Reading an Analogy Can Cause the Illusion of Comprehension

Allison J. Jaeger and Jennifer Wiley

Department of Psychology
University of Illinois at Chicago

This study explored students’ ability to evaluate their learning from a multimedia inquiry unit about the causes of global climate change. Participants were 90 sixth grade students from four science classrooms. Students were provided with a text describing the causes of climate change as well as graphs showing average global temperature changes. Half of the students also received an analogy to help support their understanding of the topic. Results indicated that overall students were overconfident about how much they learned and how well they understood the topic. Further, the presence of an analogy led to higher levels of overconfidence. Results also indicated that students with better graph interpretation skills were less overconfident even when the analogy was present. These results suggest that the presence of graphs and analogies can negatively affect students’ abilities to accurately judge their own level of understanding and may lead to an illusion of comprehension.

INTRODUCTION

With the new common core standards in the United States (National Research Council, 2012), one learning activity that is becoming more popular in science classes is asking students to answer inquiry questions, such as “What caused the eruption of Mt. St. Helens in 1986?” and “How and why are recent patterns of global temperatures different than in the past?,” from internet and printed sources
including both text and graphics (Bråten & Strømsø, 2010; Cerdan & Vidal-Abarca, 2008; Mason, Ariasi, & Boldrin, 2011; Rouet & Britt, 2011; Wiley et al., 2009). To engage in effective learning from these activities, students also need to be able to monitor and evaluate the quality of their understanding. Accurate metacognition is important because it influences self-regulated learning by guiding future learning and studying behaviors (Graesser et al., 2007; Griffin, Wiley, & Salas, 2013; Thiede, Anderson, & Therriault, 2003; Winne & Hadwin, 2008). Unfortunately, previous research has indicated that students are generally inaccurate in their ability to monitor and evaluate their own learning (Dunlosky, Rawson, & Middleton, 2005; Maki, 1998; Nelson & Dunlosky, 1991). Thus, it remains an important question as to what conditions are more likely to cause students to engage in more or less effective metacognition. The present study aimed to investigate how accurate students are in judging their learning from a multimedia unit including text and graphs. In addition, it tested whether the presence of an instructional analogy or whether the possession of graph interpretation skills might support more accurate evaluation of understanding.

What Defines Accurate Metacomprehension?

One important aspect of metacognition during reading is the subjective assessment of one’s own cognitive state (Flavell, 1979). Although metamemory accuracy refers to a person’s ability to judge his or her ability to remember information, metacomprehension accuracy refers to a person’s ability to judge his or her own learning or understanding of conceptual material (Dunlosky & Lipko, 2007; Maki, 1998; Maki & Berry, 1984; Maki, Shields, Wheeler, & Zacchilli, 2005; Wiley, Griffin, & Thiede, 2005). There are two main approaches to computing metacomprehension accuracy that both compare predictive judgments with actual performance.

Absolute accuracy refers to the difference between a student’s predicted performance and that student’s actual performance on a single test (Griffin, Jee, & Wiley, 2009; Maki, 1998; van Loon, de Bruin, van Gog, & van Merrienboer, 2013). The difference between a student’s judgment of learning and actual test score is computed and provides insight into the match or calibration that exists between perception and performance. The smaller the difference between self-judgments and actual performance, the less error there is in the judgments. Computing the signed difference between perceived performance and actual performance is referred to as confidence bias. A difference in the positive direction indicates that the student is overconfident, whereas a difference in the negative direction indicates the student is under confident. Previous research has indicated that students may have a general tendency to be overconfident (Hacker, Bol, Horgan, & Rakow, 2000; Koriat, 2011; Lipko et al., 2009). One study on eighth graders showed they overestimated their actual test performance by over
40% (Lipko et al., 2009). Overconfidence can be problematic because it may lead students to discontinue studying prematurely if they falsely believe they already understand and can explain the phenomena they are researching (Rawson & Dunlosky, 2012).

Another commonly used measure of metacomprehension accuracy is referred to as relative accuracy. Relative metacomprehension accuracy is computed as an intraindividual correlation between readers’ judgments of learning for each text in a set of texts and their actual performance on each test (Nelson, 1984). Correlations can range from \(-1\) to \(+1\), with correlations near 0 or below representing chance to poor accuracy. Correlations near \(+1\) indicate very good discrimination between texts or topics that one has understood well from those one has not. Research with this measure has also generally shown that students have poor monitoring accuracy. Typical correlations between predicted test performance and actual test performance are quite low, only around \( .27 \) (Dunlosky & Lipko, 2007; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009). Poor relative accuracy can be problematic because it may lead students to study material they already know fairly well at the expense of topics they actually need to spend more time on.

When only a single text or topic is the basis of study, then it is difficult to compute relative accuracy measures (which require independent sets of texts and tests for each individual). Although a great deal of research on metacomprehension accuracy has used relative measures and although relative measures have distinct advantages in assessing online monitoring accuracy (see Griffin et al., 2013; Nelson, 1984; Wiley et al., 2005), in the present case of learning from a single multimedia unit, only absolute accuracy measures were able to be computed. Thus, absolute accuracy is examined as the measure of metacomprehension accuracy in the present study.

However, an advantage of using measures of absolute accuracy is that it actually connects this work to yet another tradition of research on “illusions of understanding.” This area of work has shown that people are especially inaccurate at gauging their own understanding of explanatory knowledge compared with other kinds of knowledge, including procedures, narratives, and facts (Mills & Keil, 2004). Mills and Keil had participants rate their understanding of how different items work on a seven-point scale (e.g., sewing machine, cylinder lock, helicopter), with 7 representing a deep understanding and 1 representing a shallow understanding. After making their ratings, participants wrote detailed casual explanations for the phenomena or devices. An independent group of participants evaluated the quality of these generated explanations on the same seven-point scale using an expert explanation as a comparison. Participants’ self-ratings of understanding were compared with the ratings given by the independent raters and indicated that participants were generally overconfident in their self-ratings. In Mills and Keil’s second experiment,
participants rated their knowledge about procedures (e.g., how to make a pizza, how to carve a pumpkin), then wrote a description of that procedure, and read an expert’s description of that procedure, then rerated their knowledge of that procedure. Results indicated that participant’s ratings for the procedural information was more accurate and not overconfident.

Mills and Keil (2004) refer to this overconfidence in the ability to generate an explanation as the result of an “illusion of understanding.” They suggest several reasons why illusions of explanatory understanding are stronger for explanatory information than other kinds of information. First, explanatory knowledge, such as understanding how and why a natural phenomenon such as an earthquake occurs, is more causally complex than other types of knowledge and can be layered such that one can develop deeper and deeper understandings. Another reason is that many complex phenomena cannot be perceived, such as tectonic plates interacting when an earthquake occurs. This can lead to inaccurate judgments of understanding because many of the causal connections cannot be inferred through direct observation. Finally, Mills and Keil (2004) suggest that gauging understanding of explanatory knowledge is difficult because people do not have experience in self-testing this type of knowledge. In the case of procedural knowledge, you may have experienced that procedure once before and can call on that prior experience and test your current experience against it when making your judgment. However, they suggest that people may have less experience with gauging the quality of the explanations they give to others.

Possible Effects of Graphics on Metacomprehension Accuracy

Although most of the work on metacomprehension has focused on purely text-based materials, a great deal of the material that students encounter when researching or learning about a new topic is multimedia, with text usually accompanied by graphics. In many science domains, visualizations such as diagrams, schematics, or graphs are used in an attempt to support better understanding of complex phenomena (e.g., Butcher, 2006). For example, research has shown that over half of the space in middle school science textbooks is used for images or illustrations (Mayer, 1993). Although a fairly substantial amount of research has explored how and when providing visualizations such as graphs and diagrams may affect learning from text materials (Butcher, 2006; Hegarty & Kozhevnikov, 1999; Hegarty & Sims, 1994; Mayer, 2005; Tversky, 1995, 2001), much less is known about how the presence of these visualizations may affect students’ ability to accurately judge their own understanding.

One study conducted by Serra and Dunlosky (2010) did attempt to address this question and found that in general students believe they learn better from multimedia compared with text-only presentations and this belief affects the magnitude of their judgments of learning. In their study, three groups of college
students were asked to read a text about lightning formation, one in which the text was paired with diagrams, one in which the text was paired with photographs of lightning, and one group in which the text was presented alone. Participants’ beliefs about learning from multimedia and initial predictions of their learning about lightning from a text that either did or did not contain images (depending on condition) were obtained before reading. Then, while reading the six-paragraph text, participants were prompted to make a judgment of learning after each paragraph, ending with a final poststudy global judgment. Finally, participants completed a memory test.

Responses to the prereading questionnaire indicated that all participants strongly endorsed the belief that multimedia produces better learning than single media. In addition, regardless of whether predictions were made before, during, or after reading, participants gave lower judgments of learning in plain text conditions than either of the illustrated conditions. Importantly, test performance was only significantly better in the diagram condition; the photos did not improve test performance over the plain text condition. From these results, Serra and Dunlosky (2010) suggested that beliefs about multimedia learning act as a heuristic that biases judgments and that the use of this multimedia heuristic could lead to reduced monitoring accuracy in situations where it is invalid. Although they did not report the relation of judgments to test performance in terms of absolute accuracy and no differences were found in relative accuracy, they argued that because judgments did not differ between the two image conditions, but learning did, that students were likely using the multimedia heuristic instead of more valid processing cues as the basis for their judgments of learning. This observed effect of the mere presence of images on the judgment process would suggest that students could experience reduced monitoring accuracy when texts contain images, and this would be particularly likely to be observed in conditions where those images do not actually improve learning (Serra & Dunlosky, 2010).

Another recent study more directly investigated the effect that different types of visualizations may have on relative metacomprehension accuracy (Jaeger & Wiley, 2014). In their study, Jaeger and Wiley had participants read a set of five texts each describing the processes involved in a different scientific phenomenon (e.g., volcanoes, ice ages, evolution). The texts were either paired with no images, conceptual diagrams, or decorative photographs. After reading each text, participants were prompted to make a judgment asking them how well they believed they could do on a test of the material and then completed comprehension tests. Based on prior work, it seemed the presence of conceptual diagrams could provide readers with an additional basis for evaluating their comprehension or “self-testing” that would increase metacomprehension accuracy compared with the no-image condition. On the other hand, a detriment was predicted in the decorative image condition because it was believed the presence of these types of images would increase readers’ access to less valid
cues for judging their comprehension such as interest and enjoyment. Results indicated that metacomprehension accuracy was in fact harmed by the presence of the decorative images, but the presence of the conceptual images did not significantly improve metacomprehension accuracy beyond that of the no-image condition.

One possible reason for the general lack of benefit from conceptual diagrams in this study could be because the students were unable to understand or interpret the diagrams and therefore were unable to use them to help evaluate their comprehension. Although both studies did attempt to address some effects that multimedia units could have on metacomprehension accuracy, neither of them took into consideration individual differences in the skills that students may have for interpreting the information in the graphs or how these skills could affect the accuracy of monitoring judgments when learning from multimedia. Many inquiry-based science units involve the construction and interpretation of a variety of visual representations of data including tables and graphs (Bowen & Roth, 2005). Considering the role of graph comprehension skills in the context of multimedia comprehension is important because previous research has demonstrated that students may struggle to understand or interpret graphs (Henderson, 1999). For example, some people may be disadvantaged because they may not be able to interpret the information in the images or may have difficulty generating inferences from images (Canham & Hegarty, 2010; Tversky, Zacks, Lee, & Heiser, 2000). Additionally, some research has indicated that students may struggle to integrate information across text and the illustrations (Hegarty & Just, 1993). As a result, one goal of the current study was to investigate the role that graph interpretation skills may play in being able to accurately monitor one’s learning from a combination of textual and graphical information.

Possible Effects of Analogy on Monitoring Accuracy

In addition to graphics, analogies are often included in science text as another device intended help students to understand complex scientific concepts and promote mental model generation (Gentner & Holyoak, 1997; Venville & Treagust, 1996). They are used in particular to help explain difficult or unfamiliar topics (Duit, 1991; Glynn, Britton, Semrud-Clikeman, & Muth, 1989), especially needed when the phenomena being explained cannot be perceived (Jee et al., 2010; Sibley, 2009). Like graphics, analogies have also been documented to appear often in science textbooks (Curtis & Reigeluth, 1984; Orgill & Bodner, 2006). Additionally, students report finding analogy-embedded texts more understandable and interesting (Paris & Glynn, 2004).

The purpose of an analogy is to provide a more familiar exemplar phenomenon, process, or system that helps the student to understand how a novel set of relations works. The analogy thus provides a basis for understanding onto
which the new material can be mapped. From one perspective, one could predict that the presence of an instructional analogy could improve the accuracy of students’ judgments. As mentioned earlier, it has been suggested that many students struggle to accurately gauge their own understanding of complex science texts because the processes described cannot be perceived and also because students have not had direct experience with the phenomenon before reading about it (Mills & Keil, 2004). By providing students with an instructional analogy, the phenomenon that was once invisible and hard for students to visualize could become more perceptible. Similarly, if the instructional analogy provided is something students have prior experience with, it can more easily be used for testing one’s own understanding of the phenomenon. By testing one’s own understanding, students could be more likely to detect when they have not completely understood something. Additionally, analogies can be useful because the important mechanisms underlying the phenomenon are being highlighted and compared across two different domains, which may allow students to realize gaps in their knowledge. Based on these ideas, the presence of an instructional analogy could help to improve the accuracy of students’ judgments.

Yet, although analogies are generally included within science units in an attempt to improve learning and understanding, research on the effectiveness of analogy has shown mixed results. Some studies have shown that a text with an embedded analogy can facilitate learning compared with a no-analogy comparison text (Brown, 1992; Donnelly & McDaniel, 1993; Glynn & Takahashi, 1998; McDaniel & Donnelly, 1996). On the other hand, some work has found no facilitative effects for analogies (Alexander & Kulikowich, 1991; Braasch & Goldman, 2010), and in fact some work has even demonstrated detrimental effects on text comprehension and recall (Donnelly & McDaniel, 1993; Yanowitz, 2001; Zook & Maier, 1994).

Based on these mixed results, one could alternatively predict that the presence of an analogy could harm judgment accuracy. Although many researchers and educators hold the view that increasing interest and motivation in a topic is important for learning (Hidi, 1990), the presence of an analogy could detract from understanding by giving students an illusion they have understood something when in fact they have not. The presence of an analogy could be increasing readers’ access to less helpful cues that are based on their feelings about the text material or giving them a false sense of fluency if they consider their understanding of the familiar example. For example, if the presence of an analogy makes students enjoy the text more, then they could be basing their judgments on this feeling of enjoyment rather than their own understanding, resulting in less accurate judgments. Similarly, if analogies make novel phenomena seem more familiar and easy to understand, this could increase students’ access to less helpful cues such as accessibility and ease of processing, again leading to less accurate metacomprehension. Alternatively, students may hold a belief that
analogies improve learning, and the use of this heuristic could lead to poor metacomprehension accuracy.

Aims for the Current Study

Although prior research has investigated how the presence of analogies might affect learning from science text, much less is known about how they might affect students’ ability to accurately judge their own learning. Additionally, prior research has demonstrated that visualizations can affect student’s judgment accuracy, but no work has looked at how individual differences in graph comprehension skills might affect being able to make accurate judgments about learning from multimedia units. The purpose of this study was to test whether the presence of an analogy within a multimedia unit about the cause of global temperature change would affect the accuracy of sixth grade students’ judgments of learning. Additionally, this study aimed to test whether having skills in graph interpretation would help students to make more accurate judgments. Although most work in the field has used college-aged students in studies of metacomprehension accuracy, this study joins a smaller body of work that has begun to explore this topic in younger students (de Bruin et al., 2011; Redford, Thiede, Wiley, & Griffin, 2012; Thiede, Redford, Wiley, & Griffin, 2012). Because of the young age of the sample, it was especially important to measure comprehension of the unit using both a production task (an explanatory essay) and a recognition task (inference verification) so students could more fully demonstrate their understanding of what they read.

METHODS

Participants

Participants in this study were 90 sixth grade students from four science classes in an urban public middle school in the United States. The average age was 11.81 years (SD = .47). The sample was 44% female and 57% native English speakers. Self-reported ethnicity was 90.0% Hispanic, 14.4% White, 1.1% Asian, 1.1% African American, and 6.7% other (students were able to select multiple ethnicities). There were 45 students each in the analogy and no-analogy conditions.

Because of the Family Educational Rights and Privacy Act, official standardized test scores could not be obtained. To obtain a proxy for reading skill in lieu of test scores, teachers were asked to indicate each student’s level of reading skill relative to their grade level as low (1), medium (2), or high (3).

As shown in Table 1, the two conditions did not differ in proportion of girls, proportion of non-native English speakers, proportion of students who reported
learning about the topic previously in school, teacher-rated reading skill, self-rated interest in science (on scale from 1 to 5), or spatial ability (as measured by a 10-item paper folding test described below), all \( t < 1 \). There was also no difference in self-rated topic knowledge (on scale from 1 to 5), \( t(88) = 1.35, ns \).

The analogy group tended to have higher graph interpretation skills as measured by the number of important features they identified on the graphs, although this effect was marginal, \( t(88) = 1.91, p < .06 \). Also, because reading skill was related to graph interpretation skill, \( F(1, 81) = 4.37, p < .04 \), reading skill was entered as a covariate in the multivariate analyses.

### Materials and Procedure

**Reading and writing inquiry activity.** Students participated in the inquiry activity as part of their normal science classes. All materials for the inquiry activity were distributed to the students in folders, including the task essay prompt, blank writing pages, and the multimedia unit. Students were asked to read along as the inquiry task essay prompt instruction was read out loud. The full instructions for the Reading and Writing task were as follows:

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Descriptives and Measures as a Function of Analogy Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Frequencies</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion female participants</td>
<td>Analogy</td>
</tr>
<tr>
<td>.40</td>
<td>.49</td>
</tr>
<tr>
<td>Proportion native-English speakers</td>
<td>.53</td>
</tr>
<tr>
<td>Proportion taught topic before</td>
<td>.67</td>
</tr>
<tr>
<td>Words in text</td>
<td>1,088</td>
</tr>
<tr>
<td><strong>Measures</strong></td>
<td>M</td>
</tr>
<tr>
<td>Self-report prior knowledge (1–5)</td>
<td>3.33</td>
</tr>
<tr>
<td>Self-report science interest (1–5)</td>
<td>4.23</td>
</tr>
<tr>
<td>Self-report task enjoyment (1–6)</td>
<td>4.78</td>
</tr>
<tr>
<td>Paper folding score (0–10)</td>
<td>5.22</td>
</tr>
<tr>
<td>Teacher-reported reading skill (1–3)</td>
<td>2.02</td>
</tr>
<tr>
<td>Graph skill scores (no. features identified)</td>
<td>4.42</td>
</tr>
<tr>
<td>Words in explanatory essay</td>
<td>161.31</td>
</tr>
<tr>
<td>Connections in explanatory essay</td>
<td>.42</td>
</tr>
<tr>
<td>Comprehension test prediction (0–12)</td>
<td>9.31</td>
</tr>
<tr>
<td>Explanation prediction (0–10)</td>
<td>7.64</td>
</tr>
<tr>
<td>Comprehension test score (0–12)</td>
<td>6.58</td>
</tr>
<tr>
<td>Explanatory essay score (0–5)</td>
<td>1.02</td>
</tr>
<tr>
<td>Confidence bias for test</td>
<td>.23</td>
</tr>
<tr>
<td>Confidence bias for explanation</td>
<td>.56</td>
</tr>
</tbody>
</table>
The primary purpose of reading in science is to understand the causes of scientific phenomena. This means your goal for reading is to understand how and why things happen. In today’s task your goal is to learn more about the global climate system by reading about the greenhouse effect. You will eventually write an essay explaining how and why recent patterns in global temperature are different from what has been observed in the past. Before you start working on your essay, please take fifteen minutes to read through the materials.

Global temperature change multimedia unit. All participants were provided with information related to the causes of global temperature change, based on materials previously used in studies with middle school students (Griffin, Wiley, Britt, & Salas, 2012). The base text was approximately 1,000 words with an average Flesch Reading Ease score of 51.5. It included discussion of four main topics: the carbon cycle, the greenhouse effect, ice ages, and energy from fossil fuels. The material also included two graphs, one showing temperatures over the last 400,000 years and one showing data from the last 150 years. The text was either presented as one long unit with each section following the preceding one with no page breaks or as a series of pages with each main section of the unit appearing on its own page. The graphs were always presented on their own page. This presentation manipulation was fully nested within the analogy manipulation, but it had no effect on any measures in this study (all Fs < 1). All subsequent analyses collapse across presentation conditions to maximize power for testing for effects of the analogy manipulation. The multimedia unit was presented on paper, and students had access to it while writing their essays. See Appendix A for the materials.

Analogy manipulation. The main manipulation for this study was the inclusion of an analogy within the unit for half of the participants. The analogy consisted of 261 words and described the effects of increasing greenhouse gasses as being similar to rolling up the windows of a car parked in the sun (see Appendix A for complete text of the analogy, presented in italics). Half the participants received this analogy and half did not. This manipulation was randomly assigned within classes.

Judgments of learning. After reading through the unit, participants were asked to make two predictive judgments. First, they were asked if given a test on the greenhouse effect with 12 questions to predict how many they believed they could get correct on a scale from 1 to 12. Second, they were asked to indicate on scale from 1 to 10 how well they would be able to explain to another person how and why recent patterns in global temperature are different from what has been observed in the past (1, not very well; 10, extremely well).
Explanatory essay task. After making their predictive ratings, students were given this prompt to write an explanatory essay: “How and why are recent patterns in global temperatures different from what has been observed in the past? In your essay, you should include specific information and evidence from the reading to support your conclusions and ideas. Also, be sure to make connections among ideas.” Students were given approximately 40 minutes to write their essays. Students had access to the multimedia unit as they wrote.

Postessay booklet. After writing their explanatory essays, all materials and essays were collected, and each participant completed a paper-and-pencil booklet. At the start of the booklet, students responded to 12 comprehension test items. For each item they indicated whether each statement, such as “In the past 100 years, both fossil fuel use and carbon dioxide levels have increased,” seemed consistent with the reading material (see Appendix A for the complete set of verification items). These statements represented potential connections between the important causal concepts in the reading. Importantly, none of these connections was made explicitly by the text (cf., Royer, Carlo, Dufrense, & Mestre, 1996; Sanchez & Wiley, 2006). They provide an index of comprehension that tested for the verification of causal inferences. Because this task is a recognition task rather than a production task, it provides an important additional source of information about understanding that may be especially critical with younger students. The booklet ended by asking for basic demographic information (gender, age, ethnicity).

In addition, all participants were asked to indicate whether or not they had been taught about this topic before in school and rated how interested they were in science and how much they knew about the topic before reading the material on a five-point Likert scale, with 5 meaning more interest and more knowledge. They also indicated how much they enjoyed this type of task, compared with their normal science assignments on a six-point Likert scale, with 6 meaning more enjoyable, and wrote an open-ended response to the question “what did you find difficult about this task?”

Graph interpretation task. After completing the booklet, students completed a graph interpretation task in which they were re-presented the graphs they originally saw in the multimedia unit. In this task they were asked to circle features of the graphs they believed were important for being able to understand that recent patterns in global temperatures have changed and explain why they thought these features were important. They were also asked to write a summary of the overall meaning of the graph.

Paper folding task. The last task students completed was a shortened 10-item, paper-and-pencil version of the paper-folding test from the kit of
factor-referenced cognitive tests (Ekstrom, French, Harman, & Derman, 1976) to serve as a measure of spatial ability. Participants’ scores were computed as the number of correct responses. The paper folding task is a very commonly used measure in the psychometric literature. It has been included in numerous factor analytic studies and has consistently fallen into the spatial visualization factor (Carroll, 1993; Lohman, 1979; McGee, 1979). It has been shown to be a reliable test with estimates of reliability ranging between .75 and .89 (Kane et al., 2004; Kozhevnikov & Hegarty, 2001; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990). It is also highly correlated with other spatial measures, including hidden figures, paper form board, and surface development (Kozhevnikov & Hegarty, 2001; Marshalek, Lohman, & Snow, 1983). The full activity was completed over two 60-minute periods on back to back days.

Coding and Scoring of Data

Explanation essay scoring. Explanation scores were derived by coding students’ essays for the presence of five critical causal concepts relating to climate change based on an a priori causal model used in prior research with this topic (Griffin et al., 2012). These concepts were as follows:

1. We are in an unusually long warming period.
2. CO2 levels in the atmosphere are at their highest in 400,000 years.
4. CO2 is a greenhouse gas.
5. Greenhouse gases in the atmosphere cause warming.

All essays were evaluated for the presence of these target concepts and the number of explicit causal connections (Wiley & Voss, 1999) by two independent coders, who produced a high level of inter-rater reliability (Krippendorf’s $\alpha = .82$, $p < .05$). Any differences were resolved through discussion. As shown in Table 1, the average number of connections made in both the analogy and no-analogy groups was below 1, and no differences were observed between groups, $t < 1$. Based on the low incidence in explicit connections in the essays, the number of concepts students included served as the measure of explanation quality. This can be thought of as an index of coverage of important causal concepts in the essays. Proportion scores were computed by dividing the number of concepts mentioned by the total available (5).

Each essay was also coded for whether or not essays mentioned the analogy and whether students included any direct quotes from the text or referenced any documents or source of information. Essays with any references were assigned
a 1; otherwise, they received a 0. Two independent coders had 100% reliability on this coding.

**Comprehension test scoring.** Comprehension test scores were computed as the proportion correct on the 12-item verification test.

**Graph interpretation scoring.** Students identified anywhere from zero to six important features on the graphs (title, source, axis labels, measurement units, data trends, and relations between the graphs). Any students that interpreted the overall pattern or trends in the data or identified more than four of the important features were categorized as having high graph interpretation skills. All graphing tasks were scored by two independent coders, who produced a high level of inter-rater reliability (Krippendorf’s α = .95, p < .05). A total of 45 students fell into each group. Students in the high in graph interpretation skills group noticed more important features (M = 5.38, SD = .98) than those in the low group (M = 3.30, SD = .46).

**Annotation coding.** All materials were evaluated for whether or not students made annotations on them during the experiment. Only 10 of 90 students made any annotations on the materials. Because annotations were so infrequent, they were adequately captured by just three category codes (two independent coders had 100% reliability on this coding): (1) underlining, highlighting, or circling text; (2) marking on the graphs; and (3) writing comments or drawing in margins/summarizing/interpreting ideas.

**Computation of Confidence Bias**

To compute confidence bias measures, both the explanation and test prediction ratings were converted to proportions of the Likert scale. Following Maki (1998), confidence bias (sometimes referred to as over/under confidence because it indicates the direction of people’s misjudgments) was computed as the signed difference between predicted and actual performance. Predictive explanation ratings were compared with performance on the explanatory essays. Predictive test ratings were compared with performance on the verification tests.

**RESULTS**

**Effects of Analogy and Graph Interpretation Skills on Learning**

The main focus of this study was to examine how the presence of an analogy, or individual differences in graph interpretation skills, might affect
metacomprehension accuracy for learning from a multimedia unit. However, before proceeding to examining the difference between judgments and performance, the means for the performance measures are reported in Figure 1. Following Wiley et al. (2009), and because the essay (production) scores and verification (recognition) test scores are both needed to provide a complete picture of student comprehension, the effects of analogy were analyzed on both measures using a 2 (analogy: present, absent) × 2 (graph skills: high, low) × 2 (performance basis: test, essay) MANCOVA, with reading skill as a covariate. Results revealed a significant main effect for analogy such that student scores were lower overall when the instructional analogy was present than when it was not, $F(2, 79) = 5.06, p < .01$, partial $\eta^2 = .11$. Further, there was a main effect for graph interpretation skills such that students with greater graph interpretation skills scored higher than students with lower graph interpretation skills, $F(2, 79) = 3.50, p < .04$, partial $\eta^2 = .08$. The interaction between the conditions was not significant, $F(2, 79) = 1.80, p = .17$. Follow-up tests for each measure and the effect sizes and direction of effects in terms of conditions are reported in Appendix B.

Effects of Analogy and Graph Interpretation Skills on Predictions

Additionally, it was important to know if the presence of the analogy and students’ graph interpretation skills affected overall judgment magnitudes. Mean judgments
are shown in Table 1. A 2 (analogy: present, absent) × 2 (graph skills: high, low) ANCOVA looking at the magnitude of students’ test predictions found no significant effects of either the presence of the analogy or students’ graph interpretation skills (all $F$s < 1). Similarly, although students in the analogy condition gave slightly higher ratings for how well they would be able to explain global warming to others, a 2 (analogy: present, absent) × 2 (graph skills: high, low) ANCOVA on explanation ratings revealed no significant effects (all $F$s < 1).

Effects of Analogy and Graph Interpretation Skills on Metacomprehension Accuracy

The main question for the current study was whether the presence of an analogy and graph interpretation skills might improve the metacognitive accuracy of middle school students. The means for confidence bias in each condition are shown in Figure 2. To test how the presence of an analogy and students’ graph interpretation skills affected the accuracy of their judgments, a 2 (analogy: present, absent) × 2 (graph skills: high, low) × 2 (judgment basis: test, explanation) repeated-measures ANCOVA was performed. The test revealed a significant main effect for judgment basis indicating that students were more overconfident in their ability to explain the concepts to a friend as compared with
how well they thought would do on a test of the concepts, $F(1, 80) = 5.62, p < .02$, partial $\eta^2 = .07$. The analysis also revealed a main effect for analogy such that the presence of the analogy exacerbated students’ overconfidence, $F(1, 80) = 6.25, p < .02$, partial $\eta^2 = .07$. Additionally, a significant main effect for graph interpretation skills suggested that students with lower graph interpretation skills were more overconfident than students with high graph interpretation skills, $F(1, 80) = 4.94, p < .03$, partial $\eta^2 = .06$. The interaction between analogy and graphing skills was not significant, $F < 1$.

We were also interested in investigating the role that more general spatial skills might play in being able to accurately evaluate learning from a unit including both text and graphs. A regression model including analogy condition, graph interpretation skills, and paper folding scores significantly predicted explanation monitoring accuracy ($F(3, 86) = 3.15, p < .03$). Consistent with the analysis presented above, the presence of the analogy predicted confidence bias for explanations ($\beta = .24, t = 2.27, p < .03$), as did graph interpretation skill ($\beta = -.21, t = 2.02, p < .05$). Although there was no significant independent effect for paper folding scores ($\beta = .14, t = 1.39, p = .17$), higher paper folding scores tended to lead to more overconfidence.

**How Did Analogies Affect Performance?**

To gain additional insight into how the analogies might have affected students’ learning, several other aspects of their performance were examined. First, as shown in Table 1, there were no differences in the length of essays between students in the analogy condition and the no-analogy condition, $t < 1$.

Second, differences between students who explicitly mentioned the analogy in their essay versus all other students were investigated. Only 12 students explicitly mentioned the analogy in their essays. Essays that included information about the analogy did not differ from essays without mention of the analogy in terms of essay length when including the words related to the analogy (with mention of analogy: $M = 177.50, SD = 95.42$; no mention of analogy: $M = 155.21, SD = 52.30$), $t(88) = 1.21, ns$. Similarly, there was no difference in essay length when words related to the analogy were excluded (with mention of analogy: $M = 137.83, SD = 85.68$), $t < 1$. Essays including information about the analogy also did not differ from essays without information about the analogy in terms of number of concepts included (with mention of analogy: $M = 1.25, SD = 1.06$; no mention of analogy: $M = 1.15, SD = 1.11$) or the number of connections (with mention of analogy: $M = .67, SD = .78$; no mention of analogy: $M = .46, SD = .68$), all $ts < 1$.

Third, students’ ratings of enjoyment were examined. As shown in Table 1, there was a nonsignificant trend toward students in the analogy condition reporting enjoying the task more, $t(88) = 1.42, p < .16$. 
Fourth, student’s responses to the open-ended item asking them to describe what they found difficult about the task were organized into two main categories. One category represented responses that offered features of the task that were challenging (e.g., task instructions were hard to understand, they had never done a task like this before, it was hard to explain the answer), whereas the second category represented responses that were less meaningful and unrelated to the inquiry task (e.g., students simply restated the inquiry prompt, said the test questions were hard, or said nothing was hard). Students in the no-analogy condition were more likely to provide informative responses about what made the task challenging (66% in no analogy vs. 34% in analogy), whereas students in the analogy condition gave more “nothing was hard,” empty responses, or uninformative responses (59% in analogy vs. 41% in no analogy), $\chi^2 (1, n = 80) = 4.38, p = .04$. This result, taken together with the poorer monitoring accuracy shown in the previous analysis, is consistent with students in the analogy condition experiencing more illusions of comprehension than students in the no-analogy condition.

Fifth, coding the annotations that students made on their materials also provided some insight into differences in how students approached the task in the analogy and no-analogy conditions. Only 10 students made any annotations at all (6 in no-analogy and 4 in analogy). All used underlining, highlighting, and circling text, but students in the no-analogy condition were somewhat more likely to insert interpretative or summative comments (4/6 vs. 1/4), $\chi^2 (1, n = 10) = 1.67, p = .20$. Students in the no-analogy condition also gave more attention to the graphs. One student sketched a picture of the graph as he or she wrote the essay, and four students marked or summarized information on the graphs. These five students who showed evidence of having attended to the graphs all were in the no-analogy condition, $\chi^2 (1, n = 90) = 5.29, p = .02$.

Finally, students in the no-analogy condition were more likely to reference the documents in their essay than were students in the analogy condition (60% for no-analogy vs. 40% for analogy), $\chi^2 (1, n = 90) = 5.48, p < .02$. Although these qualitative results are based on a small subset of the students, they suggest that students in the analogy condition were less attentive to the graphs, less responsive to the instruction that prompted them to use evidence from the materials to support their essays, and less sensitive to the difficulties presented by this learning activity.

**DISCUSSION**

The results of this experiment suggest that the presence of an instructional analogy actually harmed the accuracy of students’ judgments of understanding and contributed to an overall trend of students being overconfident. Although the
intention behind including analogies in science text is to support student comprehension, this study showed that analogies can not only harm learning but can also interfere with students’ ability to accurately evaluate their own learning. The second major finding from this study was that having skills in graph comprehension may be necessary to support effective monitoring from inquiry units that include both text and graphs.

Why Might Analogies Hurt Metacomprehension Accuracy?

One possible explanation for this finding could be that readers possess the belief that analogies improve learning and rely on this analogy heuristic, similar to the multimedia heuristic suggested by Serra and Dunlosky (2010), when making their predictive judgments. However, because there was not an overall increase in judgment magnitude in the analogy condition in this study, this explanation does not fit the results well. Instead, the poor metacomprehension accuracy shown in this study results from the combination of poorer test performance and the lack of a downward adjustment in judgments.

An alternative explanation for why analogies harmed metacomprehension accuracy is that students may have relied on less valid cues as a basis for their comprehension judgments rather than the most valid cues that would result from an attempt to assess the quality of one’s own mental model (Koriat, 1997; Rawson, Dunlosky, & Thiede, 2000; Wiley et al., 2005). Less valid cues include reliance on general beliefs about oneself as a learner, perceptions of interest or enjoyment, or perceptions of fluency during processing. In general, analogies are presented in concrete language that may easily and naturally evoke images in the reader’s mind (McTigue & Slough, 2010) and that may increase interest in the topic (Sadoski, Goetz, & Fritz, 1993). However, interest is not a valid cue for assessing one’s own level of understanding. Similarly, the presence of the analogy may have allowed students to believe they could visualize or process the information easily, increasing their feelings of fluency without actually increasing their understanding of the underlying scientific phenomenon. If students base their comprehension judgments on feelings of interest, familiarity, or fluency activated by the presence of the analogy, rather than cues based in their actual comprehension of the concepts, that could have resulted in the inaccurate metacomprehension and illusion of comprehension found in the current study.

Why Do Graph Interpretation Skills Help?

Although this experiment revealed that analogies can harm metacomprehension accuracy, it also revealed that having good graph interpretation skills can help improve metacomprehension accuracy. Graph comprehension can be an effortful
process. Specifically, research has indicated that graph comprehension requires the recognition of visual features, the translation of those visual features into the conceptual relations that are represented by those features, and the inferring of referents from labels, captions, and titles within the graph (Carpenter & Shah, 1998). Based on the idea that our metacognitive judgments of comprehension are only as accurate as the cues on which they are based, if a student is able to interpret a graph and use this process of interpretation to form their comprehension judgments, then it should result in more accurate monitoring because their judgments would be based on valid predictors of comprehension. On the other hand, if a student does not possess the skills necessary to interpret a graph, he or she will only have access to cues related to how interesting, appealing, or familiar the graph is to him or her. These cues are unlikely to be valid predictors of comprehension. Further, students who were able to interpret graphs may also have been able to use the graphs to self-test and evaluate their comprehension. For example, if the reader thought that changes in atmospheric levels of CO₂ were partially responsible for changes in global climate, they could have looked at the graph to help confirm or refute this idea. Then, whether or not their inference matched what was seen in the graph could be used to evaluate their overall comprehension of the information.

The finding that students with better graph interpretation skills were better at gauging their own learning from this unit with both text and graphs suggests that training students how to use graphs for understanding could help to improve metacomprehension on these kinds of units. Testing this hypothesis is an important question for future research.

Although the single spatial measure (paper folding) in this study did not uniquely predict metacomprehension, and in this study graph interpretation skills were seen to be more directly related to performance, more work is needed to understand the relation of spatial skills to effective learning and metacognition from science inquiry activities. In general, learning in science has been argued to be especially dependent on the learner’s ability to visualize, manipulate, and animate spatial information (Hegarty, 2010; Uttal & Cohen, 2012). For example, it has been observed that science majors usually perform better on tests of spatial ability than nonscience majors (Casey, Winner, Brabek, & Sullivan, 1990; Humphreys & Yao, 2002; Pallrand & Seeber, 1984). There is also some evidence that students who are higher in spatial skills may show better achievement in science courses, including biology (Koroghlanian & Klein, 2004), chemistry (Bodner & McMillen, 1986; Carter, LaRussa, & Bodner, 1987; Pribyl & Bodner, 1987; Wu & Shah, 2004), and earth science (Black, 2005; Sibley, 2005). The role of spatial skills, prior knowledge, and prior experience in science are all important constructs to more fully explore in the future, because they could all either lead to worse metacomprehension accuracy (via general overconfidence in one’s skills or abilities) or more accurate metacomprehension by allowing for
Another avenue for future research is more fully exploring the role of prior knowledge in metacomprehension. One weakness of the current study is that only a singular measure of prior topic knowledge was included, and it was a self-rating measure. The design of the current study leaves open the possibility that prior topic knowledge could moderate the findings.

More work is also needed looking at the effects of analogy on metacomprehension accuracy. Because analogies are so prevalent in science, technology, engineering, and math (STEM) teaching materials across all grade levels, it is important to understand how and when they affect students’ ability to monitor and evaluate their comprehension. Although the goal of an analogy is to make understanding of a complex phenomenon easier, the current research suggests the presence of an analogy can lead to illusions of comprehension. Although analogies may be included within expository science text with the best of intentions, 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 or test their understanding of the novel phenomenon. Focusing more on seductive information without connecting it to more important conceptual information can lead to the “seductive details effect” (Garner, Brown, Sanders, & Menke, 1992). The idea behind the seductive details effect is that interesting information competes with more relevant information for readers’ attention. Although seductive details do tend to be rated as more interesting to readers (Hidi & Anderson, 1992), the presence of interesting sections of text or images have been shown to harm readers’ retention of the main ideas of the text (Harp & Mayer, 1998; Sanchez & Wiley, 2006; Wiley, Sanchez, & Jaeger, 2014). The decrement in performance due to the presence of an analogy in this study is consistent with this finding. In addition, this study demonstrates a second level on which interesting adjuncts can damage learning processes. The seductive appeal of the analogy may have prompted readers to rely on their perceptions of it rather than on more relevant cues based in their representations of the entire text when making their comprehension judgments. Thus, interesting examples intended to improve learning can backfire both in terms of comprehension and metacomprehension outcomes.

As noted previously, prior research has found that the presence of images in text can affect judgments of comprehension for a science text on lightning (Serra & Dunlosky, 2010). Other research has extended this finding using relative accuracy measures and also finding similar detrimental effects on metacomprehension accuracy when texts on science topics contain decorative images as compared with plain text (Jaeger & Wiley, 2014). In their study, Jaeger and Wiley (2014) demonstrated that presenting decorative images alongside expository text harmed metacomprehension accuracy compared with when no
images were present. They suggested that this decrement in judgment accuracy could have been the result of students relying on cues based in their interest or enjoyment of the images that would not have been valid predictors of performance on the tests. Moreover, this study also used a conceptual diagrams condition that, like the analogy condition used here, was expected to improve both learning and metacomprehension. However, neither benefit was found. To test whether students could be put in a position to use diagrams more effectively, a second study used a self-explanation condition that prompted students to try to make connections among ideas as they were reading. Importantly, this condition improved monitoring accuracy for both decorative and conceptual image conditions, such that decorative images no longer hurt metacomprehension, and a small benefit could be seen in metacomprehension accuracy in the conceptual image condition.

Although many studies have demonstrated situations in which students have poor monitoring accuracy, the Jaeger and Wiley (2014) study joins several studies showing that metacomprehension accuracy can be improved if readers are led to focus on more valid cues as the basis for their predictions. From this perspective, poor metacomprehension accuracy can be presumed to result when learners base their judgments on cues that are not valid indicators of actual comprehension (Rawson et al., 2000; Wiley et al., 2005). One successful approach that seems to help students to gain better access to valid cues while judging their level of comprehension has been to instruct students to generate key words or summaries of material after a delay (Thiede & Anderson, 2003; Thiede et al., 2003). It has been suggested that monitoring improves because of this instruction because as time passes, surface cues decay and become less accessible, whereas the situation model that predicts performance on comprehension tests is less vulnerable to forgetting (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Therefore, instructing students to generate key words or summaries after a delay helps readers use more valid situation model–based cues as the basis for their judgments of learning (Thiede, Dunlosky, Griffin, & Wiley, 2005). In other studies, having readers create concept maps or self-explanations while reading has been shown to increase relative comprehension monitoring accuracy. Again, these manipulations are thought to improve monitoring accuracy by prompting readers to use cues based in their attempts to generate connections rather than using less valid surface features of the texts or feelings of fluency (Griffin, Wiley, & Thiede, 2008; Redford et al., 2012; Thiede, Griffin, Wiley, & Anderson, 2010). A third approach that has resulted in robust improvements in monitoring accuracy is giving students practice inference questions and warning them that the comprehension test will include inference questions. Providing practice inference tests and clear reading goals seems to allow students to engage in more accurate prediction and more effective self-testing of their understanding while reading (Thiede, Wiley, & Griffin, 2011).
These results suggest several possible ways that future work can attempt to help students to avoid illusions of understanding when analogies are used in an effort to support better comprehension of scientific phenomena.

ACKNOWLEDGMENTS

We thank Anne Britt, Thomas Griffin, Andrew Taylor, Tim George, Alia Mohammad, Amy Thakkar, and the other members of Project READi for their contributions in developing the materials and assistance with data collection and coding. We also thank Gale Sinatra and Robert Danielson for suggesting the analogy. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

FUNDING

This research was funded by grant R305F100007 from Reading for Understanding Across Grades 6 through 12: Evidence-based Argumentation for Disciplinary Learning from the Reading for Understanding Research Initiative.

REFERENCES


Enhanced Greenhouse Effect

Many people have heard of the “greenhouse effect,” but not everyone knows what the “greenhouse effect” is exactly. The Earth’s greenhouse effect is part of the system that helps determine our planet’s average surface temperature. The greenhouse effect is important. Without greenhouse gases like water vapor, carbon dioxide, and methane in our atmosphere, the Earth’s average surface temperature would be about \(-1^\circ F\), which is about 60°F colder than it is today. Life on Earth would be much different without the greenhouse effect. In fact, life might not exist on Earth at all without the greenhouse effect.

So how does the greenhouse effect work? Energy in the form of visible light from the Sun enters the Earth’s atmosphere. Almost all energy that reaches the Earth originates as solar energy coming from the Sun. Clouds and other particles in the atmosphere reflect about 26% of this solar energy back into space. Some of this incoming solar energy (about 19%) is absorbed by clouds, gases, and other atmospheric particles. But because the atmosphere is mostly transparent to visible light, about 55% of the solar energy passes through the atmosphere and reaches the Earth’s surface. When energy from the Sun reaches the Earth’s surface, it is either absorbed by land and oceans or sent back out into the atmosphere as invisible infrared waves. A small portion of this infrared energy (about 4%) returns back into space. But most of this infrared energy is absorbed by greenhouse gases that prevent it from leaving the Earth’s atmosphere. In total, about 51% of the energy from the Sun that reaches the Earth is absorbed by the Earth’s surface or remains trapped in the Earth’s atmosphere by greenhouse gases.

An analogy may help illustrate this process. Imagine your car parked out in the sun with the windows slightly open. The temperature inside your car feels warmer than the outside temperature. The reason for this difference in temperature is that the Sun’s light energy enters through the car windows and is transferred to the seats, dashboard, carpeting, and floor mats. These objects give off some of this energy as infrared waves. But the windows block this infrared energy from escaping, causing it to be trapped inside
the car. Some of the trapped energy is transferred to the air inside the car, raising its temperature. This is an example of a greenhouse effect. Similarly, the Earth is surrounded by a blanket of gases, which, like the windows on the car, allow light energy from the Sun to pass through to the Earth’s surface but block some infrared energy from returning to space.

Throughout history, the Earth’s average temperature has changed in cycles. Each cycle consists of a long period when the entire Earth cools. Much of the ocean’s water becomes frozen in glaciers, and sea levels are much lower. These long cooling periods are usually followed by short warm periods during which Earth warms rapidly and the ice melts. It has been found that the cycles usually occur every 100,000 years or so. We are now in a warm period that has lasted more than 10,000 years. This is longer than most warm periods. According to the pattern of temperature cycles, we are due for the next cold phase. Over the last century, however, average global temperatures have risen instead.

Scientific measurements show that average global temperatures are rising. Surface temperatures have increased since 1880, with most of the warming occurring since the 1970s. The 20 warmest years on record have occurred since 1981 with the top 10 taking place in the last 12 years. Some believe these recent changes in average global temperatures are due to natural causes alone. However, scientists believe this is incorrect. They believe it is highly unlikely that the recent changes in average global temperatures are due to normal temperature cycles. Although it is true that climate changes can and do happen naturally, the rapid warming that the Earth is currently experiencing cannot be explained by natural factors alone. It is incorrect to believe that human activity can have no impact on the climate. The climate can be changed by humans and is being changed by humans. In fact, almost all climate scientists (97%) agree that human activities that increase the amount of greenhouse gases in our atmosphere have added to the greenhouse effect. As the amount of greenhouse gases in the atmosphere increases, more infrared energy is trapped in our atmosphere. In turn, the Earth’s surface is further warmed.

The main greenhouse gases entering the atmosphere due to human activities include carbon dioxide, methane, nitrous oxide, and fluorinated gases. The increase in carbon dioxide in the atmosphere comes mainly from burning fossil fuels (coal, oil, and natural gas) and by deforestation (the clearing of forests, which naturally absorb carbon dioxide). Methane is given off when coal and oil is produced and transported, by raising livestock and other farming practices, and through the decay of organic waste in landfills. Nitrous oxide is emitted into the atmosphere by factories and farms as well as by the burning of fossil fuels and solid waste. Fluorinated gases are also being released into the atmosphere by factories and are contributing to the greenhouse effect. Although all of these gases contribute to the greenhouse effect, you have probably heard about carbon dioxide the most because it accounts for more than 60% of greenhouse gases. The U.S. is adding carbon dioxide to the atmosphere from burning fossil fuels to generate electric power and exhaust from cars, trucks, airplanes, and trains.

The additional warming of the Earth due to increased greenhouse gas concentrations caused by human activities (such as the burning of fossil fuels) is called the enhanced greenhouse effect. It is this addition of more greenhouse gases into our atmosphere, coming directly from human activity, that is responsible for the increases in average global temperatures. To complete the car analogy, humans have the ability to influence both the windows on the car and the greenhouse gases in the atmosphere. When we slowly roll up the windows, less energy can escape from the car. Similarly, when humans add greenhouse gases into the atmosphere less energy can escape and the Earth warms up. We all know the safety warnings about leaving a child or pet in a locked car with the windows up on a
sunny day. If action is not taken to affect the amount of greenhouse gases in our atmosphere, we run the risk of Earth’s temperature rising to unsafe levels for all living things.

Which things seem true based on the text you that read?

1. All of the energy from the Sun that passes through the atmosphere reaches the Earth’s surface.
2. Global temperature changes are random and unpredictable.
3. Over the past 50 years, average global temperatures have decreased.
4. Higher levels of carbon dioxide in the atmosphere lead to higher sea levels.
5. All changes in global temperatures are caused by changes in energy from the Sun.
6. When the greenhouse effect is the strongest, sea levels will be high.
7. Cutting down forests increases the amount of infrared energy in the Earth’s atmosphere.
8. Increases in fossil fuel use increase the amount of heat that escapes to space.
9. An increase in the number of farm animals will increase greenhouse gases.
10. In the past 100 years, both fossil fuel use and carbon dioxide levels have increased.
11. Cutting down forests causes more carbon dioxide to stay in the atmosphere.
12. If the amount of energy given off by the Sun stayed the same, there would be no changes in global temperatures.
## APPENDIX B

Follow-Up Tests for MANCOVA on Learning Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Analogy</th>
<th>Graph Interpretation Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay scores</td>
<td>Analogy &lt; No Analogy* ($\eta^2 = .04$)</td>
<td>Low &lt; High** ($\eta^2 = .05$)</td>
</tr>
<tr>
<td>Comprehension test scores</td>
<td>Analogy &lt; No Analogy** ($\eta^2 = .10$)</td>
<td>Low &lt; High** ($\eta^2 = .05$)</td>
</tr>
</tbody>
</table>

Low, low graph interpretation skills; High, high interpretation skills.

**$p < .05$, *$p < .07$.**