Learning and Instruction 31 (2014) 33–45

Contents lists available at ScienceDirect

Learning and Instruction

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

The role of dynamic spatial ability in geoscience text comprehension

Christopher A. Sanchez [a,*], Jennifer Wiley [b]

[a] School of Psychological Science, Oregon State University, 2950 SW Jefferson Way, Corvallis, OR 97331, USA
[b] University of Illinois at Chicago, USA

ARTICLE INFO

Article history:
Received 15 April 2013
Received in revised form
27 December 2013
Accepted 28 December 2013

Keywords:
Spatial ability
Geoscience learning
Individual differences

ABSTRACT

The current experiment investigated the effects of a dynamic spatial ability on comprehension of a geoscience text on plate tectonics and the causes of volcanic activity. 162 undergraduates (54% female) from a large public university who had little prior knowledge of this science content area were asked to learn about plate tectonics. Measures of spatial ability and working memory capacity were used to predict comprehension from a text that contained either no images, static images, or animated versions of the static images. Only the dynamic spatial ability measure interacted with the type of illustrations contained in the text, and was shown to be especially relevant for comprehension when readers did not receive animations. These results demonstrate a novel influence of individual differences in dynamic spatial ability on comprehension of text describing dynamic spatial phenomena.

1. Introduction

Learning from expository science texts, like all comprehension processes, requires that readers go beyond a verbatim memory trace to construct an understanding of what the text is really about (Kintsch, 1994). Particularly when texts describe how or why scientific processes or phenomena occur, the goal for comprehension can be seen as the development of a situation model, causal model, or runnable mental model of the content (Gentner & Stevens, 1983; Graesser & Bertus, 1998; Johnson-Laird, 1983; Kintsch, 1998; Trabasso & van den Broek, 1985; Wiley & Myers, 2003; Zwaan & Radvansky, 1998). Further, many topics in science involve understanding phenomena with elements that move and interact across time and space. Understanding how these elements move, interact, and change over time may be more amenable to spatially-based mental representations, and may not be easily translated into verbal propositions (Friedman & Miyake, 2000; Hegarty, 1992; Hegarty, Canham, & Fabrikant, 2010; Rinck, 2005). For example, for learners to successfully understand the geological phenomena of plate tectonics, they must first identify the relevant conceptual units (e.g., plates, magma, plate boundaries, etc.), which itself is often difficult for learners (Kortz et al., 2011). Learners must then consider how these conceptual units interact and change over time, and thus they need to represent this dynamic spatial information in their own mental models of the tectonic system (Gentner & Stevens, 1983; Hegarty, 1992). This suggests that one constraint on science text comprehension may be individual differences in visuospatial skills, because they may be especially relevant for the creation of spatial situation models (Friedman & Miyake, 2000; Haenggi, Kintsch, & Gernsbacher, 1995).

1.1. Individual differences in visuospatial ability

A long tradition of factor-analytic research has provided strong evidence that 'visuospatial ability' can not only be differentiated from general intelligence and verbal ability, but also represents a complex of multiple, distinguishable, visuospatial faculties (see Hegarty & Waller, 2005, for a comprehensive review). One traditional taxonomic distinction differentiates visuospatial abilities (and the tests that measure them) into two main classes: those that tap the ability to rotate objects in space (e.g., block and figure rotation tasks) and those that tap the ability to visualize or reconceptualize an existing spatial representation into a revised new whole (e.g., paper folding, form board; Carroll, 1993; Cooper, 1975; Cooper & Shepard, 1973; Mumaw, Pellegrino, Kail, & Carter, 1984; Pellegrino & Hunt, 1991). These two sub-types of spatial ability, although generally recognized as separate abilities (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), are usually highly correlated and can sometimes be difficult to differentiate (Carroll, 1993; Just & Carpenter, 1985; Stumpf & Eliot, 1995). Hunt, Pellegrino, and their colleagues also classify both of these types of abilities within a single category that they refer to as static spatial ability (Fischer, Hickey, Pellegrino, & Law, 1994; Hunt, Pellegrino, Frick, Farr, & Alderton, 1988; Law, Pellegrino, & Hunt, 1993). They emphasize that these measures deal with transformation of a single

* Corresponding author. Fax: +1 541 737 3547.
E-mail address: christopher.sanchez@oregonstate.edu (C.A. Sanchez).

0959-4752/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2013.12.007

object and not the interaction of multiple objects over space and time. Another recent theoretical framework classifies both of these tasks into what is known as the *intrinsic-dynamic* category, which represents the transformation of spatial coding within objects relative to starting and end points (Newcombe & Shipley, 2012; Uttal et al., 2013). Although this new label includes the word 'dynamic', again the emphasis here is on within-object manipulations.

Importantly, individual differences in performance on these types of within-object-manipulation visuospatial tasks (both rotation and visualization) have been shown to explain unique variance in visuospatial memory performance over and above working memory capacity (Miyake et al., 2001) and other general reasoning/problem solving abilities (Hegarty & Waller, 2005). These individual differences in visuospatial ability have also been shown to positively predict performance on tasks that require explicit visuospatial information processing, such as using visualizations to reason mechanically about physical objects like integrated gears or pulley systems (Boucheix & Schneider, 2009; Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997), learning a new route from a map (Sanchez & Branaghan, 2009), or learning how technologies such as bicycle pumps or surfactants work from diagrams (Höffler & Leutner, 2011; Mayer & Sims, 1994). A recent meta-analysis by Höffler (2010) has corroborated that visuospatial abilities produce a general learning benefit when learners engage in processing of visuospatial information. Individual differences in within-object manipulation abilities have also been found to predict the comprehension of narrative texts where readers follow the actions of a character in physical space (Bower & Morrow, 1990; De Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005; Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004; Haenggi et al., 1995; Meneghetti, De Beni, Pazzaglia, & Gyselinck, 2011). Thus, within-object manipulation spatial abilities (WOMSA) seem especially relevant for visuospatial processing and the development of representations which contain visuospatial information.

1.2. Visuospatial abilities and science

Although many have posited that visuospatial abilities are key for understanding in science (Halpern et al., 2007; Wu & Shah, 2004), there is little direct evidence supporting this connection. It has been noted that some of the greatest theoretical discoveries in science such as the double helix, fluid dynamics, quantum mechanics, and theories of plate tectonics, have all been attributed to the ability of great individuals to think spatially (NRC, 2006). However, the evidence for the link between visuospatial ability and performance in science is largely anecdotal (such as these examples), or instead based on observational or correlational evidence such as scientists and students in advanced science courses testing better on measures of within-object visuospatial abilities than the normal population (Self & Golledge, 1994; Wu & Shah, 2004). What has been fairly well documented are the selection factors that operate as individuals make career choices, with less-visuospatially-able students typically choosing to take fewer science, math and engineering courses (Halpern et al., 2007; Shea, Lubinski, & Benbow, 2001).

There are a handful of studies that have found correlations between classroom exam performance and within-object visuospatial abilities in such topics as organic chemistry (Bodner & McMillen, 1986; Carter, LaRussa, & Bodner, 1987; Pribyl & Bodner, 1987; Wu & Shah, 2004) and earth science (Black, 2005; Sibley, 2005). However, these correlational studies fail to rule out general individual differences in ability or even prior knowledge as causal factors. Further, there are also examples of within-object visuospatial ability failing to predict learning of science content, including in biology (Koroghlanian & Klein, 2004) and physics (ChanLin, 2000). As a result, more direct investigations of the relation between spatial abilities and science learning are needed. This prompts the question as to whether or not the traditional measures of within object visuospatial abilities are indeed the most relevant for the kinds of representations that are needed for comprehension of some scientific topics. It is possible that the simple visualization and manipulation of spatial information captured by these tasks is not necessarily relevant for all kinds of complex science reasoning occurring in these domains. Given that many scientific phenomena reflect the combination or relation of multiple objects or elements over time and space, perhaps a visuospatial ability which better captures this might be a better predictor of comprehension.

A distinct ability that might better predict comprehension in this content area of science is the construct of *multiple-object dynamic spatial ability* (MODSA) which deals with the tracking of spatial information of multiple objects across space *and* time. This construct was introduced by Hunt, Pellegrino and their colleagues approximately two decades ago (Fischer et al., 1994; Hunt et al., 1988; Law et al., 1993). In tasks that measure MODSA, subjects are asked to predict not only where multiple moving objects will intercept, but also to make judgments as to when this interception might occur. Thus, integral to performing a MODSA task is the ability to represent time and use this information to calculate relative velocity, which is then used to extrapolate a viable intercept, and to make an appropriate spatial judgment (Hunt et al., 1988). Consequently, the primary relationships being processed are spatial orientations and relations over time that are updated continuously, rather than isolated judgments of relations between features within a single object.

Consistent with this theoretical classification, Hunt, Pellegrino and colleagues originally proposed MODSA as an ability that is independent from performance on single-object manipulation tasks discussed above, as well as independent from perspective taking on tasks like the Guilford–Zimmerman, and they confirmed this with several factor analyses (Hunt et al., 1988). This has been corroborated more recently by two subsequent studies which indeed found MODSA to be an independent and significant predictor of complex spatial–temporal performance, explaining performance over and above the contributions of WOMSA measured by several common tests of spatial relations and visualization (Contreras, Colom, Hernandez, & Santacreu, 2003; D'Oliveira, 2004). It is important to note, however, that tasks designed to measure MODSA and WOMSA involve the manipulation of spatial information, so naturally these measures share a modest correlation (approximately .20–.30, Contreras et al., 2003; D'Oliveira, 2004; Hunt et al., 1988).

The relationship between MODSA and other critical cognitive abilities such as working memory capacity is less clear. MODSA has been found to explain unique variance over and above verbal intelligence on performance on IQ tests (Jackson, Vernon, & Jackson, 1993) and also appears to vary independently from education level (Contreras, Colom, Shih, Alava, & Santacreu, 2001). Working memory capacity has also been implicated in performance in several spatial reasoning and intelligence tasks (Kane et al., 2004; Miyake et al., 2001), and spatial memory/manipulation tasks (Duff & Logie, 1999; Hegarty & Steinhoff, 1997; Pearson, Logie, & Gilhooly, 1999). This suggests that these constructs could potentially share variance with one another given their similar patterns of prediction. Further, it has also been shown recently that working memory may mediate the relationship between some spatial abilities and the recall of spatial texts (Meneghetti et al., 2011). Similarly, concurrent visuospatial tasks can disrupt the consolidation of information in verbal working memory (Gyselinck, Jamet, & Dubois, 2008), and similarly, verbal concurrent tasks can negatively affect high spatial individuals when also processing spatial texts (Gyselinck, Cornoldi, Dubois, De Beni, & Ehrlich, 2002). Thus, an open question is how MODSA is related to working memory capacity, and most importantly, whether it exerts an independent influence on complex performance.

1.3. MODSA and the construction of mental models

The ability to integrate spatial information within a temporal context can be seen as directly applicable to comprehension of topics that require an understanding of a spatial/temporal sequence or process. Understanding such a process requires maintaining visuospatial or conceptual information about one unit in relation to another over time, and then transforming and using this information to form some conclusion at yet another later time point. It is possible to see the application of this aptitude across numerous instances in our daily lives, ranging from perceptual-motor tasks (i.e., throwing a football to an open receiver) to the slightly more abstract (i.e., mentally animating a satellite weather report to decide whether it will rain in your area), and even to the purely conceptual such as understanding physical systems not readily apparent to the eye (i.e., plate tectonics, neurotransmitter activity, or organic chemistry). Importantly, all these examples involve the representation of spatial information and a consideration of how it changes over time.

Most relevant here, consider more deeply the development of understanding of the science topic of plate tectonics, and its potential relation to MODSA. While the identification of the actual conceptual units of the physical system (i.e., plates, magma, lithosphere) is obviously important, to form a complete understanding of the tectonic system it is necessary to not only understand how different plates interact at fault lines, but also why and how plates move across the entire system over time. Research has shown that even though students are able to grasp simple components of the system, they often struggle with integrating this knowledge within a coherent cyclical process (Smith & Bermea, 2012). Specifically, understanding of how and why the plates move (a function of how they are created and destroyed) can provide much insight as to what drives the tectonic system. In the plate tectonic system, heated magma rises from the core at certain boundary areas. This rise of hot magma simultaneously creates new plate material, and also forces the oldest plate material (at the opposite end of the plate) back into the Earth's crust, where it is then heated and recycled into magma and the process repeats. Understanding this cyclical process provides an underlying explanation for why and how the plates are moving and interacting as they do. In other words, understanding a simple relation between two elements might not require the construction of a mental model that represents dynamic phenomena. However, when one must integrate and 'mentally animate' (Hegarty, 1992) several features/elements occurring simultaneously within a larger process (and most likely repeatedly), this could be seen as requiring the formation of a dynamic mental model or a spatial situation model.

Given that understanding in this domain relies on the creation of mental models that represent dynamic phenomena, learning in this domain could be expected to be positively related to MODSA. The observation that learning in this domain may depend on the construction of mental models that represent dynamic information also suggests that providing animations may be particularly useful for this topic. Although the results of research on benefits of animation are mixed (Höffler & Leutner, 2007; Tversky, Morrison, & Betrancourt, 2002), no work has explored how individual differences in MODSA might determine their effectiveness. High MODSA individuals might be able to learn equivalently across different types of media presentations (akin to the *ability-as-compensator* hypothesis proposed by Mayer & Sims, 1994), however low MODSA individuals might especially struggle in situations that require heavy integration of spatial and temporal information (i.e., no illustrations), and may particularly benefit from situations which offload this processing demand to the environment (i.e., animations). This is also similar to the circumvention-of-limits hypothesis proposed by Hambrick et al. (2012), in which spatial abilities are more relevant for learning in situations where there are no other cognitive resources (i.e., prior knowledge) to draw upon to aid understanding. Thus, the exploration of the possible relation between MODSA and learning from text, both with and without animations, is the main question pursued in this study. To establish the independent contributions of MODSA and WOMSA to learning, measures representing both constructs will be used to predict comprehension.

1.4. Working memory capacity and mental models

As mentioned earlier, another individual differences measure that is relevant for the formation of spatial situation models is working memory capacity (WMC). From an attentional control perspective, WMC is defined as the ability to direct one's attention towards relevant information and otherwise ignore irrelevant information (Conway & Engle, 1994; Engle, 2002). While other research traditions in working memory have been concerned with the capacity of domain-specific or modality-specific attentional resources (Miyake et al., 2001; Miyake & Shah, 1999; Shah & Miyake, 1996), the controlled attention perspective uses complex span tasks to instead derive a domain-general measure of the executive control of attention (Engle, 2002; Hambrick, Kane, & Engle, 2005; Kane et al., 2004). In complex span tasks typically used to evaluate WMC, individuals are required to simultaneously process and store information, rather than maintain lists of items as is the case with simple span tasks (Conway et al., 2005). Given this requirement, the learner must not only manage switching between tasks (i.e., processing and storage), but must also resist the buildup of proactive interference across trials (Conway & Engle, 1994). In other words, they must better utilize executive control of attention to not only perform appropriately on the sub-components of the task, but also resist the influence of non-goal-relevant information. Importantly, complex span tasks which include spatial processing and storage components (e.g. Rotation Span or Symmetry Span), or instead include more verbal processing and storage components (e.g., Reading Span and Operation Span) have been shown to load at nearly equivalent levels on this domain-general construct of controlled attention (Hambrick et al., 2005; Kane et al., 2004). This again supports the notion that estimates of WMC consist of a domain-free general executive control. In fact, it is this domain-free WMC construct, especially computed as a composite score from multiple complex span tasks (and not measures related to domain-specific short-term stores) that has been shown to significantly covary with psychometric measures of general cognitive ability, such as Ravens Advanced Progressive Matrices (factor loadings ~.70; Heitz, Unsworth, & Engle, 2005; Kyllonen & Christal, 1990; Unsworth, Redick, Heitz, Broadway, & Engle, 2009), and other performance outcomes.

Numerous studies have found a positive relationship between WMC and complex cognitive processes including reading comprehension (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Murray & Engle, 2005; Turner & Engle, 1989; Waters & Caplan, 1996), learning from illustrated science texts (Geiger & Litwiller, 2005; Sanchez & Wiley, 2006), and also learning spatial descriptions from text (Brunyé & Taylor, 2008; De Beni et al., 2005).

WMC has also been implicated in performance on several spatial memory/manipulation tasks which likewise require executive control (Duff & Logie, 2001; Hegarty & Steinhoff, 1997; Pearson et al., 1999).

As such, the reason for including a measure of individual differences in executive control (WMC) is because general ability does generally predict learning outcomes, including those that have a dedicated spatial component (e.g., Ravens Advanced Progressive Matrices). Thus, to separate the influences of general ability from those of MODSA and WOMSA, we included a measure of WMC in our analyses.

1.5. Predictions and current experiment

To evaluate the impact of different visuospatial abilities on geoscience text comprehension, participants were asked to read a scientific text about volcanic eruptions that was either non-illustrated, illustrated with static conceptual images, or illustrated with animated conceptual images. Comprehension of the text was assessed by the ability to answer the question 'What caused Mt. St. Helen's to erupt?' Importantly, the answer to this question did not appear in the reading, but required participants to apply their understanding of how and why volcanic eruptions happen in this specific case. Participants were also evaluated for WOMSA, MODSA and WMC to estimate the independent influence of each of the variables by controlling for any overlapping effects in these abilities.

[*Hypothesis 1*] Given the nature of the geoscience content, it was expected that MODSA would be significantly related to comprehension in this domain. As forming understanding in this domain relies on the creation of mental models that represent dynamic phenomena, high MODSA learners should be able to better construct an understanding that preserves a high level of congruence with the external representation of the actual phenomena (Tversky et al., 2002). Thus, these individuals should perform better in situations where a dynamic representation is necessary, such as when mental animation is required to preserve the temporal nature of the scientific phenomena being learned (Jones & Scaife, 2000; Narayanan & Hegarty, 2002; Pedone, Hummel, & Holyoak, 2001).

[*Hypothesis 2*] Provided MODSA facilitates the generation and execution of a runnable mental model or dynamic spatial situation model, the importance of this ability should also interact with the nature of the illustrations contained in the learning unit. The presence of any explicit spatial material (such as illustrations or animations), which has been shown to facilitate the mental animation process (Hegarty, Kriz, & Cate, 2003), should attenuate the relationship between MODSA and comprehension. In other words, if it is assumed that MODSA represents individual differences in how well participants can mentally animate or work with dynamic, spatial representations, it is likely that any additional material that is presented to the learner (i.e., pictures or animations) that facilitates this process should lessen the processing demand on the learner's spatial abilities (Hegarty et al., 2003). For example, if MODSA supports the creation of dynamic spatial situation models, these abilities should be most relevant in cases when this need is high, such as when there are no illustrations included in a text. In these visually impoverished cases, low MODSA learners might struggle to form understanding as they are less able to integrate the temporal and spatial components inherent to the domain. In the opposite case, if the text contains images that explicitly integrate spatial and temporal information for the learner (e.g., animations) we can expect MODSA to be less critical for learning, as the animations themselves support the creation of dynamic spatial situation models for the learner. Consistent with the *ability-as-compensator* hypothesis, it would be expected that such dynamic visualizations would produce a larger benefit for low MODSA learners, as they might explicitly address the need to represent this kind of dynamic information. It is not clear how static illustrations might interact with MODSA as they likely do relieve some of the processing burden by at least providing the spatial concepts in graphic form. However, as the learner still has to connect (or perhaps decompose) the concepts within these illustrations to achieve a dynamic understanding, learning in this situation might still be dependent on MODSA.

[*Hypothesis 3*] Finally, it is also expected that WMC and WOMSA should have a positive relationship with comprehension, considering previous research relating these abilities to learning (Daneman & Carpenter, 1980; Geiger & Litwiller, 2005; Höffler & Leutner, 2011; Sanchez & Wiley, 2006).

[*Hypothesis 4*] However, if MODSA is truly a critical ability for the construction of accurate situation models for geoscience, then MODSA should significantly predict comprehension even when the effects of WMC and WOMSA are taken into account.

2. Methods

2.1. Participants

A sample of ($N = 162$) undergraduates from a large public university participated in this experiment. All participants were compensated with course credit in their Introductory Psychology class. To limit prior knowledge about the content, only participants who had taken no more than one class *combined* in any of the following: earth science, physical science or geology were eligible for the study. In previous work, this has resulted in a sample with little knowledge of this content (Wiley et al., 2009). Participants were also randomly assigned to condition.

2.2. Design

Fifty-four participants were randomly assigned to each of the three illustration conditions. The first group read a text about volcanic eruptions that did not contain any diagrams or illustrations (63% female). The second group read the same text about volcanic eruptions, except it was illustrated with relevant static diagrams (59% female). Finally, the third group read the same text as the first two groups, save that the text now contained animated versions of the illustrations seen by the second group (41% female). WOMSA, MODSA and WMC for each individual were also recorded. All participants completed the MODSA, WMC, and reading tasks on PCs with 15″ monitors, with a resolution of 1024 × 768. All other measures were administered in paper–pencil format.

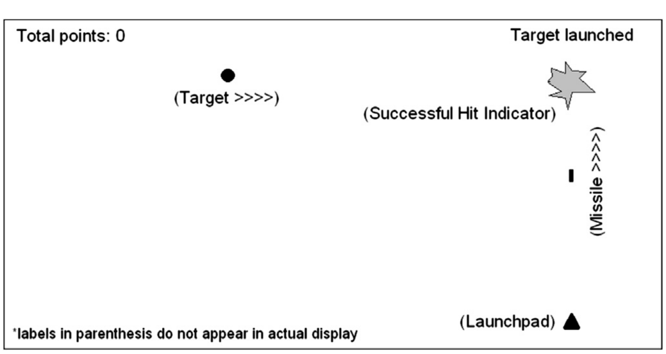


Fig. 1. Screenshot of MODSA task.

2.3. Materials

2.3.1. Multiple object/dynamic spatial ability measure: intercept task

Multiple object dynamic spatial ability was assessed using a version of the Intercept task (Hunt et al., 1988), with adjustments based on Law et al. (1993). A screenshot of this task is available in Fig. 1. The updated Intercept task consists of a game-like interface in which a small circular target (with a diameter of 10 pixels) moved horizontally across the top of the screen (from left to right) at one of three preset speeds (175 px/sec, 100 px/sec, and 75 px/sec).

In this task, the participant's goal was to 'hit' this moving circular target with a 'missile' that was launched from the lower right corner of the screen by pressing the SPACEBAR. When the spacebar was pressed, the missile moved straight up at a constant velocity of 175 px/sec, to the point of intersection that was exactly 350 pixels away. Given the constant velocity of the missile, it takes the missile 2 s to reach the intersection point. This is consistent across all trials (including practice). Thus, to successfully hit the target, the participant must release the missile when any point of the circle (e.g., target) will cross within the point of intersection.

To make the task more difficult, wait-time (and thus initial starting point) for each circular target was also manipulated by adding either 750 ms, 1000 ms, or 1250 ms to each of the three preset speeds, which resulted in 9 different trial types overall. Wait-time is defined as the time the participant must wait to release the missile to hit the circular target (e.g., how long until any portion of the target is 2 s away from the point of intersection). The starting point for every trial could be computed by the equation $2(V)+W(V)=SP$, where $V=$ velocity, and $W=$ wait time in seconds, and $SP=$ starting point.

Participants completed each of these 9 trials 7 times, resulting in 63 trials overall. Between each individual trial, there was a 5 s countdown displayed for the participant in the upper left hand corner of the screen, indicating when the next target would launch. Prior to the actual trials, participants were given 5 random practice trials of varying speeds and wait times to familiarize themselves with the interface.

Participants were awarded points for every successful collision between the missile and the target (illustrated by a small explosion). Point values ranged between 20 and 100 points (in 10 point increments), and point values were assigned based on the targets' absolute distance from the intersection point. Those trials closest to the intersection point were assigned more points than those beginning farther away, as these close trials have been shown to be more difficult (Law et al., 1993). If the participant 'missed' the target, no points were awarded, and the circle continued to move to the edge of the screen, after which the countdown to the next trial began. After all trials had been completed, the program displayed a final total score, and also listed scores for each trial type. This task was run in a browser window sized to dimensions 800×400 pixels, and the program itself was written in Java. The Intercept task takes approximately 15 min to complete, and has been shown previously to be a highly reliable indicator of MODSA (Spearman–Brown $r > .87$; Law et al., 1993). This prior work has also established that performance on the intercept task is correlated with, but separable from performance on either WOMSA or perspective taking tasks. Performance on the intercept task correlates most strongly with other measures of dynamic ability such as the race task (Hunt et al., 1988), in which participants have to also estimate and represent velocity to make a spatial judgment.

2.3.2. Within-object manipulation spatial ability measure: Paper Folding Task

The Paper Folding Task (VZ-2; French, Ekstrom, & Price, 1963) was used as the measure of WOMSA. The task consists of 2 sets of 10 trials in which the participant was shown a diagram of an irregularly folded piece of paper, with all requisite folds marked. The participants were told to imagine a single hole being punched through all the layers of paper at an indicated point, and were then asked to choose between 5 possible responses of what the paper would look like when completely unfolded. Participants were given 3 min to complete each of the 2 sets. Spearman–Brown Split-half reliability analyses were conducted on each of these 2 sets within this sample of participants, and yielded a value of .78, indicating a high degree of reliability between the two 10 item sets that is also consistent with prior research (Kane et al., 2004). Thus, the final score for this task was the overall number of correct answers out of the 20 possible, and participants were not penalized for incorrect responses. This task is generally considered a measure of spatial visualization ability.

2.3.3. WMC measure

Following the recommendations of Conway et al. (2005), a WMC measure was derived by averaging proportion scores on two complex span tasks, Operation Span and Reading Span. The WMC measures were administered in a separate session. Operation Span (OSpan) consists of several trials that are composed of a varying number of simple math equations and words that are presented individually. Participants are required to evaluate the correctness of each equation, and then remember a target word. The Reading Span (RSpan) task is functionally identical to the OSpan task, with participants evaluating whether sentences are nonsensical or not instead of equations, and remembering a target letter rather than word. Trial sizes vary between 2 and 5 equations/sentences, and there are 3 instances of each trial size in this task (randomly ordered). Participants who made more than 8 errors were not retained for any further analyses. Reliability estimates for each task were calculated following Unsworth, Heitz, Schrock, and Engle (2005), in which a proportion score for the first, second and third presentation of every set size was calculated, yielding 3 sub scores for each task. Cronbach's α was the calculated using these sub scores. Results indicated that both the OSpan ($\alpha = .80$) and RSpan ($\alpha = .77$) tasks were highly reliable, consistent with previous research (typically between .70 and .90; Conway et al., 2005).

An overall proportion correct score was then computed (# correct/trial size, averaged across all trials) for both the RSpan and OSpan measures, and these proportion scores were significantly correlated ($r = .55$, $p < .01$). These proportion scores were then averaged together and used to calculate the composite WMC score. Proportion scoring and aggregation is recommended as it increases the reliability of WMC assessment, while also reducing the influence of unique task measurement variance (Conway et al., 2005). This WMC session took no longer than 30 min to complete, and was run as part of a larger assessment of WMC ($N = 305$).

2.3.4. Reading material

The reading material was taken from the Classrooms of the Future 'Exploring the Environment – Volcanoes & the Earth' module (Center for Educational Technologies, 1997; available at http://www.cotf.edu/ete/modules/volcanoes/vlocation.html). Minor edits were made to this text to provide additional definitions of words generally unfamiliar to our sample, and also to shorten the text and reduce the complexity of sentences. The subsequent experimental text was approximately 3500 words long about volcanic eruptions and was divided into 8 separate pages that were presented through a web browser. The pages contained information about plate tectonics, location of existing volcanoes, types of plate boundaries, chemistry and volcanoes, and also types of volcanoes. In addition, the reading material included 8 concepts of an *a priori* causal model underlying volcanic eruptions in subduction zones (Fig. 2). To develop an understanding about volcanic eruptions in subduction zones, learners must incorporate and integrate these 8 concepts into their mental model. For example, they must understand that plates not only move, but also collide and subduct. This collision then produces friction and magma, which then rises into the magma chambers beneath the volcano, simultaneously building pressure until the volcano erupts. Including more of these 8 causal concepts in an explanation of why volcanoes erupt is thus likely indicative of a better or more complete appreciation of volcanic eruptions of this type.

Additionally, the text either contained no illustrations, static illustrations, or animations depending on condition. Any illustrations/animations were presented next to the text and were able to be enlarged and viewed full screen by clicking on the image. Sample text and two examples of illustrations are available in the Appendix A. Illustrations were chosen to be redundant with the text and to demonstrate the information contained within the text so that all conditions were matched for information. These images were sourced either from the Classrooms of the Future 'Volcanoes & the Earth' module, or the 'This Dynamic Planet' unit from the USGS (Kious & Tilling, 1996). The animated versions consisted of either animated .GIFs or Macromedia Flash movies that were identical in perspective and reference (e.g., spatial makeup) to the static illustrations. For example, when animating the illustration which demonstrates the subduction of the oceanic plate beneath the continental plate, the oceanic plate visibly moves beneath the continental, and magma moves toward the surface and an eruption then occurs. Again, all movement is consistent with descriptions contained within the text, and is redundant in the sense that the conceptual relationships covered in the illustrations are also referenced in the text. For those static illustrations that were merely descriptive and were not representations of actual physical movement (i.e., more decorational; Höffler & Leutner, 2007), animations were provided that presented the same information sequentially rather than all at once. In other words, the critical components of the illustrations were animated to appear in logical sequence, thus highlighting the information contained within these descriptive illustrations. Of the 11 total illustrations, only 3 were of this descriptive nature and the remaining demonstrated actual physical processes.

Table 1
Causal concepts and examples drawn from participant responses.

Concept	Coded example
1. Plates move	"…these plates make the Earth a giant jigsaw puzzle with the pieces moving violently toward or away from each other." "…the tectonic plates of the Earth's crust are constantly in motion due to the motion of the mantle beneath it."
2. Plates converge/ subduct	"The plates collided together and one plate was pushed beneath the other plate…" "…the collision between the two caused the oceanic plate to go beneath the continental plate."
3. Friction	"…because the oceanic crust rubs the continental, the temperature must increase and the rocks melt."
4. Viscous magma forms	"The melted granite causes the magma to be very viscous and sticky." "The granite from the continental crust is also heated and melted during the process."
5. More buoyant magma rises	"When [magma] is heated it becomes less dense than cold [magma], so the hot rock would start to move up, replacing the cold rock down…" "…the [magma] is so gaseous and hot that it begins to rise."
6. Magma chambers fill	"Mt. St. Helens erupted because of the buildup of magma in the magma chambers." "As the magma rises, it fills into magma chambers."
7. Cracks/weakness in crust form	"…the plates colliding also force the Earth's crust to weaken…"
8. Pressure builds and must be released	"As the magma builds up, so does pressure." "Once the pressure has reached a certain point, the gases, ash, magma and rock explode with great force…"

2.3.5. Comprehension measure

Comprehension was assessed with an essay response to the question 'What caused Mt. St. Helen's to erupt?' Participants were asked to develop an argument in response to this question, and justify their position based on what they read. This method of assessment was chosen because asking students for written

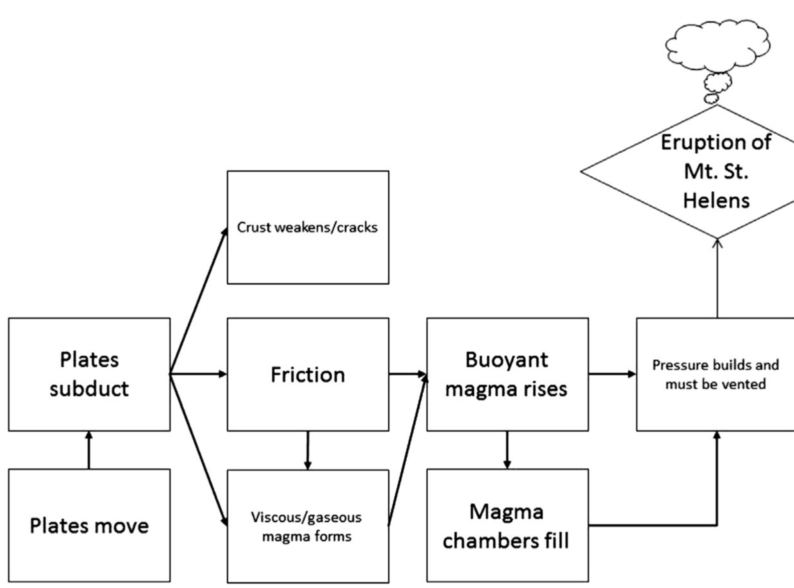


Fig. 2. Causal model of the eruption of Mt. St. Helens.

explanations or responses is a fairly common assessment tool used within classrooms. Importantly, participants needed to construct this response by connecting information across the text which was divided across 8 discrete webpages. Two example webpages are presented in the Appendix A to illustrate how participants needed to integrate information across pages to develop an understanding of this text. Responses were then evaluated for the presence of the 8 main concepts (Table 1) that should be included if readers had constructed a relatively complete situation model of volcanic eruptions from the text. For every causal concept included in the essay, participants were awarded a single point. Points were only awarded for those concepts correctly used and that matched keywords/examples drawn from prior work using these materials (Sanchez, Wiley, & Goldman, 2006; Wiley, 2001). These points were then aggregated to form an overall score on the essay response (maximum of 8). Two independent coders scored a random 60% of the overall essay responses and demonstrated a high degree of inter-rater reliability ($r_k(95) = .92$, $p < .05$). Any differences were resolved through discussion and the remaining essays were scored by a single coder.

2.4. Procedure

All participants completed the WMC measures in a separate half-hour session from the main experiment. For the main experiment, participants were each seated at individual computers, and were first required to complete the Paper Folding Task which was administered to the group in pencil-and-paper format. After completion of the Paper Folding task, participants were then asked to read the text for 15 min through an internet browser. Participants were instructed to read every page. While reading, participants were informed of the time they had left to finish reading the text by warnings given at 8 min, 3 min, and finally 1 min. These warnings were given to assure that the participants were appropriately pacing themselves to learn all the presented material.

Participants were then instructed to close the web browser, and were then asked to complete the essay task. Participants then completed the background survey, where they reported their gender and prior course experience. Participants were then required to complete the Intercept task. The entire experimental session took no longer than an hour to complete all tasks.

3. Results

There were no significant differences ($p > .05$) in WMC ($F(2, 159) = 1.21$, $MSE = .02$, $\eta_p^2 = .02$), WOMSA ($F < 1$), MODSA ($F < 1$) or the previous number of college science courses taken ($F < 1$) across illustration conditions, nor was the number of college science courses taken related to learning ($r(162) = -.11$, $p > .05$), In addition, a prior knowledge measure administered in a separate session at the beginning of the semester indicated no differences between conditions, $F(2, 124) = .13$, $MSE = 8.83$, $p > .05$. Data for this analysis were available for a majority of the participants (82%, 47 in the non-illustrated condition, 43 in the static condition, and 37 in the animated condition). The average score on the prior knowledge assessment was equivalent across the non-illustrated ($M(SD) = 19.96(2.82)$), static illustrated ($M(SD) = 19.65(3.01)$) and animated conditions ($M(SD) = 19.83(2.95)$), and there were no significant differences between the illustration conditions ($F(2, 124) = .13$, $MSE = 8.83$, $p > .05$). This suggests that these illustration groups were well matched on key variables. Additional descriptive statistics for each of these measures by illustration condition are available in Table 2.

As predicted, positive correlations between the three cognitive ability measures were found. These are shown in Table 3.

For the critical analysis, the contributions to essay performance of all individual differences measures (MODSA, WOMSA and WMC) were examined using hierarchical linear regression. All individual difference variables were centered prior to their entrance into the model. MODSA, WOMSA, WMC and prior coursework were entered into the first block of the regression to test for independent influences for each of these variables. Illustration condition was entered into the second block of the analysis. Prior to being entered into the second block of analyses illustration condition was dummy-coded into 2 discrete variables. The first dummy variable conceptually captures whether or not the text contained illustrations (hereto referred to as the illustration dummy variable). Thus, the non-illustrated text was coded 0 for this variable, and the remaining 2 conditions were coded with a 1. The second dummy variable (hereto referred to as the dynamic dummy variable) represents whether the text contained additional dynamic information in the form of visualizations. As such, both the non-illustrated and static image conditions were coded with a 0 for this variable, and the animated condition received a 1. Across these 2 dummy

Table 2
Descriptive statistics ($M(SD)$) for cognitive ability measures by illustration condition.

	WMC	WOMSA	MODSA	# Courses	Essay
Non-illustrated	.66(.12)	10.85(3.71)	982.78(339.05)	.30(.46)	2.04(1.15)
Static	.64(.13)	11.15(3.81)	998.33(322.01)	.20(.41)	2.30(1.31)
Animated	.67(.12)	11.50(3.31)	1055.00(324.31)	.22(.42)	2.67(1.33)
Total	.66(.12)	11.17(3.60)	1012.04(327.98)	.24(.43)	2.33(1.29)

Note. WMC – Working memory capacity, WOMSA – Within-object manipulation spatial abilities, MODSA – Multiple-object dynamic spatial ability.

Table 3
Pearson correlations among individual difference measures and essay scores.

	MODSA	WOMSA	WMC	ESSAY
MODSA	–	.17*	.18*	.32*
WOMSA		–	.34*	.12
WMC			–	.23*

Note: *$r(162)$, $p < .05$.
Note. WMC – Working memory capacity, WOMSA – Within-object manipulation spatial abilities, MODSA – Multiple-object dynamic spatial ability.

Table 4
Hierarchical regression results for essay performance ($N = 162$).

Predictor	Essay performance	
	ΔR^2	β
Block 1	.14**	
# Courses		−.08
WOMSA		.007
WMC		.17*
MODSA		.29*
Block 2	.03	
Dynamic dummy variable		.09
Illustration dummy variable		.10
Block 3	.04*	
MODSA × Dynamic dummy		−.69*
MODSA × Illustration dummy		−.08
Block 4	.006	
WMC × Dynamic dummy		.28
WMC × Illustration dummy		.22
Block 5	.006	
WOMSA × Dynamic dummy		−.21
WOMSA × Illustration dummy		.32
Total R^2	.23**	

Note. *$p < .05$, **$p < .01$. WMC – Working memory capacity, WOMSA – Within-object manipulation spatial abilities, MODSA – Multiple-object dynamic spatial ability.

variables, the final coding for each condition was as follows: non-illustrated (0, 0), static images (1, 0), and animations (1, 1). Interaction terms between these dummy variables and all 3 individual difference variables were then computed. Each set of interaction terms was entered into a subsequent block of the analysis to test for any type of interaction between these abilities and illustration condition, and to examine the potential individual contribution of these interactions on predicting learning performance. The interaction terms for MODSA were entered in the third block, and the interaction terms of WMC and WOMSA were entered in the fourth and fifth blocks, respectively.

Overall results are presented in Table 4. Results from the first block of analyses indicate that the overall model predicted a significant portion of variance in essay scores ($R^2 = .14$, $F(4, 157) = 6.48$, $MSE = 1.45$, $p < .01$). WMC ($\beta = .17$) and MODSA ($\beta = .29$) were both significant predictors ($p < .05$) of learning. However, neither WOMSA ($\beta = .007$), or number of previous courses ($\beta = -.08$) were predictive of performance. This suggests that overall WMC and MODSA exerted a significant positive influence on learning from the geosciences texts and produced significantly better learning. Results from the second block of analyses also indicated that the illustration dummy variable ($\beta = .10$), and dynamic dummy variable ($\beta = .09$) also failed to significantly predict performance ($R^2\Delta = .03$, $p > .05$). This suggests that illustration condition alone was not able to reliably predict learning performance.

To examine whether any individual difference variables interacted with illustration condition, interaction terms were entered in the third, fourth, and fifth blocks of the analysis. Again, MODSA was entered in the third, WMC was entered in the fourth block and WOMSA was entered in the fifth block. Results from the third block of analyses likewise produced a significant overall model ($R^2 = .21$, $F(8, 153) = 5.18$, $MSE = 1.37$, $p < .01$), and also predicted an additional significant portion of variance ($R^2\Delta = .04$, $p < .05$). While the interaction term between MODSA and the dynamic dummy variable was a significant predictor ($\beta = -.69$), the interaction term between MODSA and illustration dummy variable was not significant ($\beta = -.08$). Given the directionality of the standardized beta weight and the dummy coding, this suggests that MODSA is LESS related to comprehension in conditions which contain more dynamic information (e.g., animations), and more relevant in situations that do not contain such explicit information (e.g., non-illustrated or static), see Fig. 3. Further, the presence of illustrations does not seem to interact with MODSA in a general sense, suggesting that the use of MODSA is not limited to situations which only contain graphic information. The addition of the interaction terms for WMC in the fourth block ($R^2\Delta = .006$), and WOMSA in the fifth block ($R^2\Delta = .006$) failed to account for a significant additional portion of variance ($p > .05$). This suggests that these individual difference variables do not significantly interact with either the illustrated or dynamic nature of the visualizations.

To ensure that this pattern of results was not obfuscating potential interactions due to the order in which the blocks were entered into the model, the analysis was repeated, however the order of blocks was reversed (WOMSA in 3rd block, WMC in 4th block, and MODSA in 5th block). Results were identical, as the interaction terms for both WOMSA ($R^2\Delta = .007$) and WMC ($R^2\Delta = .002$) failed to significantly predict additional overall variance ($p > .05$), but the addition of the interaction terms with MODSA did predict a significant portion of additional variance ($R^2\Delta = .05$, $p < .05$), again driven solely by the interaction between MODSA and the dynamic dummy variable ($\beta = -.71$).

To further illuminate the differential impact of MODSA on learning across the illustration conditions, simple follow-up regressions were conducted on each illustration condition

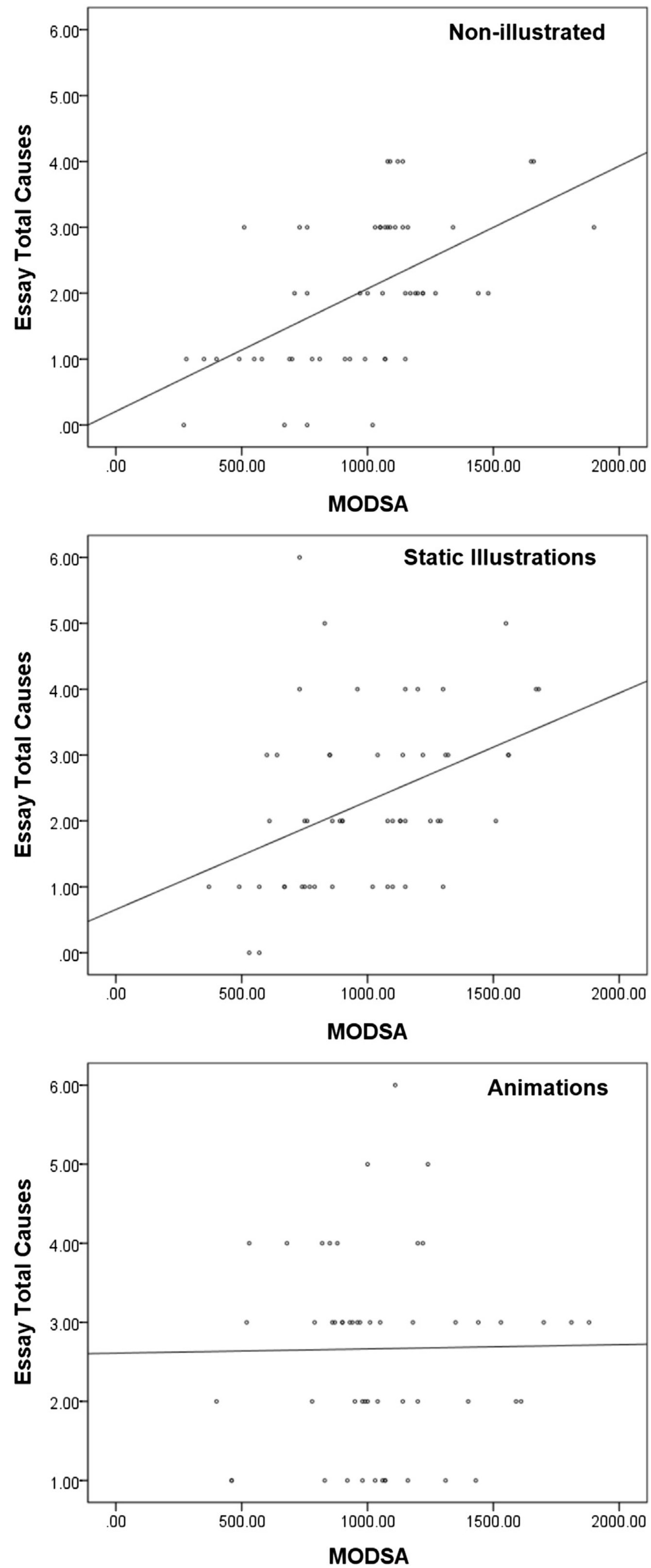


Fig. 3. Regression plots showing the relationship between MODSA and Essay recall across illustration conditions.

independently. In these analyses, MODSA alone was used to predict performance on the essay task. Results indicated that MODSA was significantly related to learning in both the non-illustrated ($R^2 = .30$, $F(1, 52) = 22.51$, $MSE = .94$, $p < .01$) and static conditions ($R^2 = .16$, $F(1, 52) = 10.09$, $MSE = 1.47$, $p < .01$), but was NOT related to learning in the animation condition ($R^2 = .00$, $F(1, 52) = .01$, $MSE = 1.81$, $p > .05$). As the animated condition produced the best learning overall (see above), this suggests that MODSA is important in learning situations that do not explicitly contain a temporal element in the illustration. In other words, this means that when the learner is forced to 'mentally animate' the materials themselves, MODSA is especially relevant. When this is done for the learner (as in the case of animations), MODSA becomes less relevant.

To build on these regression results, a final analysis examined the subset of concepts that are most explicitly connected to the dynamic causal nature of the plate interaction. This subset consisted of 5 concepts: (1) plates move, (2) plates converge, (3) heated magma rises, (4) magma chambers fill, and (5) pressure builds and is released. These concepts represent the critical ideas within the text that most succinctly capture the idea of plates moving and converging within the tectonic system, which subsequently produce a change in the system (i.e., chambers filling and pressure) that must be released. This subset was contrasted with the remaining 3 concepts (e.g., friction is generated, cracks/weakness form in the Earth's crust, magma is viscous), which although still connected to the eruption of Mt. St. Helens, are more descriptive in nature. In other words, these 3 descriptive concepts are indicative of outcomes of the key concept of plates converging (i.e., friction/magma/cracks occur because the plates converged), but are less related to the dynamic phenomena underlying volcanic eruptions.

Given the interaction between the dynamic dummy variable and MODSA found above, performance in the non-illustrated and static conditions was aggregated and compared to the animated condition in a planned comparison. Participants in the animated condition ($M = 2.33$, $SD = 1.06$) recalled significantly more of these dynamic concepts ($F(1, 160) = 10.73$, $MSE = 1.07$, $p < .01$, $\eta_p^2 = .06$) than participants in the non-illustrated/static conditions ($M = 1.77$, $SD = 1.02$). This suggests that learners struggle learning these 5 dynamic concepts when the environment does not explicitly represent them. For the remaining 3 descriptive concepts, there was no difference between the non-illustrated/static conditions ($M = .40$, $SD = .56$) and the animated condition ($M = .33$, $SD = .55$; $F < 1$). Taken together, this suggests that the learning of a particular subset of dynamic concepts might explain the interaction between MODSA and illustration condition observed in the overall analysis. As such, the differences in essay performance between conditions could be interpreted as the result of the presence of dynamic visualizations that facilitated low MODSA reader's mental representations of the dynamic phenomenon.

4. Discussion

Consistent with *Hypothesis 1*, the results of this experiment indicate that MODSA is generally related to better comprehension for geosciences texts on plate tectonics. However, consistent with *Hypothesis 2*, low MODSA individuals developed less understanding particularly in the non-illustrated and static image conditions, suggesting that these conditions required participants to generate and mentally animate their own internal representations to a higher degree than the animated condition. These findings support the notion that MODSA supports the generation of dynamic spatial mental representations, and thus MODSA influences comprehension when such representations are required. While previous studies have highlighted the importance of MODSA in performing other complex tasks (e.g., air traffic control tasks; D'Oliveira, 2004), this represents the first direct application of individual differences in dynamic spatial ability to the formation of situation models and text comprehension.

Also, it appears that the addition of animations can positively impact performance, and that this effect is moderated by MODSA. While it has been found previously that conceptually relevant images or animations can improve learning (ChanLin, 2000; Craig, Gholson, & Driscoll, 2002; Mayer & Moreno, 2002; Rieber, 1990), animations have also been found previously to differentially impact high and low ability learners (Höffler & Leutner, 2011; Mayer & Sims, 1994; Schnotz & Rasch, 2005). The results of this experiment clarify and expand on these previous results by demonstrating that animations do not facilitate situation model construction for all learners, but rather specifically for learners who are less able to mentally animate their own mental representations. An alternative explanation for the comprehension benefit from viewing animations is that perhaps providing these representations served to highlight relevant or important information for readers. However, if this were the case, one might also expect at least some kind of facilitation for those higher in MODSA in the animated condition. The data seem to fit better with the notion that animations assist low MODSA learners by providing an external support for what might otherwise be an internal process, consistent with the *ability-as-compensator* hypothesis (Höffler & Leutner, 2011; Mayer & Sims, 1994).

The implications of the results from the static illustration condition are perhaps more tentative. While learning was higher in the static condition than the non-illustrated condition, this difference was not statistically reliable. Further, MODSA was still significantly related to comprehension in the static condition, suggesting that the presence of the static diagrams did not in fact offload all demand to process such dynamic information. While this is somewhat contradictory to prior research regarding the effectiveness of static illustrations (Levie & Lentz, 1982), it perhaps highlights that such images are not of optimal utility in domains which themselves contain a dynamic component (e.g., plate tectonics). By not representing this dynamic information externally for the learner, and thus failing to appropriately match the visualizations to the task at hand, this likely caused additional demand to be placed on the learner. Failure to appropriately match visualizations to task demands has been observed previously to reduce the positive impact of visualizations on learning (Schnotz & Bannert, 2003), and so perhaps this finding is not so surprising.

Taken together, the results of this study provide an explanation for the previous inconsistency of findings regarding the benefits of visual adjuncts, with some studies suggesting a positive benefit of illustrations and animations (Park & Gittelman, 1992; Rieber, 1990) and others suggesting no benefit or even negative effects (Hegarty et al., 2003; Mayer, Hegarty, Mayer, & Campbell, 2005; Sanchez & Wiley, 2006; Tversky et al., 2002). It is possible that these disparate findings could be a result of a failure to consider both individual differences in visuospatial ability, and the visuospatial nature of material to be learned, which might lead to divergent results and interactions.

In relation to *Hypothesis 3*, WOMSA did not significantly predict essay performance. This finding is somewhat surprising given previous research that has found WOMSA to be related to developing understanding from both static and animated materials (Hays, 1996; Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997; Höffler & Leutner, 2011; Koroghlanian & Klein, 2004; Mayer & Sims, 1994), and understanding geologic features (Hambrick et al., 2011). However, the present results are consistent with other studies that have failed to find an independent contribution of WOMSA on learning of science (ChanLin, 2000; Koroghlanian & Klein, 2004). It is possible that given the temporal complexity of the content in this study, there was little demand on the simple visualization ability measured by the WOMSA task, and instead priority was placed on the more pressing need to integrate spatial information over time to form a dynamic representation, which was better captured by the Intercept task. In other words, variance typically attributed to WOMSA in other studies where MODSA is not assessed, might have instead been attributed more appropriately to the MODSA measure in this study, thus resulting in no independent contribution of WOMSA. On the other hand, essay performance was predicted by WMC, which is also consistent with previous research (Geiger & Litwiller, 2005; Sanchez & Wiley, 2006) and *Hypothesis 3*. However, WMC did not significantly interact with illustration condition, suggesting a uniform influence regardless of the presence of illustrations. The lack of an interaction might be, in part, as a result of the use of WMC measures that primarily involved the processing and storage of verbal information. Other WMC measures that primarily involve the processing and storage of spatial information (e.g., Symmetry Span), might potentially demonstrate a stronger relationship to learning with visualizations as they invoke the visuospatial store, which has shown previous relations to learning from text and pictures (Gyselinck et al., 2002, 2008). Critically, as the WMC assessments used here do not tap the visuospatial store, in the current context they represent a more pure estimate of general ability, sidestepping this issue of overlapping task requirements that are utilizing the same cognitive resource. Most importantly, however, the relation between MODSA and essay performance in this study was demonstrated over and above this overall relation with WMC and WOMSA, consistent with *Hypothesis 4*.

While the current set of results focuses on a specific facet of geoscience, it is hoped that future work will further explore this potential link between MODSA and science learning in other content domains. This study also utilized only a single measure of MODSA, the Intercept task, and it would also be useful for future work to explore how different measures of MODSA might differently predict learning. If other MODSA tasks are more diagnostic of classroom learning, it will be worthwhile to identify these alternate measures. Further, while the approach taken here assumes that even low MODSA learners are attempting to mentally animate the content, it is not clear that this is the case. Future work needs to explore how exactly learners who vary in MODSA are developing their mental models of the content, and the contexts under which low MODSA learners can successfully develop causal models of the phenomenon. Trace measures such as eyetracking or think-aloud protocols that directly focus on how concepts are being integrated during reading would cast more light on the role of MODSA in the comprehension process. Given the overall low level of learning observed here, such additional investigations might illuminate more precisely how these abilities connect to the formation of more deep conceptual understanding. It might also be useful to include other measures of learning that are not predominantly based on free recall and that might be more sensitive, such as graphical recall or measures of recognition memory. Alternative approaches to understanding the contributions of modality specific working memory capacity toward learning (i.e., using WMC measures that rely on visuo-spatial processing and storage) may also be interesting to pursue. Finally, the fact that MODSA appears to improve with practice (Law et al., 1993) suggests that a promising line of future investigation is to see whether such training ultimately might also translate into better science comprehension by improving the ability to construct dynamic mental models (cf. Sanchez, 2012).

5. Conclusions

In conclusion, this study has provided evidence for a direct link between a particular kind of visuospatial ability, *multiple object/dynamic spatial ability*, and comprehension of geoscience texts that describe physical phenomena that change over time. While WMC did significantly predict learning, WOMSA and illustration characteristics alone did not. However, it was found that illustration characteristics (specifically the presence of dynamic elements) did interact with MODSA, but not WMC or WOMSA, which potentially sheds light on previous results regarding the efficacy of adding illustrations to science text.

The demonstration of an effect of MODSA on text comprehension makes a novel empirical contribution beyond anecdotal and correlational findings which have previously been used to suggest that spatial thinking is critical for cognitive activities in science. In the broader context of concern over the lack of students who pursue careers in science math and engineering, the observation that providing animations can potentially help individuals with low MODSA reach higher levels of comprehension offers additional hope that simulations may be an effective means of lessening gaps in science learning. For example, as young students enter into courses of study early on in their educational career, supportive instruction at this initial time point could allow all students to more successfully learn the basic material. This would provide a stronger foundation which could potentially translate into higher academic success for these students, and help to address the much needed deficit in human capital related to careers in the STEM fields (Halpern et al., 2007).

Acknowledgments

This research was funded in part by an APA Dissertation Award to Christopher A. Sanchez. Additional portions of this research were supported by a grant from the National Science Foundation to Jennifer Wiley (REC 0126265 Understanding in Science; DRL 0735569 Supporting Whole-Class Science Investigations). All opinions expressed herein are those of the authors are do not necessarily reflect those of the funding organization.

Appendix A

Text from 2 discrete pages from the volcano text, and sample images:

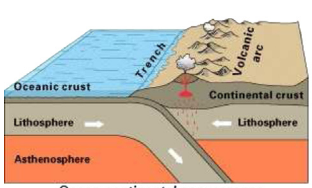
… What causes the plates to move? It turns out to be a consequence of the high temperatures inside Earth. Common experience tells us that heat flows from hot to cold, so the heat in Earth's deep interior must be flowing somehow to the surface. The hottest part of Earth's interior is the iron core. The core heats the bottom of the rocky mantle. The hottest rock near the bottom of the mantle becomes slightly less dense than the somewhat cooler rock above it, so buoyancy forces try to push the hottest rocks upward. Although the rock in the mantle is solid, the pressures and heat are so great that the rock can deform slowly, like hot wax. So the hot rock creeps upward through the cooler rock. As the hot rock rises, cooler rock flows downward to take its place next to the core, where it is heated and becomes buoyant enough to rise again later. The rising hot rock comes in contact with cold rocks near the surface of Earth where it gives off its heat, cools, and sinks again. Most of the rock in the mantle moves in this broad cyclic flow, indicated by the arrows in the figure. This zone, where rock is soft enough to flow, is called the asthenosphere. The cyclical movement of hot and cold material is called convection.

…Rock near the surface of Earth is so cold and at such low pressures that it cannot flow like mantle rock. So how does heat get through this rigid layer lithosphere, to the surface? It flows through cracks in the earth's surface. Places where liquid rock (lava) flows onto Earth's surface are called volcanoes. The movement of heat by convection in the asthenosphere causes the rock of the mantle to slowly move in huge streams. The solid (but brittle) rock of the lithosphere is resting directly on top of the solid (but soft) rock of the asthenosphere. As the rock of the asthenosphere moves in different directions, it carries parts of the lithosphere along with it. The lithospheric rock can't stretch, so it breaks into pieces – forming the plates. This whole group of observations and ideas describing the motions of the plates and their geologic features is called plate tectonics.

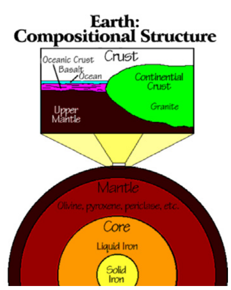
…Let's carefully examine the edges of colliding plates to see why volcanoes might be found there. Plates move because heat flowing up from Earth's core causes mantle rock to slowly flow in giant convection currents. The rigid plates are carried on the currents, crunching into each other, and pulling apart from each other.

Land volcanoes tend to occur where plates are crunching into each other. Ocean volcanoes tend to occur where plates are separating. Both are a part of a system in which crust is created and destroyed. The sea-floor spreading hypothesis says that new crust is created where plates separate in the ocean floor, while old oceanic crust is "recycled" when it collides with continental plates, and is pushed back down into the mantle. This is how new crust can be created, but the earth's size remains constant. It also explains why oceanic rocks are younger than continental rocks, and how the Atlantic Ocean is increasing while the Pacific Ocean is shrinking.

Moving plates interact in a number of different ways at their edges. When plates are colliding, it is called a convergent boundary, when plates are separating it is called a divergent boundary, and when plates are sliding past each other, it is called a transform boundary. At converging boundaries, the edges may be either oceanic crust or continental crust. The kind of plates that are part of a collision influences the geological events and features that may be found in the boundary zone.



Ocean-continental convergence



References

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

Bodner, G. M., & McMillen, T. L. B. (1986). Cognitive restructuring as an early stage in problem solving. *Journal of Research in Science Teaching, 23*, 727–737. http://dx.doi.org/10.1002/tea.3660230807.

Boucheix, J. M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction, 19*, 112–127. http://dx.doi.org/10.1016/j.learninstruc.2008.03.004.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*(4938), 44–48. http://dx.doi.org/10.1126/science.2403694.

Brunyé, T. T., & Taylor, H. A. (2008). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language, 58*, 701–729. http://dx.doi.org/10.1016/j.jml.2007.08.003.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, England: Cambridge University Presshttp://dx.doi.org/10.1017/CBO9780511571312.

Carter, C. S., LaRussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching, 24*, 645–657. http://dx.doi.org/10.1002/tea.3660240705.

Center for Educational Technologies. (1997). *Volcanoes & the earth. In exploring the environment*. Retrieved from http://www.cotf.edu/ete/modules/volcanoes/vlocation.html.

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

Contreras, M. J., Colom, R., Hernandez, J. M., & Santacreu, J. (2003). Is static spatial performance distinguishable from dynamic spatial performance? A latent-variable analysis. *The Journal of General Psychology, 130*, 277–288. http://dx.doi.org/10.1080/00221300309601159.

Contreras, M. J., Colom, R., Shih, P. C., Alava, M. J., & Santacreu, J. (2001). Dynamic spatial performance: sex and educational differences. *Personality and Individual Differences, 30*, 117–126. http://dx.doi.org/10.1016/S0191-8869(00)00015-5.

Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354–373. http://dx.doi.org/10.1037/0096-3445.123.4.354.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769–786. http://dx.doi.org/10.3758/BF03196772.

Cooper, L. A. (1975). Mental transformations of random two-dimensional shapes. *Cognitive Psychology, 7*, 20–43. http://dx.doi.org/10.1016/0010-0285(75)90003-1.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*, 428–434. http://dx.doi.org/10.1037/0022-0663.94.2.428.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. http://dx.doi.org/10.1016/S0022-5371(80)90312-6.

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology, 17*, 77–95. http://dx.doi.org/10.1080/09541440340000529.

D'Oliveira, T. C. (2004). Dynamic spatial ability: an exploratory analysis and a confirmatory study. *International Journal of Aviation Psychology, 14*, 19–38. http://dx.doi.org/10.1207/s15327108ijap1401_2.

Duff, S. C., & Logie, R. H. (1999). Storage and processing in visuo-spatial working memory. *Scandinavian Journal of Psychology, 40*, 251–259. http://dx.doi.org/10.1111/1467-9450.404124.

Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology: Section A, 54*, 31–48. http://dx.doi.org/10.1080/02724980042000011.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23. http://dx.doi.org/10.1111/1467-8721.00160.

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory and Cognition, 29*, 336–343. http://dx.doi.org/10.3758/BF03194928.

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes, 37*, 205–224. http://dx.doi.org/10.1207/s15326950dp3703_2.

Fischer, S. C., Hickey, D. T., Pellegrino, J. W., & Law, D. J. (1994). Strategic processing in dynamic spatial reasoning tasks. *Learning and Individual Differences, 6*, 65–105. http://dx.doi.org/10.1016/1041-6080(94)90015-9.

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

Friedman, N. P., & Miyake, A. (2000). Differential roles for spatial and verbal working memory in the comprehension of spatial descriptions. *Journal of Experimental Psychology: General, 129*, 61–83. http://dx.doi.org/10.1037/0096-3445.129.1.61.

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). In *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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. http://dx.doi.org/10.1207/s1532799xssr0203_4.

Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology, 16*, 665–685. http://dx.doi.org/10.1002/acp.823.

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

Haenggi, D., Kintsch, W., & Gernsbacher, M. A. (1995). Spatial situation models and text comprehension. *Discourse Processes, 19*, 173–199. http://dx.doi.org/10.1080/01638539509544913.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in mathematics. *Psychological Science in the Public Interest, 8*, 1–51. http://dx.doi.org/10.0000/j.1529-1006.2007.00032.x.

Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition. In R. Sternberg, & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 104–121). New York: Cambridge University Press.

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., et al. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: the case of geological bedrock mapping. *Journal of Experimental Psychology: General, 141*, 397–403. http://dx.doi.org/10.1037/a0025927.

Hays, T. A. (1996). Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *Journal of Educational Computing Research, 14*, 139–155. http://dx.doi.org/10.2190/60Y9-BQG9-80HX-UEML.

Hegarty, M. (1992). Mental animation: inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084–1102. http://dx.doi.org/10.1037/0278-7393.18.5.1084.

Hegarty, M., Canham, M., & Fabrikant, S. I. (2010). Thinking about the weather: how display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 37–53. http://dx.doi.org/10.1037/a0017683.

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. http://dx.doi.org/10.1207/s1532690xci2104_1.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*, 411–430. http://dx.doi.org/10.3758/BF03200867.

Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as memory in mechanical reasoning. *Learning and Individual Differences, 9*, 19–44. http://dx.doi.org/10.1016/S1041-6080(97)90018-2.

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *Handbook of visuospatial thinking* (pp. 121–169). New York: Cambridge University Press.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61–77). New York: Sage. http://dx.doi.org/10.4135/9781452233529.n5.

Höffler, T. N. (2010). Spatial ability: its influence on learning with visualizations—a meta-analytic review. *Educational Psychology Review, 22*, 245–269. http://dx.doi.org/10.1007/s10648-010-9126-7.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction, 17*, 722–738. http://dx.doi.org/10.1016/j.learninstruc.2007.09.013.

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations – evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior, 27*, 209–216. http://dx.doi.org/10.1016/j.chb.2010.07.042.

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. http://dx.doi.org/10.1016/0160-2896(88)90024-4.

Jackson, D. N., Vernon, P. A., & Jackson, D. N. (1993). Dynamic spatial performance and general intelligence. *Intelligence, 17*, 451–460. http://dx.doi.org/10.1016/0160-2896(93)90012-T.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Jones, S., & Scaife, M. (2000). Animated diagrams: an investigation into the cognitive effects of using animation to illustrate dynamic processes. In M. Anderson, & P. Cheng (Eds.), *Lecture note in artificial intelligence: Vol. 1889. Theory and application of diagrams* (pp. 231–244). http://dx.doi.org/10.1007/3-540-44590-0_22.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137–172. http://dx.doi.org/10.1037/0033-295X.92.2.137.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review, 99*, 122–149. http://dx.doi.org/10.1037/0033-295X.99.1.122.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. http://dx.doi.org/10.1037/0096-3445.133.2.189.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294–303. http://dx.doi.org/10.1037/0003-066X.49.4.294.

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

Kious, W. J., & Tilling, R. I. (1996). *This dynamic planet*. Retrieved from http://pubs.usgs.gov/gip/dynamic/dynamic.html.

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

Kortz, K. M., Clark, S. K., Gray, K., Smay, J. J., Viveiros, B., & Steer, D. (2011). Counting tectonic plates: A mixed-methods study of college student conceptions of plates and boundaries. *Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper, 474*, 171–188.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working- memory capacity?! *Intelligence, 14*, 389–433. http://dx.doi.org/10.1016/S0160-2896(05)80012-1.

Law, D. J., Pellegrino, J. W., & Hunt, E. B. (1993). Comparing the tortoise and the hare: gender difference and experience in dynamic spatial reasoning tasks. *Psychological Science, 4*, 35–40. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00553.x.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: a review of research. *Educational Technology Research and Development, 30*, 195–232. http://dx.doi.org/10.1007/BF02765184.

Mayer, R. E., Hegarty, M., Mayer, S. Y., & Campbell, J. (2005). When passive media promote active learning: static diagrams versus animation in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*, 256–265. http://dx.doi.org/10.1037/1076-898X.11.4.256.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107–119. http://dx.doi.org/10.1016/S0959-4752(01)00018-4.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extension of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401. http://dx.doi.org/10.1037/0022-0663.86.3.389.

Meneghetti, C., De Beni, R., Pazzaglia, F., & Gyselinck, V. (2011). The role of visuo-spatial abilities in recall of spatial descriptions: a mediation model. *Learning and Individual Differences, 21*, 719–723. http://dx.doi.org/10.1016/j.lindif.2011.07.015.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*, 621–640. http://dx.doi.org/10.1037/0096-3445.130.4.621.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Presshttp://dx.doi.org/10.1017/CBO9781139174909.

Mumaw, R. J., Pellegrino, J. W., Kail, R. V., & Carter, P. (1984). Different slopes for different folks: process analysis of spatial aptitude. *Memory & Cognition, 12*, 515–521. http://dx.doi.org/10.3758/BF03198314.

Murray, J. D., & Engle, R. (2005). Accessing situation model information: memory based processing versus here-and-now accounts. *Journal of Psychology, 139*, 261–273. http://dx.doi.org/10.3200/JRLP.139.3.261-272.

Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*, 279–315. http://dx.doi.org/10.1006/ijhc.2002.1019.

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

Newcombe, N. S., & Shipley, T. F. (2012). Thinking about spatial thinking: new typology, new assessments. In J. S. Gero (Ed.), *Studying visual and spatial reasoning for design creativity* (pp. 1–18). New York, NY: Springer.

Park, O. C., & Gittelman, S. S. (1992). Selective use of animation and feedback in computer-based instruction. *Educational Technology Research and Development, 40*, 125–167. http://dx.doi.org/10.1007/BF02296897.

Pearson, D. G., Logie, R. H., & Gilhooly, K. J. (1999). Verbal representations and spatial manipulation during mental synthesis. *European Journal of Cognitive Psychology, 11*, 295–314. http://dx.doi.org/10.1080/713752317.

Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory & Cognition, 29*, 214–221. http://dx.doi.org/10.3758/BF03194915.

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* (pp. 203–225). Mahwah, NJ: Lawrence Erlbaum Associates.

Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: a study of four organic courses. *Journal of Research in Science Teaching, 24*, 229–240. http://dx.doi.org/10.1002/tea.3660240304.

Rieber, L. P. (1990). Using computer animated graphics with science instruction in children. *Journal of Educational Psychology, 82*(1), 135–140. http://dx.doi.org/10.1037/0022-0663.82.1.135.

Rinck, M. (2005). Spatial situation models. In P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 334–382). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610448.010.

Sanchez, C. A. (2012). Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains. *Psychonomic Bulletin and Review, 19*, 58–65. http://dx.doi.org/10.3758/s13423-011-0177-7.

Sanchez, C. A., & Branaghan, R. J. (2009). The interaction of map resolution and spatial abilities on map learning. *International Journal of Human-Computer Studies, 67*, 475–481. http://dx.doi.org/10.1016/j.ijhcs.2008.12.003.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*, 344–355. http://dx.doi.org/10.3758/BF03193412.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. In S. A. Barab, & D. T. Hickey (Eds.), *Proceedings of the 7th international conference of the learning sciences* (pp. 662–666). Mahwah, NJ: Erlbaum.

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

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animation in multimedia learning: why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development, 53*, 1042–1629. http://dx.doi.org/10.1007/BF02504797.

Self, C. M., & Golledge, R. G. (1994). Sex-related differences in spatial ability: what every geography educator should know. *Journal of Geography, 93*, 234–243. http://dx.doi.org/10.1080/00221349408979727.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General, 125*, 4–27. http://dx.doi.org/10.1037/0096-3445.125.1.4.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *Journal of Educational Psychology, 93*, 604–614. http://dx.doi.org/10.1037/0022-0663.93.3.604.

Sibley, D. F. (2005). Visual abilities and misconceptions about plate tectonics. *Journal of Geoscience Education, 53*, 471–477.

Smith, G. A., & Bermea, S. B. (2012). Using students' sketches to recognize alternative conceptions about plate tectonics persisting from prior instruction. *Journal of Geoscience Education, 60*, 350–359. http://dx.doi.org/10.5408/11-251.1.

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. http://dx.doi.org/10.1016/0191-8869(95)00029-6.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*, 612–630. http://dx.doi.org/10.1016/0749-596X(85)90049-X.

Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154. http://dx.doi.org/10.1016/0749-596X(89)90040-5.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262. http://dx.doi.org/10.1006/ijhc.2002.1017.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505. http://dx.doi.org/10.3758/BF03192720.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: a latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654. http://dx.doi.org/10.1080/09658210902998047.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*, 352–402. http://dx.doi.org/10.1037/a0028446.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review, 103*, 761–772. http://dx.doi.org/10.1037/0033-295X.103.4.761.

Wiley, J. (2001). Supporting understanding through task and browser design. In L. R. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1136–1143). Hillsdale, NJ: Erlbaum.

Wiley, J., Goldman, S., Graesser, A., Sanchez, C. A., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal, 46*, 1060–1106. http://dx.doi.org/10.3102/0002831209333183.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes, 36*, 109–129. http://dx.doi.org/10.1207/S15326950DP3602_2.

Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88*, 465–492. http://dx.doi.org/10.1002/sce.10126.

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