientists to think spatially (National Research Council, 2006). However, the evidence for the link between spatial ability and performance in science is largely anecdotal and correlational, with findings that scientists and students in advanced science courses have higher spatial abilities than the general population, and some positive relations between exam performance and spatial ability (Black, 2005; Wu & Shah, 2004).

However, these studies cannot be taken as direct evidence that spatial ability leads to better comprehension in science unless one can rule out general individual differences in ability or prior knowledge as factors. Further, there are also examples of spatial ability failing to predict learning of science content, including topics in both biology (Koroghlanian & Klein, 2004) and physics (ChanLin, 2000). Although these studies span several different scientific domains and no doubt test learning in different ways, it is disconcerting that no consistent relationship has emerged between measures of spatial ability and measures of science learning, given the very visual and spatial nature of many topics within these domains (Gobert, 2000; Kozma & Russell, 2005). This recognition prompts the question as to whether or not the standard measures of spatial ability based in object-manipulation tasks are the most relevant for the kinds of representations that are needed for comprehension of some scientific topics. It is possible that the visualization and manipulation abilities captured by these tasks, which have been referred to as static spatial abilities or SSAs (Pellegrino & Hunt, 1991), are not necessarily the most critical for complex science reasoning in these domains.

Given the dynamic nature of many scientific phenomena, perhaps a spatial ability which better captures this dynamic characteristic, or what has been referred to as dynamic spatial ability (DSA), might be a better predictor of comprehension. Tests of dynamic spatial abilities attempt to capture this dynamic nature and represent the capacity to integrate spatial information across multiple instances and over time (Hunt, Pellegrino, Frick, Farr, & Alderton, 1988; Pellegrino & Hunt, 1991). While these abilities are modestly correlated, factor-analytic work has borne out their identity as separable constructs (Contreras, Colom, Hernandez, & Santacreu, 2003; Hunt et al., 1988).

The connection between DSA and science learning is slightly more tenuous at this point, as to date there have been only a handful of studies across any domain that have examined this ability, and most efforts have been dedicated to the establishment of the validity of this concept as distinct from SSA. Dynamic spatial ability has been shown to predict learning of an air-traffic control task (D'Oliveira, 2004), and has also been found to contribute unique variance over and above verbal intelligence on performance intelligence quotient (IQ) tests (Jackson, III, Vernon, & Jackson, 1993). However, given that most physical science topics involve understanding dynamic changes over time, DSA may be critical for the formation of dynamic scientific mental models.

To continue with our example from the earth sciences, when asked to understand the movements of plate tectonics and how this system produces volcanoes and earthquakes, learners need to understand not only that these plates exist, but also the underlying movements of magma as it circulates around the earth's core and causes these plate movements. In essence, the constant change and movement of magma, which causes the construction/destruction of the tectonic plates over time, is critical for understanding how, why, and where volcanoes and earthquakes occur. Further, maintaining spatial information about relations between isolated units or pairs of concepts (i.e., the intersection of two plates) can provide only a partial understanding of the entire process. In order to fully develop a deep understanding of the entire system, multiple relationships must be encoded and integrated within this time-based process, suggesting that DSA should be critical.

This indeed was found to be the case in a recent study by Sanchez and Wiley (2007). In one condition of this study, college students were asked to read a non-illustrated lesson about plate tectonics. Students were asked to read for the purposes of understanding what causes volcanic eruptions. Then, learning was assessed by having students write an essay on "What caused Mt. St. Helen's to erupt?" This essay was scored for the presence of eight main concepts from an a priori causal model of the phenomena. Consistent with the hypothesis that DSA constrains science text comprehension, the best predictor of overall learning from this lesson was DSA, with correlations of around .30 (Table 4.2). Specifically, it was found that in situations where the demands on learners to mentally animate the information to achieve understanding were highest (i.e., non-illustrated condition), DSA was especially important for generating and running this mental model. However, the presence of dynamic, external representations (e.g., animations) attenuated this relationship, while still producing high overall learning. This suggests that understanding dynamic relationships between conceptual units is necessary to form an accurate representation of the content area, and

Table 4.2 Essay Performance and Correlations between WMC, DSA, and SSA and Essay Performance by Condition in Sanchez and Wiley (2007)

	Non-illustrated (n = 64)	*Static Illustrations* (n = 69)	*Animations* (n = 63)
Correct concepts, mean (SD)	2.08 (1.23)	2.03 (1.45)	2.54 (1.35)
Correlations			
WMC and learning	.03	.04	.24**
DSA and learning	.30*	.26*	.04
SSA and learning	–.01	.20**	.09

* $p < .05$, ** $p < .10$.

when the learner is required to generate such relationships on his or her own, DSA is important. These results suggest that building a runnable mental model or simulation of the to-be-learned phenomena is an important part of the comprehension process, and that readers with higher DSA have an easier time translating the ideas from the text into a mental simulation or internal visualization of the phenomenon.

Supporting the Creation of Runnable Mental Models in All Readers

The next question was whether the construction of such models could be supported in students with low DSA. One obvious intervention that has been used a great deal in the literature is providing readers with illustrations to support learning. However, the impact of adding visual material to text has had mixed reviews in terms of its effectiveness. On one side, the addition of static illustrations such as charts and diagrams has been used effectively to enhance science learning (see Ainsworth & Th Loizou, 2003; Mayer, Hegarty, Mayer, & Campbell, 2005). Similarly, the addition of dynamic animations or video has also been shown to sometimes provide facilitation of learning in science (see Schnotz, 2005).

However, there is also evidence that adding visual media does not always facilitate learning, and there seems to be a great deal of specificity to the contexts in which adding visualizations will be optimal. For example, in some cases, static images produce better learning than animations (Mayer et al., 2005). In other cases, static illustrations and animations have failed to enhance learning on science topics, have sometimes led to worse performance, and often seem to interact with student abilities or explicit instructional support (Geiger & Litwiller, 2005; Sanchez & Wiley, 2006; Schnotz, 2005). In a review of studies on animation, Tversky, Morrison, and Betrancourt (2002) concluded that the evidence for beneficial aspects of animation on learning were at best mixed. Similar conclusions have been reached about static graphics (Mandl & Levin, 1989).

To date, there is not a definitive explanation of when such visualizations should improve learning. Based on their review of the mixed findings, Tversky et al. (2002) proposed two general principles: *congruence* and *apprehension*. These principles simply require that information in the visualization should match the nature of the desired internal model, and should also be able to be easily apprehended and perceived (Lowe, 2004; Schnotz, 2005). Consistent with these recommendations, readers in two other conditions of Sanchez and Wiley (2007) were given either static or animated visualizations that illustrated important conceptual ideas from the text in fairly simple schematic diagrams. The static versions gave readers a visual representation of the important entities for understanding plate tectonics such as plates, layers of earth, and faults. Notably, only the animated version presented the information in a way in which essential dynamic relations and interactions between entities were obvious.

Given the discussion above, it was anticipated that providing illustrations should improve learning, but also that the type of visualization might impact learning of this science topic. In other words, animations which capture the dynamic nature of the phenomenon might lead to the best overall comprehension, while static visualizations, as they at least represent the relevant structure of the event, might produce better comprehension than the non-visual text. Further, these adjuncts may be especially important for low spatial ability learners.

The results of Sanchez and Wiley (2007) as shown in Table 4.2 demonstrate that the presence of conceptually relevant animations improved learning overall (mean = 2.54, SD = 1.35) compared with both the no-illustrations (mean = 2.08, SD = 1.23) and static illustrations (mean = 2.03, SD = 1.45) conditions and, importantly, animations also eliminated the relationship between DSA and learning. In the animated condition, only WMC was a predictor of learning. However, in the static illustrations condition, overall learning was not improved significantly over the no-illustrations condition and DSA was still a significant predictor of learning.

Results from across the three conditions in this experiment provide evidence that individual differences in spatial ability, and also WMC to some extent, impose constraints on the comprehension of expository science text. First, the inclusion of animations alongside the text led to the best performance overall, and this can be taken as evidence that representing the dynamic properties of process described in the text supported better comprehension of a topic that is an inherently dynamic phenomenon. Consistent with this interpretation, the relation between DSA and learning was eliminated in this condition. Interestingly, the addition of static illustrations did not enhance learning above the non-illustrated condition, and the relationship between DSA and learning was still strong in this situation. It appears that learners were able to represent the relevant conceptual entities from the text even without visual support. This produced similar learning effects in the no-illustrations and static illustrations conditions. Given that both of these non-dynamic conditions still placed heavy requirements on DSA, for at least this topic, a key difficulty was mentally animating the dynamic changes that occur across time.

In sum, it appears that DSA is indeed a constraint on understanding in science and is connected to how well individuals can construct mental models of a dynamic science topic. The presence of relevant conceptual animations can improve understanding and support visualization skills among readers with low DSA. However, although individual differences in DSA were eliminated in the animation condition, the role of individual differences in WMC became more apparent, consistent with previous research showing that under some conditions WMC will also act as a constraint on comprehension processes (Geiger & Litwiller, 2005). In this case, as in previous research, WMC may help readers to learn effectively from animations.

Multiple Routes to Improving Student Learning from Science Texts

The approaches discussed above have delineated two major classes of difficulties that may prevent readers from developing adequate understanding from scientific texts. The first is that readers may generally misunderstand that their goal for reading expository science texts is to develop an explanatory model of how or why phenomena happen, and that they should monitor their comprehension using that standard. The second is that some readers will have difficulty creating mental models when reading about science topics, particularly for texts about dynamic physical phenomena, due to a lack of dynamic spatial ability. However, interventions that clarify the desired nature of processing, and that support visualization and the creation of runnable mental models, are promising directions for the future which may ultimately help all students become more effective learners of science. It is also an interesting question for future research as to how the early support of expository comprehension and spatial skills in elementary school students would affect later learning in science.

Acknowledgments

A portion of the research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305H030170 and R305B070018 to Keith Thiede, Thomas D. Griffin, and Jennifer Wiley. Additional portions of this research were supported by the APA Dissertation Research Award to the second author. The opinions expressed are those of the authors and do not represent views of these institutions.

References

Ainsworth, S., & Th Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science*, 27, 669–681.

Anderson, M. C., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, 128, 110–118.

Black, A. A. (2005). Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education*, 53(4), 402–414.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge: Cambridge University Press.

ChanLin, L. J. (2000). Attributes of animation for learning scientific knowledge. *Journal of Instructional Psychology*, 27(4), 228–238.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5). Mahwah, NJ: LEA. pp. 161–238.

Contreras, M. J., Colom, R., Hernandez, J. M., & Santacreu, J. (2003). Is static spatial performance distinguishable from dynamic spatial performance? A latent-variable analysis. *Journal of General Psychology*, 130, 277–288.

D'Oliveira, T. C. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *International Journal of Aviation Psychology*, 14, 19–38.

Fincher-Kiefer, R., & D'Agostino, P. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes*, 37, 205–224.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton: Educational Testing Service.

Geiger, J. F., & Litwiller, R. M. (2005). Spatial working memory and gender differences in science. *Journal of Instructional Psychology*, 32, 49–58.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Hillsdale, NJ: LEA.

Gobert, J. D. (2000). A typology of causal models for plate tectonics. *International Journal of Science Education*, 22(9), 937–977.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, 2, 247–269.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation. *Memory & Cognition*, 36, 93–103.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations. *Journal of Educational Psychology*, 89, 92–102.

Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as memory in mechanical reasoning. *Learning and Individual Differences*, 9, 19–44.

Hunt, E., Pellegrino, J. W., Frick, R. W., Farr, S. A., & Alderton, D. L. (1988). The ability to reason about movement in the visual field. *Intelligence*, 12, 77–100.

Jackson, D. N., III, Vernon, P. A., & Jackson, D. N. (1993). Dynamic spatial performance and general intelligence. *Intelligence*, 17(4), 451–460.

Kane, M. J., Bleckley, K. M., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130,169–183.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Koroghlanian, C., & Klein, J. D. (2004). The effect of audio and animation in multimedia instruction. *Journal of Educational Multimedia and Hypermedia*, 13, 23–45.

Kozma, R., & Russell, J. (2005). Multimedia learning of chemistry. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning*. New York: Cambridge University Press. pp. 409–428.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94, 778–84.

Lowe, R. K. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction*, 14(3), 257–274.

Mandl, H., & Levin, J. R. (Eds.) (1989). *Knowledge acquisition from text and pictures*. Amsterdam: North-Holland.

Mayer, R. E., Hegarty, M., Mayer, S. Y., & Campbell, J. (2005). When passive media promote active learning. *Journal of Experimental Psychology: Applied*, 11, 256–265.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38, 1–30.

National Research Council. (2006). *Learning to think spatially*. Washington, DC: The National Academies Press.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76, 413–448.

Pellegrino, J. W., & Hunt, E. B. (1991). Cognitive models for understanding and assessing spatial abilities. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement*. Mahwah, NJ: LEA. pp. 203–225.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34, 344–355.

Sanchez, C. A., & Wiley, J. (2007). Spatial abilities and learning complex science topics. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, Nashville, TN.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning*, New York: Cambridge University Press. pp. 49–70.

Singer, M., & O'Connell, G. (2003). Robust inference processes in expository text comprehension. *European Journal of Cognitive Psychology*, 15, 607–631.

Stensvold, M. S., & Wilson, J. T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. *Science Education*, 74, 473–480.

Stumpf, H., & Eliot, J. (1995). Gender-related differences in spatial ability and the *k* factor of general spatial ability in a population of academically talented students. *Personality and Individual Differences*, 19, 33–45.

Sweet, A. P., & Snow, C. E. (Eds.) (2003). *Rethinking reading comprehension*. New York: Guilford Press.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experiment Psychology: Learning, Memory & Cognition*, 31, 1267–1280.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, 47, 331–362.

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education*. New York: Routledge.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24, 612–630.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human–Computer Studies*, 57(4), 247–262.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook on research in teaching* (3rd edn). New York: Macmillan. pp. 315–327.

Wiley, J., Griffin, T., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, 132, 408–428.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, 36, 109–129.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources. *Journal of Educational Psychology*, 91, 301–311.

Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88(3), 465–492.