

When Analogies Harm: The Effects of Analogies on Metacomprehension

Jennifer Wiley, Allison J. Jaeger, Andrew R. Taylor & Thomas D. Griffin

University of Illinois at Chicago

1007 W Harrison St MC 285

Chicago IL 60647 USA

Corresponding Author: Dr. Jennifer Wiley, jwiley@uic.edu

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.

This article should be cited as:

Wiley, J., Jaeger, A. J., Taylor, A. R., & Griffin, T. D. (2018). When analogies harm: The effects of analogies and valid cues on the metacomprehension of science text. *Learning & Instruction*, 55, 113-123. <https://doi.org/10.1016/j.learninstruc.2017.10.001>

Acknowledgements

Portions of this work including preparation of this manuscript were supported by the National Science Foundation (NSF) under DUE grant 1535299, and by the Institute of Education Sciences, U.S. Department of Education, through Grants R305B07460 and R305A160008. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of either of these organizations. The authors thank Tricia Guerrero, Marta Mielicki, Tim George, and David Sarmento for their assistance and comments on this project and manuscript.

When Analogies Harm: The Effects of Analogies on Metacomprehension

Abstract

The main goal of the present research was to test whether the presence of analogies would affect the relative accuracy of metacognitive judgments about learning from expository science texts, and whether any effect would depend on the type of cues that readers used as the basis for their judgments of comprehension. In a series of experiments, students read texts that either contained or did not contain analogies; were asked to judge how well they understood each text; took comprehension tests for each topic; and were asked to self-report the basis for their judgments. Relative metacomprehension accuracy was computed as the intra-individual correlation between judgments and test performance. Results showed that the presence of analogies can lead to poor relative metacomprehension accuracy for students who fail to use situation-model-based cues to judge their understanding of text.

Keywords: Metacomprehension, Analogical Examples, Expository Science Texts

1. Introduction

Many individuals suffer from poor metacomprehension accuracy for science texts, meaning they are generally quite poor at gauging how well they have understood what they have read (Dunlosky & Lipko, 2007; Maki, 1998; Wiley, Griffin, & Thiede, 2005). A major strand of research on metacomprehension accuracy follows the lead of initial studies by Maki and Berry (1984) and Glenberg and Epstein (1985). To assess how well people can evaluate their own understanding of texts, they developed a paradigm to assess the accuracy of comprehension monitoring (metacomprehension) in which individuals study a set of expository texts, judge their understanding for each of the texts, and then complete comprehension tests. From this data, they computed *relative metacomprehension accuracy* (Maki, 1998) as a measure of whether a person knows which texts they understood the best, relative to those they understood the least. *Relative metacomprehension accuracy* is operationalized as the within-person or intra-individual correlation between an individual learner's judgments of understanding for a number of different topics, and that same learner's actual understanding for each of those topics as assessed by objective tests of comprehension.

Relative accuracy mimics the regulation of learning that students need to engage in on a daily basis. Students are routinely tasked with learning about a variety of topics during any given night of homework, as well as gauging their understanding across all topics when they study for cumulative tests. Being able to have an accurate sense of one's relative levels of understanding for various concepts is critical to effectively regulating how much effort to devote to studying and re-studying on each topic. Most studies examining *relative metacomprehension accuracy* using this standard paradigm have used sets of expository texts on scientific topics consisting of 4-6 texts over 500 words each. Typically, baseline conditions (without additional instructions or

manipulations) in these studies have found low levels of relative accuracy hovering around $r = .27$ (see Dunlosky & Lipko, 2007; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009; and Griffin, Mielicki, & Wiley, in press, for reviews). Perfect accuracy would be a correlation between judgments and performance of 1.00.

1.1. Why do individuals tend to have poor relative metacomprehension accuracy?

To explain poor relative metacomprehension accuracy, theories of metacomprehension such as the *situation-model approach to metacomprehension* (Wiley, Thiede, & Griffin, 2016) have integrated the cue-utilization approach (Koriat, 1997) which makes a distinction between the use of valid and invalid cues as a basis for monitoring judgments, with Kintsch's (1994, 1998) comprehension paradigm which makes a distinction between memory for text and learning from text. In combination, these frameworks suggest that accurate comprehension monitoring depends on the use of valid cues, and specifies that situation-model-based cues will be most valid when learning will be evaluated through comprehension questions that require representing key relations among ideas (i.e. the situation model). Following Kintsch's comprehension framework, readers should derive their predictive judgments of performance on comprehension questions based on their situation models and not based on their memory for discrete ideas or details mentioned in the text. Thus, these theories predict that situation-model cues will be the most valid basis for making accurate predictions of performance on upcoming tests to the extent that comprehension is tested by measures that depend on the quality of the situation model.

In light of these theoretical assumptions, one prevailing explanation for the low levels of relative metacomprehension accuracy that are typically observed is that students do not seem to know what kind of information they should use as a basis for their judgments of comprehension. Koriat's (1997) cue-utilization framework asserts that learners engage in an inferential process

when they are asked to make predictive judgments of how they will do on upcoming tests. That is, learners attempt to gauge their likely test performance from a variety of different types of cues. These cues can include features of the learning context that are intrinsic or extrinsic to the learning materials (e.g., amount of information, nature of the items, or number of study episodes), but they can also include cues that reflect subjective experiences of attempting to process the information. Considering both intrinsic and extrinsic features of a learning context results in theory-based inferences, where learners use their beliefs or assumptions about the effect that such features may have on learning to determine their judgments. These kinds of cues can be thought of as rule-based or heuristic because they rely on *a priori* knowledge and assumptions about how contextual features should impact the learning process in general. On the other hand, internal subjective indicators of performance derived from the actual learning experience provide a basis on which to determine the extent to which a particular item has been learned (Fischer & Mandl, 1984; Flavell, 1979; Griffin, Jee, & Wiley, 2009; Griffin et al., 2013; Nelson & Narens, 1990). In Koriat's (1997) research on judgments of learning in a metamemory context, he demonstrated that relative accuracy improves as learners shift from greater reliance on theory-based cues to greater reliance on experience-based cues when making predictive judgments about future performance on a memory test. Conceptually, this makes sense because *a priori* theories will tend to be general and thus will fail to predict variation in comprehension within the same person from one instance to the next.

Extending this logic to a context in which a reader is tasked with learning from texts, the *situation-model approach to metacomprehension* (Wiley et al., 2016) further posits that subjective *experience-based* cues that are anchored in the quality of the mental representation constructed during reading will provide the most valid basis from which to infer the extent to

which one has understood a text on a particular topic (Rawson, Dunlosky, & Thiede, 2000; Weaver et al., 1995; Wiley et al., 2005). That is, greater reliance on experience-based cues should also increase the relative accuracy of predictive judgments about future performance on a comprehension test. Further, Kintsch's comprehension framework (1994, 1998) suggests that not all subjective experiences will be equally valid indicators of understanding. Some perceptions, such as a sense of fluency, an apparent ease of processing while reading, or the feeling that one will remember details from a text, can sometimes be useful; but in other cases (i.e. when the comprehension will be tested with questions assessing understanding of key relations among ideas, rather than memory-for-details questions) they will be less useful because they are not indicative of the quality of the mental model of the phenomena that is being constructed.

Therefore these perceptions would be less valid as cues for making accurate judgments about an individual's understanding of the topic. Similarly, heuristic cues such as the length or difficulty of a text would also be less valid, because they are not invariably linked to the quality of an individual's understanding about the topic. Cues that are specifically based in the quality of the situation model, or the mental model of the scientific system or process that one constructs while reading, will be the most valid basis for making accurate predictions about performance on upcoming comprehension tests that include questions assessing understanding of key relations among ideas.

Some examples of ways that readers can generate valid cues are to engage in self-explanation about how or why a process works, and to think about the extent to which a causal explanation is complete and coherent (c.f. Griffin, Wiley, & Thiede, 2008; Jaeger & Wiley, 2014; Wiley, Griffin, Jaeger, et al., 2016). Similarly, readers can think about whether or not they could explain how and why the process works or phenomenon happens to another student. Or

they could consider whether they might be able to make new inferences based on the information, or apply the principles in the text to a new situation. Importantly, readers do not need to possess a high quality situation model or to be able to construct a complete and coherent causal explanation in order to have good relative metacomprehension accuracy. Readers just need to *attempt* to engage in these activities because the results of these attempts represent *experiences* that will provide them with cues about the quality of their mental models, which in turn will provide them with a valid basis for making their judgments of understanding.

On the other hand, relying on superficial cues such as text length, fluency, interest in the topic, or perceived difficulty should be less likely to lead to accurate judgments of understanding, and thus these superficial cues can be seen as less valid. Thiede, Griffin, Wiley, and Anderson (2010) found evidence consistent with these theoretical assumptions by asking readers to self-report the types of cues that they used as a basis for their judgments of understanding. They found that students who reported using heuristic cues were generally less accurate than those who reported using experience-based cues. Similarly, Jaeger and Wiley (2014) found that students who reported using comprehension-based cues (using their ability to explain, summarize or make connections while reading the text, or whether they thought they could answer questions like those they saw on the practice test as the basis for their judgments) had higher levels of relative metacomprehension accuracy than those who used non-comprehension-based cues.

1.2. How do instructional adjuncts affect metacomprehension accuracy?

To date, most studies on metacomprehension have investigated accuracy in the context of reading non-illustrated expository text passages. Work has just begun to assess how different adjuncts to text, or features of texts, might influence metacomprehension accuracy. The main

question for the current line of investigation is whether the presence of analogies, a common instructional device, might impact metacomprehension accuracy.

Because not much work has been done yet with instructional analogies and metacomprehension, it is useful to consider a related line of research exploring the effects of including visualizations alongside text. In some domains like geology, biology and chemistry, a popular way of supporting understanding is through providing visualizations such as diagrams, schematics, animations, or simulations that are thought to be helpful in conveying invisible phenomena or relations that may be implicit in written text (Ainsworth & Loizou, 2003; Butcher, 2006; Larkin & Simon, 1987). Despite their intuitive promise, many studies have shown that visualizations can fail to improve or even impede learning outcomes compared to plain text (Hegarty, 2014; Höffler & Leutner, 2007). Much depends on the topic, the quality of the visualizations, and the learning task, as well as on the learner. The theory of multimedia learning suggests that when designing multimedia-learning materials, such as texts accompanied by visualizations, the presented material should have a coherent structure and be presented in a manner that provides guidance as to how to build an accurate mental model. In particular, Mayer (2005) lays out several principles of multimedia instructional design that are aimed at minimizing the effects of extraneous processing by emphasizing main ideas and increasing access to cues that indicate how visual and textual information should be coordinated.

Theoretically, including diagrams displaying conceptually important relations could improve metacomprehension accuracy. The presence of diagrams could facilitate metacomprehension accuracy by providing students with a standard for evaluating their own comprehension. Diagrams can serve as a benchmark against which students could use to judge their own understanding and evaluate gaps in their knowledge. Students could compare the

causal relations presented as part of the illustrations with the relations represented in their own mental models. Similarly, illustrations could provide a basis for evaluations or comparisons, or could help to illustrate standards and comprehension goals to students, all of which in turn should allow for more accurate judgments.

Alternatively, one could hypothesize that including images alongside expository text could harm metacomprehension accuracy. Images can impact readers' enjoyment and interest in learning about a topic (Harp & Mayer, 1998). Also, many people generally believe that images improve learning (Serra & Dunlosky, 2010). People may gain the impression that they comprehend images well since the gist can be extracted in a glance (Loftus, 1972; Potter, 1975), or because causal diagrams used in instruction typically simplify processes (Keil, 2006). Perceptions of fluency, enjoyment, interest, and *a priori* beliefs about images' utility for learning can serve as heuristic cues that can undermine the use of more predictive experience-based cues tied to the quality of one's mental representation of the process described by particular text. Additionally, sometimes the images included alongside science texts are not directly relevant to understanding the underlying causal model of the phenomenon (Harp & Mayer, 1998, Sanchez & Wiley, 2006). In these cases, if students still rely on their *a priori* beliefs that images are helpful for learning even when they are not, this may also harm their metacomprehension accuracy.

Recent studies have found support for the latter hypothesis that images may be harmful to the judgment process. Serra and Dunlosky (2010) found that students strongly endorse the belief that learning is greater from multimedia than from text alone. This belief was related to higher predictions of test performance when a six paragraph text on lightning formation was accompanied by images than when the same text was presented without images, regardless of

whether the images were a series of six conceptual diagrams (depicting relations or processes presented in the text) or six decorative photos (photographs of lightning strikes). Students showed higher judgments for illustrated texts prior to reading the texts, and did not re-adjust those ratings after reading, despite the fact that the decorative photo condition did not yield greater learning compared to a no-image control. From these results, Serra and Dunlosky suggested that readers used *a priori* beliefs about multimedia learning as a general heuristic that biased their judgments. Metacomprehension accuracy was not computed for this study, but the authors argued that use of this multimedia heuristic could lead to reduced accuracy when there is no actual learning benefit from images.

Lenzner, Schnotz and Mueller (2013) found a similar result that decorative pictures reduced the perceived difficulty of a lesson on “Light and Shadow” for middle school students, but they did not actually affect learning. Ikeda, Kitagami, Takahashi, Hattori, and Ito (2013) also found that the presence of instructional graphics (functional magnetic resonance (fMRI) images or bar graphs) can affect the magnitude of metacomprehension judgments. In a first experiment, they showed that undergraduates who read a text about brain activity in patients with depression accompanied by fMRI images gave higher metacomprehension judgments than a text-only condition. However, performance on a comprehension test did not differ between conditions. In a second experiment they found that fMRI brain images increased the magnitude of judgments more so than bar graphs. However, again there were no actual differences in comprehension due to graphic type.

The above studies used limited types of images and only a single image or single topic. Another study using a larger set of topics and images suggests that readers do harbor an expectation that diagrams with realistic qualities are particularly useful for supporting

understanding. Wiley, Sarmento, Griffin, and Hinze (2017) asked participants to make pre-reading predictions on how much they thought a variety of images that appeared in authentic biology textbooks would impact their interest in a stated topic. They also asked participants to predict how much they thought these same images would impact their understanding if they were included as part of a text. There were 10 different biology topics presented, each paired with 7 types of images that varied on the dimensions of form (elements of realism or abstract conventions) and function (mere depiction to causal explanation), this resulted in a total of 70 images. Of central interest was a subset of explanatory diagrams that were similar to each other in their depiction of relations among important causal ideas, but differed from each other in that some were purely abstract while others included realistic depictions of biological organisms or their features. Independent of the conceptual function of the image, the presence of realistic elements impacted topic interest, and variations in interest ratings for the images were related to variations in expected effects on understanding. Participants expected explanatory diagrams to benefit their understanding only if they included realistic elements. Although no learning outcomes were collected in this study, this finding suggests that the presence of realistic features in some images could alter judgments of understanding. This may be problematic to the extent that other studies have failed to find learning benefits due to the presence of realism in images (Butcher, 2006; Ikeda et al., 2013).

Another study (Jaeger & Wiley, 2014) has demonstrated the effects of including visual adjuncts on metacomprehension. In contrast to the work reviewed above, this study was able to compute measures of relative accuracy. Jaeger and Wiley (2014) presented a set of six texts on six different topics that were either each illustrated either with a single decorative image (i.e. a photograph of lightning) in one condition, a single conceptual illustration (simple schematics

representing relations and processes described by the text) in another condition, or no images in a third condition. They demonstrated that the presence of decorative images led to poorer relative accuracy compared to plain text. Ackerman and Leiser (2014) found similar results. However, including conceptual images neither helped nor harmed relative accuracy compared to the plain text condition.

On the other hand, a benefit was found in the second experiment of Jaeger and Wiley (2014) when conceptual images were paired with an instruction for students to explain each text silently to themselves while studying the material. This latter result coheres with Ainsworth and Loizou's (2003) finding that diagrams can facilitate self-explanation during learning, and suggests that conceptual images may only help monitoring accuracy if readers are instructed to engage in tasks like explanation, and are prompted to make use of the metacognitive affordances of the images, such as by using the processes and elements depicted in the images as a way to evaluate the quality of their own mental models. Without such direction, readers may use any variety of cues derived from the images as a basis for their judgments of understanding, including irrelevant features of images such as their aesthetic appeal.

These findings related to visualizations can be used for projecting the potential effects that including instructional analogies within science texts might have on metacomprehension (Jaeger & Wiley, 2015). If learners rely upon heuristic cues (such as interest or familiarity) to judge their understanding of a topic, they may be likely to incorporate only superficial information from the instructional analogies that will impede the accuracy of their judgments across a set of texts. In contrast, if learners rely upon experience-based cues tied to their construction of a situation-model (such as those that could be generated by efforts to self-explain the relevance of the instructional analogy, or attempts to integrate the analogy and target

concepts), they may be more likely to make accurate judgments of understanding across a set of texts.

1.3. How do instructional analogies affect metacomprehension accuracy?

The present research examines the effects of instructional analogies on the ability of readers to accurately judge their understanding across a set of texts. The same tension exists with analogies as with visualizations. Analogies are generally believed to aid learning of unfamiliar concepts, and to improve understanding when they are included in expository science texts. This belief is reflected in the frequency with which textbooks draw on analogies to explain concepts (Curtis & Reigeluth, 1984). It has been suggested that analogies may be most effective for supporting the understanding of novice learners who lack prior familiarity or knowledge of topics, or low-ability learners because they provide more explicit guidance or scaffolding (Bean, Singer, & Cowan, 1985; Duit, 1991; Mayer, 1989). Despite the implicit belief that analogies aid learning from science text, research findings in this area are mixed, sometimes supporting modest advantages for instructional analogies, and sometimes not demonstrating clear advantages (Alexander & Kulikowich, 1991; Bean, Searles, & Cowen, 1990; Donnelly & McDaniel, 2000; Gilbert, 1989; Hammadou, 2000; Jaeger, Taylor, & Wiley, 2016; Jaeger & Wiley, 2015).

According to the structure-mapping theory, processing complex analogies in science takes place by mapping the relations between the base (or familiar) domain to the target (or new) domain (Gentner, 1983). Mapping refers to how knowledge about the base is carried over to the target, and allows readers to generate inferences and construct a mental model of the target. The primary goal is not just to provide an anchor, but also to invoke comparison between the two domains. The process of comparison can help the reader to comprehend something new or

complex by pointing out its similarities, or differences, to something familiar (Kurtz, Miao, & Gentner, 2001).

Just as there are principles associated with designing optimal multimedia, there are also some general criteria for what makes a good analogy (Gentner, 1982). Generally, it is accepted that in order for an analogy to be effective, readers must have a good understanding of the base domain prior to engaging in analogical processing (Duit, 1991; Wilbers & Duit, 2006). Another important consideration is the clarity of the analogy, or how precisely the alignments (or alignable differences) can be defined. Further, analogies can vary in their abstractness, that is, whether the mappings are between relations or between attributes. Analogies that include too many attributes or spurious attributes may be less effective because readers could focus on these alignments rather than the more important relational alignments (Iding, 1997).

Similar to visualizations, while the presence of instructional analogies as part of a science text may improve understanding of the topic under some circumstances, it also has the potential to invoke a misleading sense of understanding. Analogies are thought to help students to learn difficult, abstract, temporal-spatial concepts by providing a concrete or familiar case (Curtis & Reigeluth, 1984; Orgill & Bodner, 2006; Thiele & Treagust, 1994) that can be used as a basis for understanding a novel target concept (Kurtz et al., 2001). Indeed, students report that texts with analogies are more interesting, more enjoyable to read, and easier to understand than texts that do not reference analogous cases (Jaeger & Wiley, 2015; Paris & Glynn, 2004). These very aspects of analogies could be responsible for a misleading sense of familiarity, ease, and fluency for the novel to-be-learned concepts. Analogies vary in how familiar they are to readers, but ideally will be more familiar to readers than the target concept. If readers rely on heuristic cues like familiarity for the analogical example, then the presence of analogies may make judgments

about understanding of the target concept less predictive of actual comprehension. To the extent that the presence of an instructional analogy within a text gives readers access to a wider range of cues that they may use to make judgments of understanding (including invalid cues such as the familiarity or interest in the example), then providing an analogical example may actually undermine metacomprehension accuracy. On the other hand, if readers use the analogical cases to test their own understanding, or if analogical comparisons prompt efforts to self-explain how the example and target concepts relate to each other, then this could generate valid experience-based cues that should improve their metacomprehension accuracy.

Given the critical role of cue-basis in determining how instructional adjuncts can impact metacomprehension accuracy, the main goal of the present set of studies was to test theoretical predictions that the presence of analogies might affect relative metacomprehension accuracy, as well as to test whether any effects on accuracy depend upon the type of cues that learners rely on for their judgments. If learners rely upon heuristic cues (such as interest or familiarity), they may be likely to attend to or incorporate only superficial information from the analogies that will impede the accuracy of their judgments across a set of texts. In contrast, if learners rely upon experience-based cues tied to their construction of a situation-model (such as those that could be generated by efforts to self-explain the relevance of the analogy), they will be more likely to make accurate judgments across a set of texts.

2. Experiment 1

Experiment 1 provided an initial test as to whether the presence of an analogy would affect the relative accuracy of judgments of understanding for a set of expository science texts, and whether any effect would depend on the type of cues that readers used as the basis for their judgments.

2.1 Method

2.1.1. Participants

Ninety students from the Introductory Psychology Subject Pool at the University of Illinois at Chicago participated in partial fulfillment of a course requirement. This population was studied precisely because most students in the pool would have low prior familiarity with these science topics, making this an appropriate population in which to examine learning from these texts (as opposed to studying students enrolled in a college-level course in natural science who are more likely to already possess knowledge about these topics). Seven participants were dropped due to a lack of variance in their judgments, preventing computation of relative accuracy. The final sample was 83 participants, which was 52% female, and were typically first-year college students with an average age of $M = 18.72$ ($SD = 0.94$). Although all students were fluent in English, 21% were born outside the US, and only 66% reported being native English speakers. The average score on the American College Testing reading test (ACT READING) was $M = 23.44$ ($SD = 4.53$) and did not differ across conditions, $t < 1$. (This data was self-reported and scores were missing for 20 participants.) This suggests that on average students were able to read at the college level. However, as is typical in this subject pool population, close to a third of the sample (38.4%) reported scores at or below the college readiness benchmark for reading. For this reason, these studies used texts written at a high school reading level.

2.1.2 Materials

2.1.2.1. Texts. Participants read science texts on five topics (Atoms, Vision, Circulatory System, Weather Patterns, Lightning) adapted from Hinze, Pellegrino, and Wiley (2013). The texts were presented in 12 point Times New Roman font, and each was split across two consecutive pages. The base texts were approximately 550 words each and were written at a high school level

(Flesch-Kincaid grade level of 9-10; see the Appendix for an example). The analogy versions of the texts included an average of 188 additional words that related the scientific phenomenon described in the text to a familiar base concept. The atom analogy described how the sun is at the center of the solar system with the planets orbiting around it, the vision analogy described how a camera captures images, the circulatory system analogy described how water travels around a city through plumbing systems, the weather pattern analogy described the movement of air while inflating and deflating a balloon, and the lightning analogy described how static is created when our feet rub on carpet.

2.1.2.2. *Judgments of Comprehension.* Participants were prompted to make predictive judgments ranging from 0 to 5: “If you were to take a test on the material you just read, how many questions out of 5 would you answer correctly on the test?” Each participant made one judgment of comprehension (JOC) after each of the five texts.

2.1.2.3. *Tests.* For each text, a five-item multiple-choice comprehension test was created. Each item was designed to assess understanding of important relations among ideas in each text (Donnelly & McDaniel, 1993; Royer, Carlo, Dufresne, & Mestre, 1996; Wiley et al., 2005). Two example items for the text about weather patterns are included in the appendix. In the first example question, the reader is asked where a warm water bulge will occur. The answer to this question is not explicitly stated in a single sentence in the text, but can be inferred based on information from these sentences in the text, “The surface layer is the warmest because there is a concentrated amount of sunlight hitting the surface at the equator. This surface water is what moves in response to the trade winds. As the trade winds blow from east to west they push steadily against the sea for thousands of miles, dragging the surface water along that same path. As a result, warm surface water accumulates in the western Pacific” Other questions required

recognizing key relations that were mentioned in the text, such as the connection between air movement and air pressure, as targeted by the second example test question “Across the Earth (not just the Pacific Ocean), air always moves from...”.

The tests were designed so that test scores for each text would neither be at floor nor at ceiling, with actual performance covering the full range of scores, but average performance falling between 2 and 3 items correct for each topic. Performance on these types of comprehension tests have been shown to reliably correlate with other comprehension assessments, including performance on “how” and “why” essay questions (Hinze & Wiley, 2010; Hinze et al., 2013; Jaeger et al., 2016; Sanchez & Wiley, 2006, 2010; Wiley et al., 2009; Wiley & Voss, 1999), as well as with scores on the ACT and Nelson Denny Reading Comprehension Test (Griffin et al., 2008; Wiley, Griffin, Jaeger, et al., 2016). Reliability of this assessment approach is also evidenced by the highly replicable effects on relative metacomprehension accuracy scores in studies using these types of tests and various interventions including delayed generation tasks, explanation activities, and test-expectancy instructions (Griffin et al., 2008; Jaeger & Wiley, 2014; Redford, Thiede, Wiley, & Griffin, 2011; Thiede, Wiley, & Griffin, 2011; Wiley, Griffin, Jaeger, et al., 2016).

Topics were presented in the same order for testing as they had been seen during reading. Because the tests were all multiple choice, participants either got a 1 (correct) or 0 (incorrect) for each test item, resulting in scores ranging from 0 to 5 for each test.

2.1.2.4. Final Survey. Participants completed a paper-and-pencil measure of their cue basis that asked them to describe what information or cues they used when trying to predict their future test performance. The exact prompt they were given was, “After you read each text, you were asked

to predict how many items out of 5 you would get correct on a test. What information or cues did you use to make these predictions after reading each text? What did you consider?"

2.1.3. Procedure

Prior to beginning the experiment, each participant completed an agreement-to-participate form. Participants were randomly assigned to either the analogy or no-analogy condition. The main portion of the experiment was completed on the computer. The experimenter instructed participants to click a link that began the task. Introductory instructions were the same across conditions: "In this study, you will be reading a series of texts, estimating how many questions you can get correct on a five item multiple-choice test, and then taking a series of tests to see how well you actually do."

After the instructions, participants moved on to the texts and made their judgments immediately after reading each text. Once the last judgment was made, students completed the multiple-choice tests. Lastly, each participant completed the final survey then were debriefed and thanked for their participation.

2.1.4. Coding and Reliability

Relative metacomprehension accuracy was computed as an intra-individual Pearson correlation between JOCs (0-5) and comprehension test scores (0-5) for each text. Griffin et al. (2008) recommends the use of the Pearson correlation as a measure of relative accuracy when both judgments and test scores represent a continuous range of values rather than dichotomous data. The original recommendation to use the Gamma correlation as a test of relative accuracy (Nelson, 1984) was made in relation to tests and judgments with dichotomous values (i.e. when people are predicting whether or not they will recall individual memory test items). In such an instance, it makes sense to compute a tally of hits and misses as an accuracy measure, which is

essentially what is involved in computing Gamma. Gamma attends only to the frequency of concordances and discordances between judgment-performance pairs, while completely ignoring the magnitude of the differences that arise when both variables are measured on scales with a range of values. When predictions and performance are measured on non-dichotomous scales, Gamma correlations essentially treat them as dichotomous, reducing the range of observed values, and increasing the possibility of ties, ceiling, and floor effects with many observations being pushed to the maximum and minimum values of +/- 1.0.

For example, in this study there were only 20 unique values for relative metacomprehension accuracy as computed with Gamma, including 16 scores of 0, 22 scores of +1.0 and 13 scores of -1.0. In contrast, there were 68 unique values for relative metacomprehension accuracy as computed with Pearson, with only 8 scores of 0, 1 score of +1.0 and 0 scores of -1.0. Thus, Pearson provides a more continuous metric of covariance between predictions and test scores, which is more appropriate for use as the dependent variable in general linear models.

Gamma (Nelson, Narens, & Dunlosky, 2004) and d_a (a similar measure to Gamma developed by Masson & Rotello, 2009) are also reported. Absolute accuracy was computed as the average absolute difference between JOCs and test scores for each text for each individual. Although absolute accuracy results were analyzed, it is important to acknowledge that absolute accuracy is largely determined by overall levels of test performance, which is related to test difficulty, and overall levels of judgment magnitude, which are dependent on learners' general heuristic assumptions about themselves and the task (Griffin, Wiley, & Salas, 2013). In contrast, relative accuracy is conceptually and statistically orthogonal to both overall test performance and average judgment magnitude (Nelson, 1984).

In addition, because the effect of analogies on metacomprehension accuracy could hinge upon the type of cues learners use to form their judgments of understanding, a measure of cue use was collected during a final survey to test whether it would moderate the effect of analogies on metacomprehension accuracy. Responses to the cue-basis question were expected to range from comments related to superficial text features such as length and interest, to verbatim memory processes, to comments about deeper integration processes. Based on Thiede, Griffin, Wiley, and Anderson (2010), the comments were coded into two categories reflecting whether they implicated *Superficial* use of heuristic or memory-based cues (e.g., “If I was interested in the text I felt I would do better on the test.”) versus use of *Situation-Model* cues (e.g., “I mentally tried to explain the main ideas”). Responses were coded by two independent raters blind to condition, with a Krippendorff’s $\alpha = .90$. The first coder’s data were used for analyses.

With only five texts, judgment-performance correlations could be unreliable and overly influenced by one particular text. To address this, the reliability of the correlations between the texts and tests was demonstrated using the subsets procedure developed in Maki, Jonas, and Kallod (1994). To create five distinct subsets, a different text and test was dropped from each subset. Relative accuracy scores were then computed using intra-individual Pearson correlations among the remaining four texts and tests. The Intra-class Correlation (ICC) among the relative accuracy scores for the five subsets was estimated as Cronbach’s $\alpha = .86$, indicating a high level of consistency (i.e., the findings are not being driven by one text or judgment).

2.2. Results

2.2.1. Metacognitive Judgments and Test Performance

Mean judgment magnitudes, test performance, and absolute accuracy for the four conditions are shown in the first three rows in Table 1. None of these differed due to analogy conditions or cue type conditions, nor were there significant interactions (all F s < 1).

2.2.2 Relative Metacomprehension Accuracy

The intra-individual Pearson correlation reflecting average relative accuracy for each individual in each condition is shown in the fourth row of Table 1. A 2 (condition) x 2 (cue type) ANOVA revealed no significant main effect for analogy condition, $F < 1$. However, there was a significant main effect for cue type, $F(1, 79) = 4.26, p < .04, \eta_p^2 = .05$, which was subsumed by a significant interaction between cue type and analogy condition, $F(1, 79) = 4.69, p < .03, \eta_p^2 = .06$. Relative accuracy computed using gamma and d_a (shown on rows 5 and 6 of Table 1) showed the same patterns as relative accuracy computed using Pearson. Results did not change when test performance was included as a covariate in the model. Follow-up tests using Tukey's Highly Significant Difference (HSD) tests indicated that the interaction was due to readers who reported using superficial cues doing more poorly in the analogy condition than readers who reported using situation-model cues ($p = .01$).

2.3. Discussion

The central finding from Experiment 1 was the interaction, which revealed that analogies can have different effects on relative accuracy depending on the cues that readers use to judge their comprehension. A limitation of this study was that the instructional analogies did not have a positive effect on comprehension itself. To examine the effects of instructional analogies on relative metacomprehension accuracy in a context where the analogies were actually effective in facilitating learning of the target phenomena, the materials were revised so that the analogy-embedded texts led to better performance on comprehension tests in Experiment 2.

3. Experiment 2

Experiment 2 was largely an attempt to replicate the results seen in Experiment 1, with revisions to the stimuli so that understanding of the topics was improved due to the presence of the instructional analogies. If analogies only negatively impact relative metacomprehension accuracy when they are ineffective for learning as was the case in Experiment 1, then the results of Experiment 1 should not replicate. On the other hand, if readers who rely on more superficial cues still rely on less valid features of the analogies as a basis for their judgments even when the analogies do facilitate comprehension, then the analogy condition should still result in lower relative accuracy.

3.1. Method

3.1.1. Participants

One hundred and eight students from the Introductory Psychology Subject Pool at the University of Illinois at Chicago participated in partial fulfillment of a course requirement. Ten participants were dropped due to lack of variance in their judgments resulting in a final sample of 98 participants which was 60% female, and who were typically first year college students with an average age of $M = 19.18$ ($SD = 1.23$). In this sample, 22% were born outside the US, and only 50% reported being native English speakers. The average ACT READING score was $M = 23.80$ ($SD = 5.13$) and did not differ by condition ($t < 1$). (This data was self-reported and scores were missing for 28 participants.) Close to a third of the sample (34.2%) reported scores at or below the college readiness benchmark for reading.

3.1.2. Materials

The materials were similar to Experiment 1 except that the text set was revised to include six topics (Weather Patterns, Lightning, Endocrine System, Vision, Global Warming, Circulatory

System). Test performance in Experiment 1 suggested that the only analogy that did not improve understanding was the solar system analogy in the text on the atom. This entire text was dropped and two new texts were added covering the Endocrine System (with a satellite television analogy), and Global Warming (with a parked car analogy). To further increase learning from the analogies (Jaeger et al., 2016), the analogical material was moved from the beginning of the passages to being interleaved and more integrated into the discussion of the target phenomena, as shown in the appendix. In addition, participants received a seventh text on digestion as a practice passage. The practice passage was the same for both conditions and did not include an analogy.

3.1.3. Procedure

The procedure was the same as in Experiment 1 except that prior to reading the target passages, participants read a practice passage on digestion (with no analogy), completed a practice judgment, and then took a practice test on that first passage to better understand the types of comprehension questions they would be receiving. The practice passage and test was added because prior work has indicated that students generally expect test questions that rely on verbatim memory for the text, and that instilling a clear expectancy that tests will contain comprehension questions requiring attention to relations among ideas can improve relative metacomprehension accuracy (Thiede et al., 2011; Wiley, Griffin, & Thiede, 2008; Wiley, Griffin, Jaeger et al., 2016).

3.1.4. Coding and Reliability

Cue basis was coded using the same procedures as in Experiment 1. Responses were coded by two independent raters who were blind to condition, which resulted in a Krippendorff's $\alpha = .90$. The first coder's data were used for analyses.

Reliability was computed using the same subset procedure as in Experiment 1. Relative accuracy scores were computed using intra-individual Pearson correlations for each of the six possible subsets of five texts and tests for each participant. The ICC among the five relative accuracy scores was estimated as Cronbach's $\alpha = .95$, indicating a high level of consistency (i.e., the findings are not being driven by one text or judgment).

3.2. Results

3.2.1. Metacognitive Judgments and Test Performance

Mean judgment magnitudes, test performance, and absolute accuracy for the four conditions are shown in the first three rows in Table 2. The mean magnitude of judgments did not differ due to analogy conditions or cue-type conditions ($F_s < 1.14$). Test performance was higher in the analogy condition, $F(1, 94) = 6.47, p < .02, \eta_p^2 = .06$. However, there was no effect of cue type, nor was there an interaction, $F_s < 1$. Absolute accuracy did not differ due to analogy or cue-type conditions, nor was there a significant interaction (all $F_s < 1$).

3.2.2. Relative Metacomprehension Accuracy

The intra-individual Pearson correlation reflecting average relative accuracy for each individual in each condition is shown in the fourth row of Table 2. No significant difference in relative metacomprehension accuracy was seen for analogy condition, $F < 1$. However, there was a significant main effect for cue type, $F(1, 94) = 6.68, p < .02, \eta_p^2 = .07$, which was subsumed by a significant interaction between cue type and analogy condition, $F(1, 94) = 5.72, p < .02, \eta_p^2 = .06$. Relative accuracy computed using Gamma and d_a are shown in rows 5 and 6 of Table 2. They showed the same patterns as relative accuracy computed using Pearson. Results did not change when test performance was included as a covariate in the model in an attempt to account for better performance in the analogy condition.

Follow-up tests using Tukey's HSD indicated that the significant interaction was due to two significant differences among conditions. Readers who reported using superficial cues did more poorly in the analogy condition than readers who reported using situation-model cues ($p = .004$); and readers who reported using superficial cues did worse with the analogies than without analogies ($p = .03$). As in Experiment 1, relative accuracy was worst when readers used superficial cues and read texts with analogies.

3.3. Discussion

Thus, Experiment 2 provides a replication and extension of the findings from Experiment 1, showing that analogies had different effects on relative accuracy depending on the cues that readers used to judge their comprehension. Readers who relied on superficial cues were harmed by the presence of the analogy, while readers who used situation-model cues were not. Importantly, this replication occurred in a context where the analogy versions of the passages did lead to better overall performance on the comprehension tests, and the effect was robust enough to withstand changes in text topics, how the analogies were integrated into the text, and inclusion of a practice text and test.

4. Experiment 3

An important limitation of both Experiments 1 and 2 is that participants were not randomly assigned to cue-use conditions. Rather, readers were categorized by their natural tendencies toward using superficial versus situation-model cues as a basis for their judgments. To address this limitation, Experiment 3 introduced a manipulation that prompted participants to either use situation-model cues or superficial cues as the basis of their judgments. Following earlier work, an explanation instruction and greater detail about the nature of upcoming tests (Griffin et al., 2008; Jaeger & Wiley, 2014; Thiede et al., 2011; Wiley, Griffin, Jaeger et al.,

2016) was provided to half of the participants in order to encourage the use of valid, situation-model cues as a basis for judgments. The other half of the participants were instructed to consider their interest in the texts, familiarity with the topics, and perceptions of the memorability and difficulty of the texts to encourage the use of less valid cues as the basis for judgments. The goal of this study was to further explore the most reliable effect found in the first two studies which was the difference between cue-type groups within the analogy condition. Given this focus, both cue-use conditions received only the analogy versions of the texts in this study.

In addition, the results of Experiments 1 and 2 suggested that some readers may have been misled by the analogies, not because they were using superficial or heuristic cues, but rather because readers may have relied on cues tied more directly to their comprehension of the analogical example (e.g., how heat gets trapped in a car) rather than their comprehension of the target phenomena (e.g., global climate change). To limit this possibility, in Experiment 3, we added an analogy to the digestion practice text while keeping the comprehension practice test items the same, which (like all the actual tests) had no questions referencing the analogical example. This was intended to reduce any misleading expectations that the analogy itself would be part of the test questions.

4.1. Method

4.1.1. Participants

One hundred and eleven students from the Introductory Psychology Subject Pool at the University of Illinois at Chicago participated in partial fulfillment of a course requirement. Eleven participants were dropped due to incomplete data or a lack of variance in their judgments resulting in a final sample of 100 participants, with 50 in each condition. The sample was 49%

female, and typically in their first year of college with an average age of $M = 18.89$ ($SD = 1.16$). In this sample, 21% were born outside the US, and 64% reported being native English speakers. The average ACT READING score was $M = 22.33$ ($SD = 4.55$) and did not differ by condition ($t < 1.05$). (This data was self-reported and scores were missing for 42 participants.) Almost half of the sample (51.7%) reported scores at or below the college readiness benchmark for reading.

4.1.2. Materials and Procedure

The materials and procedure were the same as in Experiment 2 with a few exceptions. First, both cue-use conditions received only the analogy versions of the texts, and the practice text on digestion was altered to include an analogical comparison to a recycling center. This contrast was designed to focus on the most reliable effect found in the first two studies, namely the difference between cue-type groups within the analogy condition.

Second, participants were randomly assigned to receive an instruction that prompted them to consider either situation-model cues or superficial cues. After taking the practice test on digestion, and before reading the target texts, half the participants received this instruction, designed to prompt them to make their judgments using cues based in their situation model.

“The comprehension tests are meant to test your understanding of each of the topics. The questions might ask you about connections that you could make between parts of the text, or conclusions that you could draw based on the reading. Some questions might ask you to think about explanations of how or why a process is occurring. For example, some of the questions about the digestion passage asked you about how the digestion system works, and what could alter the digestion process. When you are asked to make your judgments of understanding for each passage, you should think about whether or not you think

you could answer these types of questions. For example, do you think you could explain the causal process of digestion to a friend?”

The other half were received this instruction, designed to prompt them to make their judgments using more superficial cues:

“The comprehension tests are meant to test your understanding of each of the topics. When you are asked to make your judgments of understanding for each passage, you could think about whether or not you thought the text was easy to read. You could think about whether you found the material to be interesting or enjoyable to read, or how quickly you were able to read the text. You could ask yourself whether the ideas seemed familiar, or whether you have prior knowledge about the topic. You could also make your judgments based on your memory for the text. For example, how much of the text do you feel like you would be able to recall? How much of the content do you feel like you can remember?”

As in Experiments 1 and 2, reliability was computed using the same subset approach. The ICC among each of the 6 subsets was estimated as Cronbach’s $\alpha = .85$, indicating a high level of consistency (i.e., the findings are not being driven by one text or judgment).

4.2. Results

Descriptive statistics and results for all analyses are reported in Table 3. No significant differences were seen in either judgment magnitude or test performance due to the judgment manipulation. No significant difference was seen between conditions in absolute accuracy.

However, relative metacomprehension accuracy was significantly higher when readers were prompted to use situation-model cues than when they were prompted to use more superficial cues. While the pattern of means was similar when relative accuracy was computed

with gamma and d_a , those analyses did not reach significance. Results did not change when test performance was included as a covariate.

4.3. Discussion

These results show that when readers were prompted to use situation-model cues tied to the comprehension of the target phenomena as a basis for their judgments, it led to higher relative metacomprehension accuracy from expository texts containing analogies than when readers were prompted to use superficial cues.

5. General Discussion

In the first two experiments, the presence of analogies in expository science texts led to poor relative metacomprehension accuracy among readers who relied upon more superficial cues that were not tied to the processes of constructing, using, or evaluating mental models for the phenomena described by the texts. This result occurred regardless of whether the analogies facilitated actual comprehension of the target phenomena, whether the analogies were contained in a single paragraph or interleaved throughout the text, and whether or not readers were provided with a prior practice text and test to create an appropriate expectancy for inferential comprehension tests. On the other hand, readers who utilized judgment cues tied to their situation model showed no negative effects of analogies. In Experiment 3, readers who were prompted to use situation-model cues showed a benefit in monitoring comprehension from expository texts containing analogies over those who were prompted to use more superficial cues. The results lend further support to the situation-model approach to metacomprehension (Wiley et al., 2016), in which the key to making accurate predictive judgments of performance on comprehension tests (that will assess their understanding of relations among ideas) is

utilization of subjective experiences during reading that are tied to the quality of one's situation model.

While these studies demonstrated differences in relative metacomprehension accuracy due to the interaction between analogy and cue-use conditions, no significant differences were seen in absolute accuracy. This is not surprising because cue use was unrelated to overall test performance and overall judgment magnitude, and analogies did not impact judgment magnitude. An oft-cited problem with absolute accuracy as a measure of metacomprehension is its high statistical dependence upon mean test performance and mean judgment magnitude (Nelson, 1984). The divergent results on relative versus absolute accuracy in all three studies further support recent recommendations that these two measures be treated as measures of statistically and conceptually distinct constructs (Griffin et al., 2013; in press). Researchers should avoid the tendency to assume that results for one measure have any implications for the other. In addition, that fact that the effect for analogies was only seen on relative accuracy shapes the interpretation of how they are having their impact. Contrary to work suggesting that students' *a priori* general beliefs about images inflate their judgments of understanding when images accompany text (Serra & Dunlosky, 2010), the current results do not support such a generalized heuristic regarding the presence of analogies in instructional texts. A generalized heuristic would predict generally inflated judgment magnitudes overall and an impact on absolute accuracy. Instead, analogies seem to impact judgments in a more idiosyncratic manner that is specific to the particular topic and the particular analogical example that is used. Text-to-text variability in how familiar or interesting each particular analogy is to the reader would have inconsistent impacts on judgments. This inconsistency would not show up in overall judgment magnitude or

absolute accuracy, but would impact the relative differences in judgments, and thus relative accuracy.

5.1 Limitations and future directions

Of course, the results of these studies are limited by the particular sets of expository science texts, analogical examples, and the particular undergraduate population that was used. To generalize these results, other instructional materials and reader samples will need to be tested. It is also important to note that this work did not find overall detriments or benefits in monitoring due to the presence of analogies in expository science texts. Rather, the effects of analogies depended on the types of cues that readers used to judge their own understanding. The results of Experiments 1 and 2 showed that readers who relied on superficial cues were harmed by the presence of analogies, while readers who used situation-model cues were not. Like other studies on learning with analogies, this suggests that individual differences among learners are going to be important to consider.

It has been suggested that analogies may be especially beneficial for low prior knowledge or low ability students. For example, Jaeger, Taylor, and Wiley (2016) demonstrated that the presentation format of an analogy within a science text interacts with spatial thinking skills. In particular, their results indicated that students who scored low on a test of spatial thinking skills demonstrated better understanding of the global weather phenomenon of El Nino when an analogy was interleaved throughout the base text; whereas students who scored high on the spatial measure demonstrated better understanding in conditions when no analogy was present. The authors argued that low spatial individuals benefitted from the additional mapping provided by the interleaved presentation and suggested that this format may have assisted them in generating a spatial mental model, something that low spatial students tend to struggle with. On

the other hand, they argued that students who demonstrated stronger spatial thinking skills might have been better off without additional support, since they possessed the skills needed to engage in active construction of a spatial mental model on their own. In a similar vein, Newton (2003) has suggested that younger students may struggle to develop coherent mental models because they do not yet possess the full set of cognitive abilities required to do so. Based on this idea, younger readers may benefit from the added support similar to the low spatial students in Jaeger et al. (2016). However future work needs to test whether lower ability or younger readers might also show this effect, or if they may be even more likely to attend to superficial features and thus may require different conditions than older students to benefit from analogies.

Prior knowledge is another important learner characteristic to consider. This work explored science learning among students who were unlikely to hold extensive prior knowledge about the topics. Donnelly and McDaniel (2000) suggested that the presence of analogies could have different effects among learners with extensive background knowledge. They suggest that analogies may actually mislead, constrain, or simplify the kind of elaboration that more knowledgeable students would typically engage in and thus, could result in disrupted or reduced learning rather than the benefits in comprehension that were seen here. The further effect that the presence of analogies would in turn have on the metacomprehension accuracy of more knowledgeable students is an open question. To date, relatively little research has been done on how and when the use of analogies may support or interfere with learning, and even less has explored how they affect the monitoring of learning. Much more work is needed to understand how these adjuncts may affect both comprehension and comprehension-monitoring processes when students are tasked with learning from expository science texts.

More work is also needed to understand how we can support accurate monitoring. The results of Experiment 3 showed that analogies may only facilitate relative metacomprehension accuracy if readers are specifically prompted to use situation-model cues. The question that cannot be answered from these results is whether a similar level of facilitation would also have occurred in a condition where readers were prompted to use situation-model cues without the analogies being present. It would be interesting to see what might happen in this condition, and whether the presence of the analogies in the texts is adding anything to the equation, such as by supporting efforts to map the analogies to the targets, or to explain the target concepts.

5.2. Conclusions

The results of the current studies show that the presence of analogies in expository science texts can sometimes lead to poor relative metacomprehension accuracy. In both studies where readers reported the basis for their judgments, nearly two-thirds of readers reported using superficial judgment cues. Among these readers, mean relative accuracy on all three relative measures was negative when texts contained analogies. Negative correlations mean that readers did poorest on the topics they thought they had understood the best, which means that they likely would make very poor decisions about which topics to restudy. Without special directions, many readers tend to rely upon invalid cues, and the presence of instructional analogies may mislead readers to make even less accurate judgments of comprehension than they typically do in baseline conditions. Instructional analogies may only facilitate relative metacomprehension accuracy if readers are specifically prompted to engage in activities that help them to be able to make use of the metacognitive affordances of the instructional adjuncts. An embedded analogy may improve metacomprehension by helping to clarify goals or standards for comprehension. Or, an embedded analogy may serve as an example of an explanatory framework that learners

can use to map their understanding onto, to contrast their understanding against, or to detect gaps in their own understanding. However, the benefits of such adjuncts may only be seen when readers are basing their self-assessments on situation-model cues that reflect understanding of the key relations among concepts that will be tested. Taken together, these studies suggest that adjuncts like embedded analogies need to be paired with instructions that prompt readers to consider the quality of their situation-models related to the target phenomena in order to avoid illusions of comprehension, and to support accurate comprehension monitoring.

Appendix

Example Text and Comprehension Questions (with analogy portion in added italics)

Weather Patterns

Weather patterns are governed by the relationship between the movement of air in the atmosphere and the movement of water in the ocean.

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. The standard relationship is that air moves from areas of high pressure to areas of low pressure. This creates wind patterns that move air around the Earth. In the context of the Equatorial Pacific Ocean, pressure is usually higher in the east than it is in the west. Air moves from the higher pressure systems in the Eastern Pacific to the lower pressure systems in the Western Pacific. This air pressure difference is what drives the air from the east to the west and creates a weather pattern called the “trade winds” along the equator.

Imagine that you have just blown up a balloon with air and you are holding the balloon shut with your fingers. You loosen your grip on the mouthpiece so that air begins to come out. What causes the air to come out of the balloon? The air comes out because there is higher air pressure inside the balloon and lower air pressure outside the balloon, and air always moves from areas of higher pressure to areas of lower pressure.

The movement of air also has an impact on movement of the waters of the Pacific. Generally, the Pacific Ocean has many layers of water that are defined by their temperatures. The surface layer is the warmest because there is a concentrated amount of sunlight hitting the surface at the equator. This surface water is what moves in response to the trade winds. As the trade winds blow from east to west they push steadily against the sea for thousands of miles, dragging the surface water along that same path. As a result, warm surface water accumulates in the western Pacific. This bulge of warm seawater can become massive, extending out for many thousands of miles.

Now imagine that you have re-inflated the balloon, and this time you are dangling a sheet of paper in front of the mouthpiece. If you again loosen your grip on the mouthpiece so that air begins to come out, what will happen to the piece of paper? The stream of air coming out of the balloon will push the paper away from the balloon. This is because when you open the balloon, air will move away from areas of higher pressure to areas of lower pressure. The movement of air as illustrated by this balloon example is a central concept in understanding many weather patterns.

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 and a great deal of upward air movement, as the warm, moist air rises.

As the moist air evaporating from the ocean 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 lot of rain cloud formation and precipitation becomes more likely. Thus, the large bulge of warm water in the west increases the likelihood of precipitation in that part of the Pacific. These countries have very humid and tropical conditions. Conversely, in the Eastern Pacific drier conditions are the norm.

The process of *upwelling* is a major reason why the Eastern Pacific has cooler surface water temperatures than the Western Pacific. As the trade winds move surface water from east 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. As a result there is a constant renewal at the surface with cooler water. This is the process of upwelling. Eventually however this surface water will meet the same fate as the water it replaced. It will also get heated by the sun and pushed towards the west by the trade winds. The cooler surface water in the Eastern Pacific inhibits the evaporation of moisture into the air and makes it less likely to rain. This is why the coast of Peru typically experiences drought conditions.

Example test items:

1. A warm water bulge will occur...
 - a. in areas of high air pressure.
 - b. in areas of low air pressure.
 - c. across the Pacific ocean.
 - d. near upwelling.

2. Across the Earth (not just the Pacific Ocean), air always moves from...
 - a. east to west.
 - b. west to east.
 - c. areas with low air pressure to areas with high air pressure.
 - d. areas with high air pressure to areas with low air pressure.

References

- Ackerman, R., & Leiser, D. (2014). The effect of concrete supplements on metacognitive regulation during learning and open-book test taking. *British Journal of Educational Psychology, 84*, 329-348.
- Ainsworth, S., & Loizou, A.T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*, 669-681.
- Alexander, P. A., & Kulikowich, J. M. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Reading Behavior, 23*, 165-190.
- Bean, T. W., Searles, D., & Cowen, S. (1990). Text-based analogies. *Reading Psychology: An International Quarterly, 11*, 323-333.
- Bean, T. W., Singer, H., & Cowan, S. (1985). Analogical study guides: Improving comprehension in science. *Journal of Reading, 29*, 246 – 250.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*, 182-197.
- Curtis, R. V., & Reigeluth, C. M. (1984). The use of analogies in written text. *Instructional Science, 13*, 99–117.
- Donnelly, C. M., & McDaniel, M. A. (1993). Use of analogy in learning scientific concepts. *Journal of Experimental Psychology: Learning, Memory, Cognition, 19*, 975–987.
- Donnelly, C. M., & McDaniel, M. A. (2000). Analogy with knowledgeable learners: When analogy confers benefits and exacts costs. *Psychonomic Bulletin & Review, 7*, 537-543.
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228-232.

- Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, 75, 649-672.
- Fischer, P. M., & Mandl, H. (1984). Learner, text variables, and the control of text comprehension and recall. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 213–254). Hillsdale, NJ: Erlbaum.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34, 906–911.
- Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), *Metaphors: Problems and perspectives* (pp. 106-132). Brighton: Harvester Press.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gilbert, S. W. (1989). An evaluation of the use of analogy, simile, and metaphor in science texts. *Journal of Research in Science Teaching*, 26, 315-327.
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, Cognition*, 11, 702-718.
- Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, 37, 1001–13.
- Griffin, T. D., Mielicki, M. K., & Wiley, J. (in press). Improving students' metacomprehension accuracy. To appear in J. Dunlosky & K. Rawson (Eds.), *Cambridge handbook of cognition and education*. New York, NY: Cambridge University Press.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36, 93-103.

- Griffin, T. D., Wiley, J., & Salas, C. R. (2013). Supporting effective self-regulated learning: The critical role of monitoring. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 19-34). New York, NY: Springer Science.
- Hammadou, J. (2000). The impact of analogy and content knowledge on reading comprehension: What helps, what hurts. *The Modern Language Journal*, *84*, 38-50.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*, 414-434.
- Hegarty, M. (2014). Multimedia learning and the development of mental models. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 673-701). New York, NY: Cambridge University Press.
- Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, *19*, 290-304.
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, *69*, 151-164.
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*, 722-738.
- Iding, M. K. (1997). How analogies foster learning from science text. *Instructional Science*, *25*, 233 – 253.
- Ikeda, K., Kitagami, S., Takahashi, T., Hattori, Y., & Ito, Y. (2013). Neuroscientific information bias in metacomprehension: The effect of brain images on metacomprehension judgment of neuroscience research. *Psychonomic Bulletin & Review*, *20*, 1357-1363.

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy?

Learning & Instruction, 34, 58-73.

Jaeger, A. J., & Wiley, J. (2015). Reading an analogy can cause the illusion of comprehension.

Discourse Processes, 52, 376-405.

Jaeger, A. J., Taylor, A. R., & Wiley, J. (2016). When, and for whom, analogies help: The role of spatial skills and interleaved presentation. *Journal of Educational Psychology, 108*, 1121-

1139.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227-254.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*,

294-303.

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

University Press.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to

judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The*

Journal of the Learning Sciences, 10, 417-446.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words.

Cognitive Science, 11, 65-100.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning.

Instructional Science, 41, 811-831.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology,*

3, 525-551.

- Maki, R. H. (1998). Test predictions over text material. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *10*, 663-679.
- Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, *1*, 126-129.
- Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 509-527.
- Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, *1*, 43-64.
- Mayer, R. E. (2005). Principles for managing essential processing multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 169–182). New York: Cambridge University Press.
- Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin*, *95*, 109-133.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic.

- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*, 53-69.
- Newton, L. D. (2003). The occurrence of analogies in elementary school science books. *Instructional Science, 31*, 353-375.
- Orgill, M. K., & Bodner, G. M. (2006). An analysis of the effectiveness of analogy use in college-level biochemistry textbooks. *Journal of Research in Science Teaching, 43*, 1040-1060.
- Paris, N. A., & Glynn, S. M. (2004). Elaborate analogies in science text: Tools for enhancing pre-service teachers' knowledge and attitudes. *Contemporary Educational Psychology, 29*, 230-247.
- Potter, M.C. (1975). Meaning in visual search. *Science, 187*, 965-966.
- Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*, 1004-1010.
- Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning & Instruction, 22*, 262-270.
- Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction, 14*, 373-408.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*, 344-355.
- Sanchez, C. A., & Wiley, J. (2010). Sex differences in science learning: Closing the gap through animations. *Learning and Individual Differences, 20*, 271-275.

- Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief that diagrams improve learning from text. *Memory, 18*, 698-711.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*, 331-362.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D.J. Hacker, J. Dunlosky, & A.C. Graesser, (Eds.) *Handbook of metacognition and self-regulated learning* (pp. 85-106). New York, NY: Routledge.
- Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81*, 264-273.
- Thiele, R. B., & Treagust, D. F. (1994). The nature and extent of analogies in secondary chemistry textbooks. *Instructional Science, 22*, 61-74.
- Weaver, C. A., Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C.A. Weaver, S. Mannes, S., & C. Fletcher (Eds.), *Discourse comprehension. Essays in honour of Walter Kintsch* (pp. 177-193). Mahwah, NJ: Erlbaum.
- Wilbers, J., & Duit, R. (2006). Post-festum and heuristic analogies. In P.J. Aubusson, A.G. Harrison, & S.M. Ritchie (Eds.), *Metaphor and analogy in science education*, (pp. 37-49). Springer: Netherlands.
- Wiley, J., Goldman, S., Graesser, A., Sanchez. C., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal, 46*, 1060-1106.

- Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., & Thiede, K. W. (2016). Improving metacomprehension accuracy in an undergraduate course context. *Journal of Experimental Psychology: Applied*, *22*, 393-405.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, *132*, 408-428.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2008). To understand your understanding one must understand what understanding means. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Washington D.C.
- Wiley, J., Sarmiento, D., Griffin, T. D., & Hinze, S. R. (2017). Biology textbook graphics and their impact on expectations of understanding. *Discourse Processes*, *54*, 463-478. DOI: 10.1080/0163853X.2017.1319655
- Wiley, J., Thiede, K. W., & Griffin, T. D. (2016). Improving metacomprehension with the situation-model approach. In Mokhtari, K., (Ed). *Improving reading comprehension through metacognitive reading instruction for first and second language readers* (pp. 93-110). Lanham, MD: Rowman & Littlefield.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*, 301-311.

Table 1.

Means and Standard Errors for Judgment Magnitudes, Test Performance, Absolute Accuracy, and Relative Metacomprehension Accuracy across Conditions in the Experiment 1

Cue-Type	No Analogy		Analogy	
	Superficial	SM-cues	Superficial	SM-cues
N	25	14	25	19
Judgment Magnitude	3.02 (.17)	2.84 (.22)	3.01 (.17)	3.05 (.19)
Test Performance	2.11 (.12)	2.09 (.16)	2.28(.12)	2.06 (.14)
Absolute Accuracy	1.55 (.11)	1.41 (.16)	1.48 (.11)	1.33 (.10)
Relative Accuracy Pearson	.20 (.10)	.19 (.14)	-.13 (.10)	.36 (.12)
Relative Accuracy Gamma	.17 (.13)	.30 (.17)	-.18 (.13)	.49 (.15)
Relative Accuracy d_a	.18 (.35)	.15 (.47)	-.83 (.35)	.75 (.41)

Note: SM-cues = Situation-Model cues

Table 2.

Means and Standard Errors for Judgment Magnitudes, Test Performance, Absolute Accuracy, and Relative Metacomprehension Accuracy across Conditions in Experiment 2

Cue-Type	No Analogy		Analogy	
	Superficial	SM-cues	Superficial	SM-cues
N	33	17	30	18
Judgment Magnitude	2.50 (.15)	2.91 (.19)	2.79 (.15)	2.71 (.20)
Test Performance	2.26 (.12)	2.21 (.16)	2.62 (.12)	2.58 (.17)
Absolute Accuracy	1.06 (.07)	1.22 (.10)	1.19(.08)	1.05 (.10)
Relative Accuracy Pearson	.19 (.09)	.21 (.12)	-.12 (.07)	.35 (.08)
Relative Accuracy Gamma	.24 (.12)	.32 (.15)	-.22 (.10)	.43 (.12)
Relative Accuracy d_a	.20 (.30)	.41 (.41)	-.85 (.31)	.68 (.40)

Note: SM-cues = Situation-Model cues

Table 3.

Means, Standard Errors, and Results of Statistical Tests for Judgment Magnitudes, Test Performance and Metacomprehension Accuracy on Analogy Texts across conditions prompted to use Superficial and Situation-Model (SM) Cues in Experiment 3

	Superficial	SM-cues	<i>F</i>	<i>p</i>	η_p^2
N	50	50			
Judgment Magnitude	2.57 (.10)	2.84 (.14)	2.55	.11	.03
Test Performance	2.14 (.09)	2.35 (.09)	2.67	.11	.03
Absolute Accuracy	1.20 (.05)	1.17 (.08)	.11	.74	.00
Relative Accuracy Pearson	.06 (.06)	.34 (.04)	16.11	.0001	.14
Relative Accuracy Gamma	.07 (.08)	.18 (.08)	1.00	.32	.01
Relative Accuracy d_a	.06 (.23)	.34 (.18)	.91	.34	.01

Note: SM-cues = Situation-Model cues