When, and for Whom, Analogies Help: The Role of Spatial Skills and Interleaved Presentation

Allison J. Jaeger, Andrew R. Taylor, and Jennifer Wiley
University of Illinois at Chicago

Understanding many scientific phenomena, processes, or systems may be especially dependent on a student’s ability to visualize or manipulate spatial information in order to construct mental representations. One instructional technique often included in science texts to help students understand difficult concepts is the use of concrete or familiar analogies. Two experiments tested whether individual differences in spatial skills may impact the effectiveness of learning by analogy, if an analogy might particularly improve the learning of low-spatial students, and if the way in which the analogical comparison is presented matters. In these studies, students read a text about the processes and causes of the global weather phenomenon known as El Niño. For some students, the text also contained an analogy that compared El Niño to the inflating a deflating of a balloon; this analogy was either presented at the beginning of the text or interleaved throughout the text. Across both experiments, results indicated that spatial skills generally improved learning from a text about El Niño, but that interleaving an analogy changed the relationship between spatial skills and learning, and improved performance for low-spatial learners.

Keywords: analogical reasoning, spatial skills, text comprehension, science learning

Students are called upon to learn a vast landscape of science concepts during any particular academic year. It has been suggested that understanding many scientific phenomena, processes, or systems may be especially dependent on the student’s ability to visualize or manipulate spatial information and to understand visual representations, making them especially challenging for students who are poor at spatial thinking (Hegarty, 2010; National Research Council, 2012). Consistent with these assumptions, spatial skills have been shown to be positively related to performance and entry into the science, technology, engineering, and mathematics (STEM) domains (for review, see Uttal & Cohen, 2012).

One instructional technique often included in science texts to help students understand difficult concepts is the use of concrete or familiar analogies (Brown, 1993; Curtis & Reigeluth, 1984; Duit, 1991; Gentner & Stevens, 1983; Iding, 1997; Orgill & Bodner, 2006; Vosniadou & Ortony, 1989; Zeitoun, 1984). Although research on instructional analogies has revealed mixed results in relation to facilitating understanding of target concepts, one potential reason for inconsistent effects could be because of individual differences in spatial visualization skills. The question asked in these experiments is whether an analogy might be particularly useful for learners lacking spatial visualization skills, who generally struggle the most with constructing mental models from scientific texts. To the extent that analogies improve understanding by providing readers with a provisional model of the target construct, low-spatial students might show the most benefit from an analogical comparison. Two experiments tested whether individual differences in spatial skills may impact the effectiveness of learning by analogy, and whether the way in which the analogical comparison is presented matters.

Spatial Skills and Science Text Comprehension

A central assumption in research on learning from text is that comprehension requires the reader to construct multiple representations of the information conveyed by the text (Kintsch, 1998). The most relevant level of representation for understanding the complex processes described by many science texts is the mental model or situation model (Gentner & Stevens, 1983; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). The mental model captures what the text is about and supports performance on inference or application questions on tests of comprehension (Donnelly & McDaniel, 1993; Kintsch, 1994; Mayer, 1989a; Wiley, Griffin, & Thiede, 2005). More specifically, this level of representation includes a causal model of the phenomena being described, and generally involves the generation of inferences or connections across multiple pieces of information to explain how or why these phenomena occur (Chi, 2000; Graesser & Bertus, 1998; Millis & Graesser, 1994; Singer & Gagnon, 1999; Wiley & Myers, 2003).

Theories of text comprehension also generally assume that successful comprehension is affected by both characteristics of the reader (prior knowledge and abilities) and affordances or features of a text or its topic (Kintsch, 1988; McNamara, Kintsch, Songer, & Kintsch, 1996; van den Broek, 2010; Wiley, Sanchez, & Jaeger, 2014). For learning in science, individual differences in spatial visualization skills seem to be highly relevant for success (Uttal & Cohen, 2012). For example, performance on spatial measures has

Allison J. Jaeger, Andrew R. Taylor, and Jennifer Wiley, Department of Psychology, University of Illinois at Chicago.

Allison J. Jaeger is now at the Spatial Intelligence Learning Center (SILC), Temple University. Andrew R. Taylor is now at the Division of Public Health Sciences, Henry Ford Health Systems, Detroit, Michigan.

This research was completed at the University of Illinois at Chicago.

Correspondence concerning this article should be addressed to Allison J. Jaeger, Department of Psychology, Temple University, 1701 North 13th Street, Weiss Hall, Philadelphia, PA 19122. E-mail: allison.jaeger@temple.edu
been used to predict success in spatially related industrial jobs (Ghiselli, 1973); course grades in mechanics, art, math, and physics (McGee, 1979); as well as in training courses for pilots and air crew (Guilford & Lacey, 1947). Lord (1987) showed that science majors have higher spatial scores than nonscience majors, and similarly, Siemankowski and MacKnight (1971) demonstrated that science, math, and art majors outperformed nonscience majors on tasks measuring spatial skills. Differences in spatial scores have also been related to performance in specific science domains, including chemistry (Carter, LaRusso, & Bodner, 1987; Staver & Jacks, 1988), engineering (Hsi, Linn, & Bell, 1997), and geosciences (Black, 2005).

It has been suggested that spatial visualization skills contribute to learning for some topics in science because they relate to one’s ability to mentally represent and transform complex objects or entities, as well as to represent the relations between these entities (Hegarty, 2010). For some scientific phenomena, processes, or systems, the construction of accurate mental models would require representation of the configuration of the main components, the spatial relations between components, as well as information about the movements or causal interactions among the components (Gentner & Stevens, 1983; Hegarty & Just, 1993; Mayer, 1989b). For example, to represent an understanding of plate tectonics, the reader needs to create a mental model that accurately represents the layers of the earth and the locations of tectonic plates, as well as information about how those plates move and interact with each other to cause phenomena such as earthquakes (Sanchez & Wiley, 2014, in press).

Challenges of Learners With Low-Spatial Skills

Despite the importance of creating internal representations when reading science texts, many studies have shown that readers fail to generate a coherent mental or causal model of phenomena on their own, and that this may be especially true for students who have weaker spatial skills (Graesser & Olde, 2003; Hegarty, Narayanan, & Freitas, 2002; Wiley & Sanchez, 2010). Because so closely a student’s internal representation matches that of the actual scientific phenomenon is generally a good indicator of their understanding, then it may also be true that the more accurately one can create a mental representation, the better one is able to understand or is capable of understanding the topic. From this perspective, students who struggle to visualize or mentally simulate the processes occurring in a scientific phenomenon may be particularly disadvantaged when it comes to comprehending these topics.

Consistent with this perspective, Mayer and Sims (1994) showed that high-spatial students developed better understanding of how a bicycle tire pump works and how the human respiratory system works from narrated animations than low-spatial students. Höfﬂer and Leutner (2011) also found differences in understanding related to spatial ability. They found that high-spatial individuals developed deeper comprehension of surfactants from a narrated series of pictures than did low-spatial individuals. Similarly, several studies presenting students with texts instead of narrations have found relationships between spatial skills and mental model generation. For example, Hegarty and Just (1993) found that low-spatial individuals performed less well on comprehension questions about a mechanical text on pulley systems, reread more clauses in the text, and relied more heavily on diagrams when attempting to create mental models. Similarly, Narayanan and Hegarty (2002) found that when reading an illustrated text about how a flushing cistern works, high-spatial individuals had higher scores on both causal questions and questions that required being able to explain what may be going on in the system when the machine is not functioning properly.

Potential Support for Comprehension From Analogies

Instructional analogies are often included as a device that is intended to help explain difficult or unfamiliar topics, and to promote inferences and mental model generation (Blanchette & Dunbar, 2002; Gentner & Stevens, 1983; Thagard, 1992; Vosniadou & Ortony, 1989). Analogies may be especially needed when the phenomena being explained are hard to imagine, abstract, or cannot be directly perceived, such as with very-small-scale phenomena like submicroscopic particles, or large-scale phenomena like plate tectonics, or are “invented” theoretical phenomena, like “quarks” (Brown, 1993; Curtis & Reigeluth, 1984; Duit, 1991; Jee et al., 2010; Lawson, 1993; Thiele & Treagust, 1994, 1995). Several studies have examined analogy use in science textbooks (Curtis & Reigeluth, 1984; Glynn, Britton, Semrud-Clikeman, & Muth, 1989; Newton, 2003; Orgill & Bodner, 2006; Thiele & Treagust, 1994, 1995), and students report finding analogy-embedded texts more understandable and interesting (Glynn & Takahashi, 1998; Paris & Glynn, 2004).

In the analogy literature, the familiar domain is called a source or base domain, whereas the difficult or novel domain is called the target domain. It is generally accepted that in order for an analogy to be effective, readers must have a good understanding of the source domain prior to engaging in analogical processing (Curtis & Reigeluth, 1984; Duit, 1991; Gentner & Gentner, 1983; Iding, 1997; Jee et al., 2010; Thagard, 1992; Zeitoun, 1984). Although there are differences between specific theoretical models of analogy (Blanchette & Dunbar, 2002; Gentner, 1983; Holyoak & Thagard, 1989; Vosniadou & Ortony, 1989), all generally recognize comparison, mapping, and inference as the crux of analogical reasoning. Mapping is how knowledge about the source is carried over to the target, and allows to readers to generate inferences and construct a more complete mental model. The goal of an instructional analogy is not just to provide an anchor for understanding by providing the source or base domain, but also to invoke comparison between the source and target domains. A comparison can help the reader to comprehend something difficult or novel by pointing out its similarities to something more familiar.

In the analogy literature, the familiar domain is called a source or base domain, whereas the difficult or novel domain is called the target domain. It is generally accepted that in order for an analogy to be effective, readers must have a good understanding of the source domain prior to engaging in analogical processing (Curtis & Reigeluth, 1984; Duit, 1991; Gentner & Gentner, 1983; Iding, 1997; Jee et al., 2010; Thagard, 1992; Zeitoun, 1984). Although there are differences between specific theoretical models of analogy (Blanchette & Dunbar, 2002; Gentner, 1983; Holyoak & Thagard, 1989; Vosniadou & Ortony, 1989), all generally recognize comparison, mapping, and inference as the crux of analogical reasoning. Mapping is how knowledge about the source is carried over to the target, and allows to readers to generate inferences and construct a more complete mental model. The goal of an instructional analogy is not just to provide an anchor for understanding by providing the source or base domain, but also to invoke comparison between the source and target domains. A comparison can help the reader to comprehend something difficult or novel by pointing out its similarities to something more familiar.
and easy to understand (Kurtz, Miao, & Gentner, 2001). A popular example of an analogy in science is how the camera can be used to teach how the human eye works. The camera and the human eye can be analogous in several ways: The pupil is similar to the shutter of the camera because it controls how much light gets in, and the retina is similar to the film because it is responsible for sensing and capturing the image. Another popular analogy in science is comparing the structure of the atom to the solar system. In this comparison, the nucleus of the atom is like the sun at the center of the solar system and the electrons surrounding the nucleus are like the planets orbiting around the sun. A successful comparison process can be seen as occurring in three steps. In the first step, the reader retrieves prior knowledge about the source domain. In the second step, the reader compares the source domain with the target domain as a function of the relations that they share, and attempts to create an alignment across the two domains. To do this, the reader must identify the relations that are common to both the source and target, and make the proper correspondences between the two. In the third step, the reader utilizes these correspondences between the source and target to make inferences about the new domain, which become part of their situation model or understanding of the phenomena. Thus, completing this process improves students’ comprehension of the target by helping them to generate inferences and construct a more complete mental model of the target.

Subsumed within this theory is the need for readers to be able to differentiate analogs from irrelevant attributes as they attempt to transfer knowledge from the source to the target. However, one of the main issues with analogies is that students often do not select the correct features to transfer (Day & Goldstone, 2011; Halpern, Hansen, & Rieber, 1990; Ieding, 1997; Thagard, 1992). When analogs are concrete, familiar, and easily visualizable, individuals are able to take advantage of the intuitive causal relationships between physical objects in order to understand more complex or less transparent situations (Brown, 1993; Curtis & Reigeluth, 1984; Day & Goldstone, 2011; Mayer, 1989b; Zeitoun, 1984). Concrete similarities between the objects, parts, and attributes of the source and target that support the desired relational match can help the student align the two examples. Put another way, an analogy can serve as an early or provisional model for the target phenomena, which provides a basis for the generation of inferences and mental model construction, and in turn improves student understanding (Glynn & Takahashi, 1998; Sibley, 2009; Wilbers & Duit, 2006).

Although analogies are common in practice, and should facilitate learning in theory, empirical research on the effectiveness of instructional analogies has gained mixed results (Dagher, 1995; Duit, 1991; Orgill & Bodner, 2006). Some studies have shown that receiving an analogy can facilitate better memory for information (Glynn & Takahashi, 1998; Paris & Glynn, 2004; Royer & Cable, 1975, 1976; Vosniadou & Schommer, 1988). Other studies have been able to show better understanding of target concepts because of the presence of an analogical comparison (Bean, Searles, Singer, & Cowen, 1990; Chiu & Lin, 2005; Donnelly & McDaniel, 1993; Kurtz et al., 2001; Matlen, Vosniadou, Jee, & Ptochukina, 2011; Yanowitz, 2001). Although this research has been able to provide support for the conclusion that including analogies in science texts can benefit learning, most of these studies have used materials or instructions that are more involved than simply presenting an analogy as part of a textbook passage. For example, several studies have found that being given multiple source analogies instead of only a single analogical example can improve learning of the target concept. Kurtz et al. (2001) demonstrated benefits from comparing two possible analogies for heat flow. Similarly, Chiu and Lin (2005) either gave students who were learning about electricity a single water-flow analogy, or two similar or two complementary analogies. In this study, analogies were signaled in special font and students were specifically directed to attend to them. Overall, students who received any analogy performed better than students who received no analogy, but best performance on comprehension and integration questions was seen with two complementary analogies. Likewise, Bean, Searles, and Cowen (1990) and Matlen et al. (2011) have found benefits from combining visual and verbal analogies. Matlen et al. presented students with an analogy-enhanced text on plate tectonics. Accompanying this text was either a set of pictures depicting both the source and target of each analogy (visual analogy condition) or pictures of only the target concepts of each analogy (target picture condition). Results indicated that students in the visual analogies condition performed better than students in the target-picture-only condition on both an immediate and delayed comprehension test. Both Glynn and Takahashi (1998) and Paris and Glynn (2004) also found a benefit of analogies, but these analogies were elaborated (articulating multiple points of comparison) and presented with pictorial illustrations. The analogies used by Yanowitz (2001) also articulated multiple mappings between source and target. Further, similar to the instruction in Chiu and Lin (2005), the benefits seen in Donnelly and McDaniel (1993) on inference questions were in the context of an explicit instruction of how to use the analogy for learning. Across all of these studies, researchers have been able to find benefits from an “analogy” condition; however, most of these studies employed more complex manipulations, instructions, or materials than typically provided to students by textbook passages.

Alternatively, a substantial number of articles have either reported inconsistent effects on comprehension outcomes across studies or measures (Alexander & Kulikowich, 1991; Ieding, 1993; McDaniel & Donnelly, 1996; Simons, 1984), failed to find facilitative effects (Bean, Searles, & Cowen, 1990; Braasch & Goldman, 2010; Gilbert, 1989), or demonstrated comprehension benefits only under specific conditions (Brown, 1992; Halpern et al., 1990; Mayer & Bromage, 1980). Mayer and Bromage (1980) found greater benefits in conceptual understanding from an analogy presented at the start of a text about FORTRAN than at the end. Brown (1992) compared understanding of force from a text that included several isolated concrete examples versus several concrete examples chained together to provide an explanation (bridging from the concrete to the abstract). Students developed better understanding from the bridging explanation text than from the text that just contained concrete examples. Halpern et al. (1990) found benefits only from distant analogies (near analogies did not improve understanding beyond the no-analogy condition). Still other studies have shown detrimental effects of analogies on understanding of the target concept (Donnelly & McDaniel, 2000; Jaeger & Wiley, 2015), including that analogical examples can sometimes cause misconceptions about the target domain (Spiro, Feltovich, Coulson, & Anderson, 1989; Zook & di Vesta, 1991). Donnelly and McDaniel (2000) suggested that analogies can be detrimental particularly for students who are already familiar with the target domain (or have more science background; Donnelly & McDaniel, 1993). Spiro et al. (1989) found that students can
overextend information when using an analogy to learn a new concept. Likewise, an analogy condition can cause more misconceptions or inappropriate inferences compared with a no-analogy control (Zook & di Vesta, 1991).

One potential reason for the inconsistent effects of analogies on comprehension outcomes could be because of individual differences in spatial visualization skills and their interaction with the instructional context. When no general effect of analogy is found, there could still be a benefit more specific to low-spatial readers who need the additional support for inference generation and mental model construction that the familiar analogy provides. Thus, the present studies aimed to investigate whether differential benefits of providing an instructional analogy could be found for students with varying levels of spatial visualization skills.

Effects Related to How Analogies Are Presented

Another factor that could be responsible for the lack of consistent benefits from instructional analogies is that the presentation of analogies can differ in many ways. Curtis and Reigeluth (1984) identified several dimensions on which analogies used in textbooks vary including their degree of elaboration (simple to enriched to elaborated), the basis of the comparison (structural or functional), and the relative abstraction of the source and target (concrete to abstract vs. concrete to concrete, etc.).

In addition, they also noted that analogies can differ in terms of their placement with respect to target information: appearing before the target content, after the target content, or as an "embedded activator" with the analogical comparison interleaved with target content. (In a subsequent analysis, Thiele and Treagust (1994) also added a fourth option for placement in the margins.) Although analogies do seem to be used as a preface or in the margins some portion of the time, the most common way that analogies are presented in science textbooks is in the embedded or interleaved format (Curtis & Reigeluth, 1984; Orgill & Bodner, 2006; Thiele & Treagust, 1994). In fact, a review of elementary level science textbooks (Newton, 2003) found that all of the analogies were presented in this way. Newton (2003) suggests that this presentational format is especially beneficial for younger and lower ability students because it provides a clearer and more concrete mapping from the known to the unknown at the time that it is needed. Specifically, embedded analogies are generally inserted into the materials at points at which new or difficult information is being introduced. By presenting them in this way, they can help to clarify what should be mapped to the new and novel information. This suggests that when low-spatial students are provided with a concrete or familiar analogy, and the mapping is made clear via interleaved presentation, it could be predicted that they may be able to use the familiar domain to begin to form a mental model of the new, novel domain.

Although observed less frequently in the textbook analyses, there is also reason to suspect that analogies presented before target content could improve learning because they may serve as advance organizers (Mayer & Bromage, 1980; Royer & Cable, 1975, 1976). Originally defined by Ausubel (1968), advance organizers can be defined as relevant introductory materials that are presented in advance of the main learning material and tend to be at a higher level of abstraction and generality. Ausubel describes two different functions of advance organizers: expository organizers, which are meant to provide relevant background knowledge, and comparative organizers, which are meant to be used to point to relations between some familiar information and the new, to-be-learned information. Although perhaps not as abstract as suggested by Ausubel’s original definition, analogies presented prior to the target concept can be argued to function as comparative advance organizers (cf. Mayer & Bromage, 1980). Ausubel (1960) suggests that advance organizers benefit learning and retention because they allow for readers to have a structure on which to “anchor” the incoming information. A meta-analysis conducted by Luiten, Ames, and Ackerson (1980) showed that they do have an overall facilitative effect on both learning and retention; however, the results also suggested that individual differences in ability could impact their effectiveness. Specifically, the meta-analysis indicated that advance organizers were more effective for high-ability students. One can imagine that presenting the source analogy at the start of a passage could make processing more difficult both because the analogy needs to be remembered and then applied when the new material is encountered. Because it does not make the mapping as clear as via interleaved presentation, it could be predicted that analogies placed before target material may be less facilitative for low-spatial readers.

To address these hypotheses, the two studies reported here assessed individual differences in spatial skills and examined their relation to learning from an analogy-enhanced text in which the analogy either appeared before the target material, or interleaved throughout the target material, or did not appear. Because low-spatial individuals may struggle to create mental models on their own, these individuals should benefit most from interleaved analogies that provide the most possible support and opportunity for mapping. Although providing an analogy before target material may provide low-spatial individuals with more opportunity for mapping and mental model generation than no analogy would, this format still requires the student to determine what pieces of information from the base are important for understanding the target and where those pieces of information should be mapped on. Consequently, it is hypothesized that interleaved analogies will be the most beneficial placement for low-spatial individuals. High-spatial individuals, on the other hand, do not require such extensive support for mapping and mental model generation. High-spatial individuals should be able to develop a mental model of the information regardless of whether and where an analogy is presented. Based on this idea, it is hypothesized that high-spatial individuals will show no differences in learning across analogy presentations. That is, whether the analogy is presented in an interleaved fashion, before the target material, or not presented at all will not impact high-spatial individuals’ learning because they should possess the skills necessary to mentally represent the complex scientific phenomena on their own. An interesting alternative hypothesis for high-spatial individuals is that they may in fact be harmed by the presence of an analogy if the representation they can create on their own is better than the representation facilitated by the analogy.

Experiment 1

The main goal of Experiment 1 was to test whether individual differences in spatial visualization skills and the position of an analogy within a science text (analogy before, interleaved analogy, or no analogy) would interact and affect understanding of a complex scientific phenomenon. All readers received the same base text describing the El Niño weather phenomenon in the Pacific Ocean. Participants in the analogy conditions also received a short
analogy describing air movement patterns in the context of inflating and deflating a balloon. Key measures of learning and mapping were collected following reading, and spatial skills were also assessed. The main prediction that was tested was whether effects from the placement of analogies and individual differences in spatial skills would interact such that low-spatial individuals would benefit most from the interleaved analogy condition.

Method

Participants. Seventy-two students at a large Midwestern university participated as part of a subject pool requirement for introductory psychology. Participants were randomly assigned to one of three conditions (n = 24 per condition): Analogy before, interleaved analogy, and no analogy. As shown in Table 1, conditions did not differ by age, distribution of gender or bilingual status, paper folding scores, total number of science courses, ratings for prior knowledge, interest or ease of reading, or the amount of time spent reading the text. El Niño was chosen as a topic because students in the same introductory psychology course in a prior semester were asked to rate how much they knew about the El Niño weather pattern on a 1-to-10 Likert scale (1 = not much, 10 = very much) during mass-testing at the start of the semester. This pilot data indicated that prior knowledge of El Niño was quite low in this population (self-report ratings, M = 2.51, SD = 2.13). Also, performance on the concept verification task (CVT) administered to a sample that received no instruction on this topic showed performance was at chance (M = .50, SD = .10).

Materials.

El Niño text. All participants read the same 1,366 word base text on El Niño weather patterns in the Pacific Ocean (presented in Appendix A). The analogy-embedded versions of the text included 15 additional sentences (305 words) that described the movement of air while inflating and deflating a balloon (represented by the bolded text in Appendix A). The location of the analogical information in the interleaved version was based on locations in which it most closely mapped on to the target domain. Locations for the interleaved analogy information were similar to those used by Braasch and Goldman (2010), who also used an El Niño text, but included a tire analogy instead of the balloon analogy. A pilot study suggested that a balloon analogy would be more effective than the tire analogy because students were more familiar with it. Sentences about the analogy were inserted at three locations in which the concept was explained abstractly and where the concept was explained directionally (East, West) in terms of the Pacific Ocean. The first portion of the analogy appeared after the initial description of air pressure movement around the Earth (movement from areas of high to low pressure, creating trade winds), and described how the air pressure inside a blown-up balloon is high and the pressure outside the balloon is low. The second portion of the analogy appeared after the description of surface water movement related to trade winds and described how the movement of air from inside the balloon to outside creates a stream of air, which pushes the dangling piece of paper. The final section of the analogy appeared after the initial description of the El Niño phenomenon. Specifically, in El Niño, the pressure gradient is weakened and the trade winds cease allowing large bulges of water to flow back to their original locations. The portion of the analogy following this information described a deflated balloon that is no longer creating a stream of air, allowing the piece of paper to slowly fall back to its original location. These locations were selected in order to provide the most support for the mappings between the target topic of El Niño and the balloon analogue for the reader. For the analogy-before condition, all of the balloon analogy sentences were inserted as a block at the start of the base text. The no-analogy condition presented only the base text.

Essay task. Participants were asked to write an essay on paper in response to this prompt: “What are the causes of typical weather patterns across the Pacific Ocean, and what is different about these patterns during the El Niño phenomenon?” The essay served as one measure of understanding.

CVT. As a second measure of understanding, an 18-item CVT (based on the measure used by Braasch & Goldman, 2010) asked participants whether statements seemed true or false based on the information they read. These statements represented conceptual connections and connections between ideas that were not made explicitly in the text, such as “Under typical weather conditions, areas with higher air pressure experience cooler air temperatures” (TRUE) and “The western Pacific Ocean experiences less upwelling during an El Niño phase” (FALSE). There were a total of nine true items and nine false items, and participants’ scores were computed as the percent correct (see Appendix C).

The verification technique, pioneered by Royer and his colleagues, has been used in many studies as a measure of comprehension. Reliability of verification tests has been demonstrated in several studies (Greene, Royer, & Anzalone, 1990; Marchant,

Table 1

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>No analogy M (SD)</th>
<th>Analogy before M (SD)</th>
<th>Interleaved M (SD)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.58 (4.48)</td>
<td>19.00 (1.09)</td>
<td>18.92 (1.14)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>54</td>
<td>70</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Bilingual (% bilingual)</td>
<td>50</td>
<td>65</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Paper folding score</td>
<td>.60 (.21)</td>
<td>.65 (.16)</td>
<td>.59 (.17)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td># Science courses</td>
<td>1.64 (1.94)</td>
<td>2.13 (2.13)</td>
<td>1.90 (2.86)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Prior knowledge rating</td>
<td>2.04 (1.73)</td>
<td>2.09 (1.91)</td>
<td>1.67 (1.17)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Interest in topic rating</td>
<td>5.25 (3.22)</td>
<td>5.48 (2.64)</td>
<td>4.96 (2.01)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Ease of reading rating</td>
<td>6.75 (2.40)</td>
<td>6.43 (1.90)</td>
<td>6.92 (1.91)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Reading time (seconds)</td>
<td>555 (138)</td>
<td>531 (175)</td>
<td>585 (222)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>9.79 (6.80)</td>
<td>9.25 (4.27)</td>
<td>9.54 (5.25)</td>
<td>F &lt; 1</td>
</tr>
</tbody>
</table>
Royer, & Greene, 1988; Royer, Sinatra, Greene, & Tirre, 1989). Although many more studies have used verification measures, reliabilities of these comprehension measures have generally not been reported, especially in studies investigating learning from a single topic or passage. In general, when Cronbach’s alpha has been reported, verification tests based on three or fewer passages (with approximately 16 sentences per passage) often have reliabilities in the .5 to .6 range. As the number of passages increases, so does internal reliability of the tests. Verification tests based on four passages generally have reliabilities in the .7 to .8 range, and tests based on six passages or more have reliabilities in the .8 to .9 range. Lower alphas can be expected when items come from a single passage, as test creators intentionally try to create tests with items that assess understanding of different parts of a passage to test for breadth of understanding for the topic, rather than testing only one single construct which would lead to higher internal reliability. In the present study, the goal was for the 18 items to assess understanding of many different aspects of the El Niño system, not to create 18 items that were all testing the same idea. The Cronbach’s alpha for the 18-item test was .60. The modest internal reliability for this test can be interpreted as reflecting the fact that the test strove for coverage of the content.

Instead of using internal reliability as a basis for evaluation, measurement quality of verification tests has more typically been demonstrated through the relation of verification task performance to other measures, including individual differences in reading ability or comprehension skill, differences in performance because of instruction, or the correlation of verification scores with other measures of comprehension (Braasch & Goldman, 2010; Griffin, Wiley, & Thiede, 2008; Royer, Carlo, Dufrense, & Mestre, 1996; Royer, Lynch, Hambleton, & Bulgareli, 1984; Royer, Marchant, Sinatra, & Lovejoy, 1990; Sanchez & Wiley, 2006, 2010; Wiley et al., 2009; Wiley & Voss, 1999). Importantly, performance on the verification task was well aligned with essay scores in this study, \( r = .60, p < .0001 \), suggesting they were both capturing aspects of student understanding about El Niño.

**Mapping task.** A mapping task was developed as a third measure for readers in the analogy-embedded conditions. Students were asked, “In the text you were told that the balloon analogy would be helpful for learning about weather patterns. What connections did you notice between the balloon analogy and weather patterns?”

**Paper folding test.** Each student completed a computerized version of the Paper-Folding Test from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). In this test, the participant must determine which one of five possible patterns of holes will result after a piece of square paper goes through a sequence of folds and then is punched. Since the 1940s, this test has commonly been used as a measure of spatial visualization skill, which, broadly defined, represents one’s ability to mentally transform or manipulate objects (Carroll, 1993). The test consisted of two parts, each containing 10 items, presented one at a time. Participants had 3 min to complete each part. Participants’ scores were the proportion of correct responses out of those attempted across both parts. This task was selected as the measure of spatial visualization skill because it has been demonstrated to be a strong predictor of performance or aptitude in STEM areas (Höffler & Leutner, 2011; Hsi et al., 1997; Lord, 1987; Mayer & Sims, 1994; Siemankowski & MacKnight, 1971), and more specifically, when learning from science text (Narayanan & Hegarty, 2002; Sanchez, 2012; Sanchez & Wiley, 2010). Split-half reliability on this measure was .85 in this sample.

**Final survey.** All students completed a paper-and-pencil final survey. This survey asked them to provide ratings on scales of 1 to 10, reporting how much they knew about El Niño before reading the text (1 = not much, 10 = very much), how easy they found the text to read (1 = not at all, 10 = very much), as well as how interesting they found the text on El Niño to be (1 = not at all interesting, 10 = very interesting). The final survey also asked participants to report basic demographic information, including gender, age, whether they are bilingual, and the number of science courses taken.

**Procedure.** The experimental sessions were run in groups of one to five participants. Participants first read a set of instructions stating that they would be asked to later write an essay on the text they were about to read. They were also informed of the exact essay prompt they would be given. Participants then read the text on an IBM-compatible PC on E-Prime 2.0, moving through the text at their own pace by pressing the spacebar. Once participants finished reading, they were asked to write their essays without access to the text. After completing the essay, participants returned to the e-prime program to complete the CVT, and then the paper folding test. After the paper folding task, participants in the analogy conditions completed the mapping task. Finally, all students completed the final survey. No portion of the experiment except the paper folding task was timed. Participants generally completed the study in an hour, but sometimes went longer.

**Essay coding.** The essay coding scheme gave credit for the inclusion of six primary causal concepts that form the basis of an accurate mental model of both normal weather patterns in the Pacific and the changes that occur during El Niño:

1. Air moves from areas of high to low pressure.
2. Air movement drags warm surface ocean water with it.
3. Displacement of warm surface ocean water causes upwelling.
4. Lower air pressure brings an influx of warmer surface ocean water, higher air temperatures, more evaporation, more condensation.
5. El Niño reverses the pressure gradient/trade winds/ocean water movement.
6. Weather patterns are part of a larger circulation system that goes through cycles.

These concepts were scored as either present or missing, with essay score representing the total number of these concepts mentioned in each essay. Two coders scored all essays with interrater reliability (Krippendorff’s alpha) of .94 (Krippendorff, 2013).

**Mapping task coding.** There were a total of five possible points in the mapping task, which represented the five major components of the balloon analogy. These included connecting the weather pattern and El Niño information to the inside of the balloon, the outside of the balloon, air moving from inside the
balloon to outside of the balloon, the piece of paper, and the deflated balloon. Participants received a point if their response explicitly made a connection between a part of the balloon analogy and weather patterns. For example, if a participant stated that the air inside the balloon represented the high-pressure East Pacific, they received a point. However, they would not have received a point for simply stating that the inside of the balloon is high pressure. Mapping scores were computed as a total number of points. Two coders scored all responses with interrater reliability (Krippendorff’s alpha) of $\alpha = .88$.

All three dependent measures were correlated with each other (essay and CVT scores: $r = .60$, $p < .001$; essay and mapping scores: $r = .33$, $p < .03$; CVT and mapping scores: $r = .31$, $p < .04$), demonstrating convergent construct validity.

### Results

#### Essay performance.

On average, participants wrote 9.53 ($SD = 5.46$) sentences in their explanations of El Niño, and included an average of 2.76 ($SD = 1.67$) of the six main concepts. Other ideas that were mentioned in the essays included details from the text on how El Niño affects the anchovy crop, and how it causes bad weather and flooding, economic and social problems, and death. A little over half of the participants (58%) included at least one of these ideas in their essays, but this did not vary by condition or spatial skills, nor did it relate to any of the dependent measures. Only four participants mentioned the balloon analogy in their essay. This also did not relate to either spatial skills or performance on any of the dependent measures. Additionally, there were five participants whose essays indicated confusion about the location of the countries mentioned in the text and where the East Pacific and West Pacific were located. This directional confusion was also not related to spatial skills or performance on any of the dependent measures.

The effects of analogy placement and spatial skills on the number of six main concepts included in the essays were analyzed using linear regression. The no-analogy condition was used as the referent condition, with dummy codes to test for the effect of the interleaved analogy, and the effect of placing the analogy before the target information, and their interactions with spatial skill (see Table 2 for paper folding scores made more connections in the analogy-before condition than the interleaved analogy condition, whereas participants at $+1 SD$ did better on the CVT in the no-analogy condition than the interleaved analogy condition.

#### CVT performance.

A one-sample $t$-test testing CVT scores against chance (50%) demonstrated that performance was significantly above chance on this measure ($M = .56$, $SD = .16$, $t(71) = 2.92$, $p < .01$). Parallel regression analyses to those used for essay scores were conducted to test for the effects of analogy placement and spatial skills on CVT scores. Paper folding scores ($\beta = .29$, $p < .02$) and interleaving ($\beta = .41$, $p < .03$) both significantly predicted verification task scores. Presenting the analogy before the target information marginally predicted verification task score ($\beta = .36$, $p < .08$). There was also a significant interaction in the relation of performance to spatial skills between the interleaved analogy condition and the no-analogy condition ($\beta = .44$, $p < .02$), but no interaction in the relation of performance to spatial skills between the analogy-before condition and the no-analogy condition ($\beta = .25$, $ns$). As shown in Figure 2, participants at $-1 SD$ in paper folding scores did better on the CVT in the interleaved analogy condition than the no-analogy condition, whereas participants at $+1 SD$ did better on the CVT in the no-analogy condition than the interleaved analogy condition.

#### Mapping task.

A third analysis was conducted to test for the effects of analogy placement and spatial skills on the ability to articulate the connections they saw between the balloon example and El Niño. On average, participants listed only .46 ($SD = .87$) connections. Because the mapping measure could only be collected in the analogy conditions, the analogy-before condition was used as a referent in this analysis. Paper folding scores ($\beta = .38$, $p < .08$) and interleaving ($\beta = .40$, $p < .07$) only marginally predicted mapping scores. However, there was a significant interaction in the relation of performance to spatial skills between the interleaved analogy condition and the analogy-before condition ($\beta = .53$, $p < .04$). As shown in Figure 3, participants at $-2 SD$ in paper folding scores made more connections in the interleaved analogy condition than the analogy-before condition, whereas participants at $+2 SD$ made more connections in the analogy-before condition than the interleaved analogy condition.

### Table 2

**Performance on Learning Outcomes as a Function of Spatial Skill and Analogy Condition in Experiment 1**

<table>
<thead>
<tr>
<th>Learning measures</th>
<th>No analogy</th>
<th>Analogy before</th>
<th>Interleaved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High ($n = 11$)</td>
<td>Low ($n = 13$)</td>
<td>High ($n = 15$)</td>
</tr>
<tr>
<td>Essay</td>
<td>$3.55 (1.51)$</td>
<td>$1.38 (1.26)$</td>
<td>$3.40 (1.50)$</td>
</tr>
<tr>
<td>CVT</td>
<td>$.65 (.21)$</td>
<td>$.50 (.12)$</td>
<td>$.59 (.43)$</td>
</tr>
<tr>
<td>Mapping</td>
<td>—</td>
<td>—</td>
<td>$.67 (1.04)$</td>
</tr>
</tbody>
</table>

*Note.* High and low spatial groups determined by median splits. CVT = concept verification task.
Discussion

The main goal of Experiment 1 was to explore how the presentation of an analogy within a science text may interact with spatial skills to influence learning about El Niño. In general, spatial skills were found to lead to better understanding of El Niño from the text as demonstrated on both essay and verification task measures, consistent with prior work that has shown relations between spatial skills and the generation of mental models about scientific systems (e.g., Hegarty & Just, 1993; Narayanan & Hegarty, 2002; Sanchez, 2012; Sanchez & Wiley, 2014). The most important finding of the study was a consistent interaction that was seen among spatial skills and analogy placement. Interleaved analogies were especially beneficial for low-spatial individuals. Interleaving facilitated the mapping to the source domain, which helped with understanding and building mental models of the target domain, and decreased reliance on spatial skills. At the same time, high-spatial learners showed a better ability to articulate the mapping between the balloon analogy and El Niño when they were forced to engage in the constructive mapping process on their own (in the analogy-before condition) versus when the mapping was made more transparent by the interleaved condition. This finding is reminiscent of other work that has shown that when readers possess the abilities needed to engage in active construction processes, they are better off when they are not given additional supports that other readers benefit from (Hegarty & Kriz, 2008; McNamara et al., 1996; Voss & Silfies, 1996). It also seems consistent with Donnelly and McDaniel (2000), who demonstrated that literal conditions led to better understanding when target concepts were familiar, and Donnelly and McDaniel (1993), who suggested that literal presentation was better for students with more advanced science backgrounds.

Experiment 2

The main goal of Experiment 2 was to test whether the patterns found in Experiment 1 would replicate. A few minor changes were made in the materials. To try to minimize distraction while reading, several sections of the text that contained nonessential or seductive details were removed. To reduce East–West confusions, participants were provided with a simple map of the Pacific showing the locations of Australia, Japan, and Peru, and indicating East and West. To help students focus on their main goals for comprehension, the essay prompt task instruction was revised in order to specifically cue the use of materials from the text in the response, by adding, “Use the information from the text, specifically about the differences between the Eastern and Western Pacific and the consequences of those differences, to answer this question.” The mapping task was also revised to encourage more complete responses.
Method

Participants. Seventy-two students at a large Midwestern university participated as part of a subject pool requirement for introductory psychology. Participants were randomly assigned to one of three conditions (n = 24 per condition): Analogy before, interleaved analogy, and no analogy. As shown in Table 3, conditions did not differ by age, distribution of gender or bilingual status, paper folding scores, total number of science courses, ratings for prior knowledge, interest or ease of reading. The conditions did differ in the amount of time spent reading the text, with the analogy conditions taking longer.

Materials.

El Niño text. The base text was slightly revised to remove some nonessential details (1,064 words, with deletions indicated in italics in Appendix A), and was accompanied by a map (shown in Appendix B) to help support better understanding of East and West. The placement of the analogy in the interleaved condition was not impacted by these changes.

Essay task. Participants were asked to write an essay on paper in response to the revised prompt,

What are the causes of typical weather patterns across the Pacific Ocean, and what is different about these patterns during the El Niño phenomenon? Use the information from the text, specifically about the differences between the Eastern and Western Pacific and the consequences of those differences, to answer this question.

CVT. The CVT remained the same as in Experiment 1.

Mapping task. The mapping task was revised slightly to encourage more complete responses. In addition to the prompt “In the text you were told that the balloon analogy would be helpful for learning about weather patterns. What connections did you notice between the balloon analogy and weather patterns?,” students were further scaffolded with

Use the structure of the balloon analogy to fill in what aspect of weather patterns corresponds with each:

Inside of the balloon →
Outside of the balloon →
Movement of air from inside to outside of the balloon →
Piece of paper →
Deflated balloon →

Table 3
Descriptive Measures for Experiment 2

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>No analogy M (SD)</th>
<th>Analogy before M (SD)</th>
<th>Interleaved M (SD)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.17 (1.34)</td>
<td>18.79 (1.97)</td>
<td>18.83 (1.13)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>65</td>
<td>63</td>
<td>71</td>
<td>X² &lt; 1</td>
</tr>
<tr>
<td>Bilingual (% bilingual)</td>
<td>77</td>
<td>73</td>
<td>71</td>
<td>X² &lt; 1</td>
</tr>
<tr>
<td>Paper folding score</td>
<td>.60 (.21)</td>
<td>.58 (.22)</td>
<td>.56 (.21)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td># Science courses</td>
<td>2.77 (2.35)</td>
<td>2.39 (2.81)</td>
<td>2.19 (1.54)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Prior knowledge rating</td>
<td>3.26 (2.13)</td>
<td>3.67 (2.78)</td>
<td>2.58 (2.25)</td>
<td>F &lt; 1.25</td>
</tr>
<tr>
<td>Interest in topic rating</td>
<td>5.00 (2.24)</td>
<td>4.96 (2.60)</td>
<td>4.92 (3.05)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Ease of reading rating</td>
<td>6.70 (2.44)</td>
<td>6.71 (1.94)</td>
<td>7.42 (1.67)</td>
<td>F &lt; 1</td>
</tr>
<tr>
<td>Reading time (seconds)</td>
<td>407 (122)</td>
<td>552 (233)</td>
<td>540 (136)</td>
<td>F(1, 68) = 4.10, p &lt; .02</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>7.71 (4.05)</td>
<td>7.96 (3.99)</td>
<td>8.50 (3.62)</td>
<td>F &lt; 1</td>
</tr>
</tbody>
</table>

Paper folding test. Because of time constraints, participants completed only part one (10 items) of the paper folding test in the same fashion as Experiment 1. Split-half reliability was .80 on this measure in this sample.

Final survey. The final survey remained the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1. Participants generally completed the study in under an hour.

Essay and mapping coding. The essay coding followed the same scheme as in Experiment 1. Two coders scored all essays with interrater reliability of (Krippendorf’s alpha) α = .95. The mapping task coding also followed the same scheme as in Experiment 1. Two coders scored all responses with interrater reliability of (Krippendorf’s alpha) α = .91. All three measures in this study were well aligned (essay and CVT scores: r = .57, p < .001; essay and mapping scores: r = .35, p < .01; CVT and mapping scores: r = .39, p < .01), demonstrating convergent construct validity.

Appendix A

Table 4

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.17 (1.34)</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>65</td>
</tr>
<tr>
<td>Bilingual (% bilingual)</td>
<td>77</td>
</tr>
<tr>
<td>Paper folding score</td>
<td>.60 (.21)</td>
</tr>
<tr>
<td># Science courses</td>
<td>2.77 (2.35)</td>
</tr>
<tr>
<td>Prior knowledge rating</td>
<td>3.26 (2.13)</td>
</tr>
<tr>
<td>Interest in topic rating</td>
<td>5.00 (2.24)</td>
</tr>
<tr>
<td>Ease of reading rating</td>
<td>6.70 (2.44)</td>
</tr>
<tr>
<td>Reading time (seconds)</td>
<td>407 (122)</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>7.71 (4.05)</td>
</tr>
</tbody>
</table>

Appendix B

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
Table 4

Performance on Learning Outcomes as a Function of Spatial Skill and Analogy Condition in Experiment 2

<table>
<thead>
<tr>
<th>Learning measures</th>
<th>No analogy</th>
<th>Analogy before</th>
<th>Interleaved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High (n = 12)</td>
<td>Low (n = 12)</td>
<td>High (n = 12)</td>
</tr>
<tr>
<td>Essay</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td></td>
<td>4.17 (1.11)</td>
<td>2.25 (1.91)</td>
<td>3.08 (1.78)</td>
</tr>
<tr>
<td>CVT</td>
<td>.66 (.14)</td>
<td>.50 (.10)</td>
<td>.60 (.19)</td>
</tr>
<tr>
<td>Mapping</td>
<td>–</td>
<td>–</td>
<td>1.42 (1.73)</td>
</tr>
</tbody>
</table>

Note. High and low spatial groups determined by median splits. CVT = concept verification task.

and the no-analogy condition (β = −.49, p < .01). Presenting the analogy before the target information did not predict essay score (β = −.27, ns), and there was no interaction in the relation of performance to spatial skills between the analogy-before condition and the no-analogy condition (β = .12, ns). As shown in Figure 4, participants at −1 SD in paper folding scores including more concepts in their essays in the interleaved analogy condition than the no-analogy condition, whereas participants at +1 SD included more concepts in their essays in the no-analogy condition than the interleaved analogy condition.

**CVT performance.** A one-sample t test testing CVT scores against chance (50%) demonstrated that performance was significantly above chance on this measure (M = .59, SD = .16), t(71) = 4.94, p < .001. Parallel regression analyses to those used for essay scores were conducted to test for the effects of analogy placement and spatial skills on CVT scores. Interleaving (β = .46, p < .02) significantly predicted verification task scores, whereas the relation between paper folding scores and performance was marginal (β = .19, p < .10). However, there was a significant interaction in the relation of performance to spatial skills between the interleaved analogy condition and the no-analogy condition (β = −.35, p < .04). Presenting the analogy before the target information did not predict verification task score (β = −.15, ns). There was no interaction in the relation of performance to spatial skills between the analogy-before condition and the no-analogy condition. As shown in Figure 5, participants at −1 SD in paper folding scores did better on the CVT in the interleaved analogy condition than the no-analogy condition, whereas no differences were seen among participants with higher spatial scores.

**Mapping task.** A third analysis was conducted to test for the effects of analogy placement and spatial skills on the ability to articulate the connections they saw between the balloon example and El Niño. On average, participants listed .96 (SD = 1.32) connections. Because the mapping measure could only be collected in the analogy conditions, the analogy-before condition was used as a referent in this analysis. There was only a significant interaction in the relation of performance to spatial skills between the interleaved analogy condition and the analogy-before condition (β = −.55, p < .04). As shown in Figure 6, participants at −2 SD in paper folding scores made more connections in the interleaved analogy condition than the analogy-before condition, whereas participants at +1 SD made more connections in the analogy-before condition than the interleaved analogy condition.

**Supplemental analyses collapsing across experiments.** Two supplemental analyses were performed on the data collapsing across both studies to increase power. The first investigated the effects of the analogy-before placement on learning outcomes. The
second attempted to test for unique effects of spatial skills on learning by partialing out a measure of general ability.

First, although the effect of the before-analogy condition on learning failed to reach significance on either learning measure in either experiment, it was possible that an effect might be seen in the combined sample. Although there was still no significant effect of reading an analogy before reading target information on essay scores ($\beta = -.14, p < .30$), the effect approached significance for CVT scores ($\beta = -.23, p < .09$), showing a trend such that the analogy-before condition harmed learning.

Second, because paper folding is a task that often relates strongly with general ability, a final set of analyses were conducted to test whether the interactions with spatial skills would still be seen after partialing out self-reported ACT composite scores as a proxy for general ability. Self-reported ACT scores were obtained from mass-testing data. Scores were missing for seven individuals. The interactions related to differing relationships for spatial skills and essay scores ($\beta = -.46, p < .001$) and CVT scores ($\beta = -.26, p < .05$) between the interleaved analogy and no-analogy conditions were both still significant when this variance was removed.

Discussion

The results of Experiment 2 replicated the findings from Experiment 1 that interleaved analogies were especially beneficial for low-spatial individuals, whereas high-spatial individuals performed less well in the interleaved analogy condition than the no-analogy condition. In addition, an analysis using the combined sample and covarying out ACT composite scores as a proxy for general ability showed that the interactions with spatial skills and interleaved analogies remained for both essay scores and CVT scores. This helps to suggest that there is a unique role for spatial skills (as opposed to more general ability) in contributing to performance differences in learning because of the interleaved analogy.

General Discussion

Overall, the main finding from both studies was that interleaved analogies positively impacted learning for low-spatial individuals, but negatively impacted learning for high-spatial individuals. This finding is important because it suggests that analogies may differ in their utility as a function of individual differences in spatial skills among learners. The results of the current study are well aligned with theories that suggest a role for spatial skills in science learning (Hegarty, 2010; Uttal & Cohen, 2012). As mentioned earlier, previous work has demonstrated that individuals with higher spatial skills can achieve better comprehension of science texts in some contexts, and this superior comprehension is thought to reflect the role that spatial skills play in being able to create an accurate mental model of highly spatial information (Gentner & Stevens, 1983; Hegarty & Just, 1993; Narayanan & Hegarty, 2002; Sanchez & Wiley, in press; Wiley & Sanchez, 2010). In the current set of studies, paper folding scores (either alone or as part of an interaction) were found to predict performance on all of the dependent measures. These results converge with other findings showing relations between spatial skills and learning from expository science texts, and extend those findings into the domain of understanding global weather patterns.

Further, these results indicate that there does not seem to be an optimal placement of an analogy within an expository text that will improve understanding for all students. Although analogies are generally intended as tools to promote mental model construction, the current set of studies did not find an overall benefit for the inclusion of an analogy within a text describing the causes of El Niño. Instead, the current study demonstrated that there may be individual differences that determine who benefits from analogies, and that different modes of presentation may make them more beneficial. Specifically, individuals who may struggle the most with creating their own mental models, in this case, low-spatial individuals, benefitted only from the presence of an interleaved analogy.

One potential explanation for this outcome is that interleaved analogies circumvent the need to actively align relevant information between the base and target examples, which may pose a particular challenge for low-spatial individuals. By providing low-spatial individuals with analogies in interleaved format, this directly supports the alignment and comparison process, allowing the low-spatial individuals to create accurate mental models of the target phenomenon. High-spatial individuals, on the other hand, may have fewer difficulties with creating mental models of the target phenomenon in the first place; therefore, neither the presence nor the placement of an analogy improves their comprehension. In fact, the results of these studies suggest that in some cases, individuals may benefit more from the process of generating mental models on their own (i.e., when they possess the ability to do so), consistent with other research (Donnelly & McDaniel, 1993; Duit, 1991; Hegarty & Križ, 2008; McDaniel & Donnelly, 1996; McNamara et al., 1996; Voss & Silfies, 1996).

Although an interesting pattern of results was obtained in these studies, the findings are necessarily limited by the fact that only a single topic was investigated. Although it is not uncommon for studies on instructional analogies to concentrate on the conditions that affect learning for a single topic (Bean, Searles, & Cowen, 1990; Bean, Searles, Singer, et al., 1990; Braasch & Goldman,
Another limitation of this study was that it used only a college-level population and collected data in a laboratory context, and it is possible that the same results might not obtain in younger populations or more authentic instructional settings. Newton (2003) has suggested that younger students might struggle with developing mental models of complex information because they do not possess the full set of cognitive abilities required to do so. Based on this idea, younger students may benefit from the same conditions as the low-spatial readers studied here, or they may benefit from different conditions compared with older students when learning from analogies. Differences may also occur for studies conducted in actual classroom settings compared with laboratory experiments, or from more extensive analogies, bridging analogies, or multiple-source comparison conditions. Future work taking a more process-focused approach, such as by collecting think-aloud protocols or eye-tracking traces, would also be helpful for understanding how students with different levels of spatial visualization skill actually process text with analogies in different positions (cf. Wiley, Ash, Sanchez, & Jaeger, 2011). This type of data could reveal how the interleaved analogy facilitates learning for low-spatial learners, or identify the challenges they face in comprehension when no analogy is presented. Similarly, trace data could help to show how high-spatial learners are able to learn effectively from the no-analogy text, and how the interleaved-analogy text leads to poorer comprehension.

In contrast to the effects that were observed for interleaved analogies in this study, presenting an analogy before the target information seemed to have much weaker effects. The effect of presenting the analogy at the beginning of the passage was never found to be significantly different than the no-analogy condition, suggesting that it did not improve learning. Indeed, there was a marginal trend in the combined analysis such that presenting an analogy before target information tended to harm learning outcomes for all students. This negative effect may be related a further potential issue with the use of analogical examples in science text, namely, that they may cause illusions of comprehension. Jaeger and Wiley (2015) conducted a study looking at the effect of an analogy on the ability of sixth grade students to accurately monitor their comprehension when learning about the causes of climate change. In their study, they found that the presence of the analogy actually harmed comprehension. However, at the same time, students presented with the analogy were more overconfident about how well they understood the text passage. Although analogies are meant to improve expository science text comprehension, they can introduce problems especially for novice learners who may have a tendency to focus on them as interesting examples, rather than as a way to elaborate on their mental models or to test their understanding of the novel phenomenon. Future research should continue to explore the qualities of analogies that might determine their effectiveness including familiarity, interest, and level of alignment or similarity.

Analogies are, without a doubt, valuable forms of thought. When Kepler investigated the motion of the planets, he envisioned that it could be likened to a clockwork mechanism rather than to a divine organism. When Rutherford was attempting to understand the structure of the atom, he pictured it as a mini solar system. And when Einstein noticed a similarity in mathematical formulae for an ideal gas and a black body, he followed the trail to the suggestion that light might consist of small packets of energy analogous to molecules. Case studies of great scientific thinkers provide clear examples of breakthrough discoveries about invisible or not-readily observable phenomena being because of the ability to think analogically. What is less clear, however, is whether and when presenting an analogy will be beneficial for students’ learning of science concepts.

Although it is widely presumed that providing concrete or familiar analogies to students will help their understanding of novel concepts, the literature offers mixed results. Prior work on providing analogies in text has not fully considered issues such as the way the analogy is presented within the lesson, the extent to which the mapping is signaled, the way learning is measured, and critical characteristics of the readers that may interact with all of these factors. The two experiments presented here attempted to provide a better picture of the role that spatial skills and the presentation of analogies may play in learning from science text. The results highlight that analogies presented in an interleaved fashion can be beneficial particularly for low-spatial learners. These results are consistent with the suggestion that analogies may help learning by providing students with a familiar or concrete situation that can serve as a basis for developing an understanding of the new phenomenon and may reduce the dependence on spatial skills. More work is needed to discover the extent to which these results may generalize, or the extent to which they may be specific, to contexts in which the target content is abstract, difficult, complex, or to unobservable phenomena like plate tectonics, the greenhouse effect, or the weather. It is possible that these sorts of topics may lend themselves to spatial representations, and that may provide an explanation as to why analogies, visualizations, or spatial skills seem so critical for understanding in these scientific domains.

References


Google Maps. (2016). Pacific Ocean Map. Retrieved from https://www.google.com/maps/place/Pacific+Ocean/@1.4348754,-154.1200939,21943291m/data=!3m1!1e3!4m2!3m1!1s0x76ed042c30f318eb:0x8effef1ad0f70876cb

ANALOGY AND SPATIAL SKILLS

13


This document is copyrighted by the American Psychological Association and may not be duplicated, disseminated, or electronically stored in whole or in part without written permission from the publisher except for brief quotations in critical reviews or scholarly analysis.
Appendix A

El Niño Text (bolded text appeared only in analogy conditions; italicized text appeared only in Experiment 1)

**El Niño**

**The Power of El Niño**

One of the most famous and powerful disruptions of the ocean-atmosphere system is the weather phenomenon commonly referred to as El Niño. El Niño drastically affects the weather in the equatorial Pacific Ocean and the resulting changes in weather dramatically disrupt people’s lives. Though the El Niño weather phenomenon originates in a relatively “local” area, when in full swing, it has worldwide economic, ecological, and societal consequences.

El Niño events occur in somewhat predictable cycles every three to seven years and may last for many months. During the past 40 years, 11 El Niño events have been recorded. The worst one occurred in 1997–1998. It lasted for eight months, caused an estimated 33 billion dollars in property damage, and killed approximately 2,100 people.

**How Weather Patterns Develop in the Equatorial Pacific**

Simply put, El Niño is a disruption of typical weather patterns in the equatorial Pacific. Understanding how and why this disruption occurs involves a consideration of the forces involved in weather patterns. Weather patterns are governed by the relationship between movement of air in the atmosphere and movement of water in the ocean.

Movement of Air. The movement of air in the Earth’s atmosphere is dominated by differences in air pressure. Specifically, there are high pressure systems and low pressure systems that interact with each other. The standard relationship is that air moves from areas of high pressure to areas of low pressure. This everyday example is useful for helping you to understand the typical relationships that govern weather patterns. Of course, the air in the earth’s atmosphere is not enclosed within a container, but the general principles are the same in weather systems. Air moves from areas of higher pressure to lower pressure. And, the movement of air in the atmosphere causes the movement of other things, such as water in the ocean.

As the trade winds blow from east to west and push steadily against the sea for thousands of miles, they drag the surface water along. As a result, warm surface water accumulates in the western Pacific. This bulge of warm seawater can become massive, extending out from the coasts of Australia and Indonesia for many thousands of miles. This surface water becomes the warmest ocean water on the entire planet at 75–85 °F.

The Relationship between the Ocean and Rainfall. The temperature of the ocean water has a direct effect on the weather above it. Because warm water evaporates faster than cold water, the bulge of warm water in the western Pacific creates a great amount of evaporation. When water evaporates, it warms the air. Because warm air can hold more moisture than cold air, this allows even more water to evaporate above the bulge.

Just like the ocean, there are different temperature levels in the air. Temperatures closer to the surface are warmer than temperatures high in the atmosphere. The moist air produced by the evaporation of the ocean continues to warm and rise. When this air reaches higher altitudes, it cools. As the air cools it is able to hold less moisture and as a result clouds form. The high rate of evaporation in the western Pacific produces a high rate of rain cloud formation and precipitation becomes more likely. Thus, the large bulge of warm water in the west increases the likelihood of preci-
pitation in that part of the Pacific. Countries such as Malaysia, Indonesia, Australia, and other parts of Southeast Asia typically have very humid and tropical conditions. Conversely, in the eastern Pacific drier conditions are the norm. Countries such as Peru and Chile in South America typically experience far less rainfall than those in the western Pacific.

The process of upwelling is a major reason why the eastern Pacific has cooler surface water and air temperatures than the western Pacific. As the trade winds move surface water to the west, cold, nutrient-rich waters are pulled up from the depths of the ocean to take the place of the water that has moved to the west. As a result there is a constant renewal of surface water with cooler water. This is the process called upwelling. Eventually however this surface water will meet the same fate as the water it replaced. It will also get heated and will be pushed toward the west by the trade winds. In the eastern Pacific cold water is closer to the surface than it is in the western Pacific. Together the cooler surface water and the cooler air inhibit the evaporation of moisture into the air and it is less likely to rain. This is why the Galapagos Islands and the coast of Peru typically experience drought conditions.

The airflow patterns over the equatorial Pacific Ocean are best thought of as a cyclic process. The constant evaporation of the warm water in the west creates a vacuum-like state. As the hot air rises, the trade winds are sucked toward the west. The “loop” is completed by a returning flow of air at higher altitudes from west to east. This circular pattern of air movement is referred to as convection and, in the case of the Pacific Ocean, is specifically referred to as the Walker Circulation. The Walker Circulation feeds the trade winds. In brief, the convection cycle helps explain the climate variations between the eastern and western Pacific.

**El Niño: A Disruption in Typical Weather Processes**

The El Niño weather phenomenon is a disruption in typical weather patterns. An El Niño event is triggered when the air pressure gradient weakens. The change in the gradient is due to both a decrease in air pressure the eastern equatorial Pacific and a simultaneous increase in the western equatorial Pacific. When the air pressure gradient changes, so does the strength of the trade winds.

If you think about the deflating balloon once again, you can imagine that the air coming out of the balloon will slow down over time. This is because the air pressure inside and outside the balloon are becoming more similar. Now, with no stream of air pushing against it, the paper will swing back toward the balloon.

When the trade winds decrease in strength, they can no longer keep the massive bulge of warm water in the western Pacific. As a result, the mass of warm water rushes back eastward toward the central or eastern Pacific. Accordingly, the surface water of the eastern Pacific is no longer being pulled west and upwelling is suppressed. This produces a massive pool of warm water in the east and an increasing amount of moisture evaporates into the air. As more and more moisture evaporates, the hot air rises, condenses into clouds, and, ultimately, precipitates out as rain. The result is that areas that typically experience drought conditions, like Peru and Chile, instead experience heavy rains. Massive flooding often results. Exactly the opposite occurs in the western equatorial Pacific. It experiences colder ocean water closer to the surface. Consequently, there are not massive amounts of moisture evaporating into the air and the probability of rain decreases dramatically. As a result, countries in the western Pacific often experience drought conditions during an El Niño event.

*El Niño events disrupt oceanic ecologies, economic and social systems. For example, in Peru, the economy and diet depend on the large anchovy populations that inhabit the typically cold and nutrient-rich coastal waters. During an El Niño event the anchovy fish population all but disappears because warm water displaces the cold water for one or two years. The lack of anchovies affects the prosperity of the local fisherman and reduces the amount of fishmeal exported to other countries that use it to feed poultry and livestock. In the absence of sufficient fishmeal supply, more expensive alternative feed sources must be used and poultry prices increase worldwide. On the other side of the Pacific the relative lack of rainfall creates fresh water shortages, deficits in hydroelectric energy resources, crop losses, and devastating brushfires. Thus, although El Niño is a relatively local weather phenomenon it has worldwide consequences.*
Appendix B

Pacific Ocean Map (only appeared in Experiment 2)

Map Data Copyright 2016 Google, INEGI (Google Maps, 2016). See the online article for the color version of this figure.

(Appendices continue)
Appendix C

Concept Verification Task Items

Please read each statement and mark each as being either true or false based on what you read about weather patterns.

1. Within the Pacific Ocean, upwelling occurs in regions with low air pressure.
2. In regions with low air pressure, air tends to be cooler and drier.
3. An increase in the air pressure over the western Pacific Ocean increases rainfall in the eastern Pacific.
4. Under typical weather conditions, areas with higher air pressure experience cooler air temperatures.
5. The western Pacific Ocean experiences less upwelling during an El Niño phase.
6. Cloud formation increases over the eastern Pacific Ocean during El Niño events.
7. During an El Niño event, the relaxation of the trade winds results in decreased cloud formation in the eastern Pacific Ocean.
8. The relaxation of the trade winds coincides with cooler ocean temperatures in the western Pacific Ocean.
9. There is less evaporation over the western Pacific Ocean during El Niño events.
10. Water temperature increases in the western Pacific Ocean during El Niño events.
11. Under typical weather conditions, areas with lower air pressure experience warmer ocean temperatures.
12. There is always a warm water bulge in the western Pacific Ocean.
13. Under typical weather conditions, the trade winds cause upwelling in the eastern Pacific Ocean.
14. A strengthening of the trade winds results in increased evaporation in the western Pacific Ocean.
15. Warm water suppresses the upwelling of cooler water from the ocean depths.
16. Under typical weather conditions, the eastern Pacific Ocean receives comparatively more rainfall than the western Pacific.
17. El Niño events coincide with a strengthening of the trade winds.
18. Areas of lower air pressure have lower evaporation levels.

Received August 12, 2015
Revision received February 11, 2016
Accepted February 13, 2016