Biology Textbook Graphics and Their Impact on Expectations of Understanding

Jennifer Wiley, David Sarmento, Thomas D. Griffin & Scott R. Hinze

To cite this article: Jennifer Wiley, David Sarmento, Thomas D. Griffin & Scott R. Hinze (2017) Biology Textbook Graphics and Their Impact on Expectations of Understanding, Discourse Processes, 54:5-6, 463-478, DOI: 10.1080/0163853X.2017.1319655

To link to this article: http://dx.doi.org/10.1080/0163853X.2017.1319655

Published online: 08 May 2017.

Submit your article to this journal

Article views: 35

View related articles

View Crossmark data
Biology Textbook Graphics and Their Impact on Expectations of Understanding

Jennifer Wileya, David Sarmentoa, Thomas D. Griffinb, and Scott R. Hinzeb

aDepartment of Psychology, University of Illinois at Chicago; bDepartment of Psychology, Virginia Wesleyan College

ABSTRACT

Graphics presented alongside expository science texts can have a number of positive effects for instruction, including facilitating engagement, arousing interest, and improving understanding. However, because students harbor expectations about which contexts are likely to support better understanding, the mere presence of graphics also has the potential to lead to inaccurate judgments of understanding when those graphics do not actually lead to presumed levels of performance. Previous work has demonstrated that including graphics alongside text can alter the judgment process. The present work explores different categories of instructional graphics found in biology textbooks and tests how different graphic types, classified by their form and function, can affect expectations of understanding prior to actual reading. Experiment 1 found that realistic, depictive graphics predominated in a middle school text, whereas more abstract and explanatory graphics predominated in a college text. Experiment 2 demonstrated that different categories of graphics led to differences in expectations of how helpful graphics would be for understanding.

Introduction

Despite their intuitive advantage and some evidence that visualizations can enhance learning, many studies show that adding graphics1 to text can fail to improve, and even sometimes hinder, learning compared with plain text lessons (Harp & Mayer 1997; Höffler & Leutner, 2007). The particular graphics that are used seem to impact whether harm or benefits are seen in comprehension (Carney & Levin, 2002; Levin, Anglin, & Carney, 1987). For example, despite some mixed results (Rey, 2012), the presence of interesting but irrelevant graphics (often referred to as “decorative” or “seductive” images) can undermine comprehension of expository texts about scientific processes compared with plain text or with texts containing relevant “explanatory” or “instructional” graphics that attempt to explain or articulate concepts (Harp & Mayer, 1997, 1998; Korbach et al., 2016; Park et al., 2005; Sanchez & Wiley, 2006; Wang et al., in press). These findings suggest that certain types of graphics may cause comprehension difficulties, in contrast to having more positive effects on comprehension. However, many students may have general a priori beliefs about the benefits of graphics for learning. The purpose of this research was to test whether students have different expectations about how different types of instructional graphics might impact their understanding and the extent to which differences in expectations of understanding may be seen due to interest.

Studies have begun to consider the impact of the inclusion of graphics in instructional text on the ability to monitor comprehension. In one study (Serra & Dunlosky, 2010), undergraduates either
received a plain text lesson on lightning formation, a lesson illustrated with explanatory process diagrams, or with photographs of lightning strikes. Before reading, participants rated how well they expected to understand the text and then also judged their understanding after reading, but before taking the actual test. The presence of either photographs or explanatory diagrams increased the magnitude of their ratings for understanding at both time points; however, test performance was only better than the plain text condition for the explanatory diagrams but not the photographs. This implies worse monitoring because test scores did not actually improve from the photographs, even though students expected it to. Because differences in ratings arose even prereading and were similar to differences postreading, these results suggest that students relied on a generalized multimedia heuristic to form an expectation of understanding that persisted to impact their postreading judgments of understanding. In addition, participants were asked before they knew the topic whether they believed the presence of graphics generally improves learning from science texts. Nearly all (94%-99%) students endorsed this generalized multimedia heuristic. Lenzner, Schnitz, and Müller (2013) found similar effects. These findings suggest that graphics can create expectations about learning that may undermine a reader’s ability to accurately judge how well they actually understood a text.

Although these studies have established that graphics can have effects on judgments when learning about a single topic, the effects of graphics on judgments have also been examined when readers are given a set of expository science texts on a range of topics. Jaeger and Wiley (2014) had undergraduates read texts on five topics (eruptions, evolution, lightning, ice ages, and cheese-making) that were each illustrated with explanatory process diagrams, decorative photographs (e.g., of a glacier, eruption, or lightning), or no graphics. They made readers predict how well they would do on an upcoming test after reading the texts but before taking the tests. Graphic condition did not have an impact on the overall magnitude of either the predictive judgments or test scores. However, readers in the photograph condition made judgments that were less correlated with their own performance on the comprehension tests across topics. This intra-individual correlation between performance predictions and actual performance across an array of texts serves as a measure of metacomprehension accuracy (Maki, 1998). Jaeger and Wiley (2014) found negative correlations in the photograph condition but positive correlations in other conditions, suggesting that readers attended to some cue in the decorative photographs they used to inform their judgments, but that cue ended up relating negatively to comprehension. Ackerman and Leiser (2014) found similar results. Taken together, the above studies show that when graphics are interesting but irrelevant, merely decorative rather than instructional, or intended to spark engagement more so than to explicate important processes for understanding a topic, metacomprehension may suffer. That is, students may experience illusions of understanding likely because they are using a priori, and possibly inaccurate, beliefs about the educational utility of graphics to guide their judgments of understanding or are mistaking perceptions of interest for understanding.

These observed illusions of understanding related to decorative graphics are consistent with theories of metacomprehension (Griffin et al., 2013) that have integrated the cue-utilization approach (Koriat, 1997) with Kintsch’s (1998) levels of comprehension framework. These theories suggest that accurate monitoring depends on the use of valid cues (Koriat, 1997). In terms of comprehension monitoring, this means that to be accurate readers need to base their judgments of understanding on cues related to the quality of the situation model that has been constructed for each text (Wiley, Griffin, & Thiede, 2005; Wiley, Thiede, & Griffin, 2016). Yet, learners are often misled about their level of comprehension as a result of relying on generalized heuristics such as beliefs about themselves or more superficial features of the learning materials like topic interest and familiarity (Thiede et al., 2010). Because these superficial cues do not align with the processes required for comprehension, reliance on these cues leads to illusions of understanding. Using a generalized multimedia heuristic can create an expectation of understanding that blinds readers to the more valid cues tied to the quality of the processing they are engaging in while reading. Further, since graphics vary in features that are more (e.g., causal arrows) or less (e.g., realism) related to the causal/logical relations in text, judgments of understanding may be differently impacted by different types of graphics.
Although a number of studies have explored the harmful effects of decorative graphics, much less research has explored effects among different types of instructional graphics that are intended to support understanding of the topic. In one such study, Ikeda et al. (2013) asked whether differences in monitoring might be found between two types of instructional graphics (functional magnetic resonance images vs. bar graphs) when undergraduates were given a text about brain activity in patients with depression. Readers were asked to judge how well they would do on a test about the text after reading but before taking the test. Although the two graphics were essentially equivalent in their instructional function, Ikeda et al. found that the functional magnetic resonance images increased the magnitude of judgments more so than the bar graphs. Further, despite differences in the magnitude of these predictive judgments, there were no actual differences in learning due to graphic type. Other studies have also shown a lack of learning benefit from adding realism to conceptually relevant explanatory diagrams (Imhof, Scheiter, & Gerjets, 2011; Mason et al., 2013). The Ikeda et al. results suggest that students may experience illusions of understanding even when the graphics presented are conceptually relevant for understanding the text. The authors hypothesized this was due to the realistic properties of the functional magnetic resonance images, which suggests that realism may cause people to inflate their judgments of understanding.

Other studies have found that more abstract explanatory process diagrams can produce better comprehension than more realistic ones (Butcher, 2006; Brucker, Scheiter, & Gerjets, 2014). For example, Butcher (2006) found advantages for learning about the circulatory system from simplified over detailed diagrams. Scheiter et al. (2009) also showed a learning benefit from abstract schematics of mitosis over more realistic graphics. They asked learners to evaluate their learning after seeing a lesson but before taking tests on the content, which revealed that learners tended to perceive the schematics as more helpful to learning. This suggests that learners may be sensitive to differences in how realistic versus abstract graphics affect learning. Although these studies used a single topic, they provide initial evidence that it is important to explore which particular features or types of graphics may evoke different expectations about how graphics might affect understanding, which can lead to either accurate or illusory estimates depending on how much the graphics actually impact learning outcomes.

Which kinds of graphics might cause illusions of understanding?

Rather than creating a new set of graphics specifically for this research, the current study relied on existing graphics that appeared in actual biology textbooks for several reasons. First, biology was selected as a focus for this research because it is one subject matter for which visual representations are an important form of communication and represent a common instructional device (Treagust & Tsui, 2013). Further, it is also already suspected that students often misuse visual representations or fail to use them effectively when learning in biology (Griffard, 2013). Second, using authentic biology textbook materials ensures some ecological validity for this work and allows for the assessment of students’ expectations about instructional graphics to which they have likely been exposed. Further, we were interested in the naturally occurring prevalence of different graphic types and features and how they might change across different grade levels. Because biology is covered at multiple grade levels in kindergarten through grade 12 education and is taken by a large percentage of college students to fulfill either major or general-breadth requirements, this allowed us to study both variations in types of graphics across grade levels and differences in perceptions of different graphic types.

To document variability in expectations about how different graphic types might affect understanding, it was necessary to first identify a scheme for categorizing them. Prior work by Levin and colleagues (Carney & Levin, 2002; Levin, Anglin, & Carney, 1987) demonstrated that learning from illustrated prose differs due to picture type. Although the Levin work provides a valuable precedent for expecting differences in actual comprehension due to graphic types, it was problematic to apply this scheme to the graphics found in biology textbooks. Instead, more recent work with science textbooks (Dimopoulos et al., 2003; Koulaidis et al., 2005; Lemoni et al., 2011) has followed the Kress
and van Leeuwen (1996) scheme and coded graphics for two dimensions: the functions they are intended to serve and whether their form is realistic/naturalistic versus abstract/conventional. The potential for the form and function of graphics to impact judgments of understanding is motivated by the theories of metacomprehension (Griffin et al., 2013) described earlier. Realistic pictures may evoke the use of superficial cues on which readers may base their judgments of understanding because the images are likely to seem more familiar, relatable, and interesting to typical novice biology students than abstract conventions requiring specialized knowledge. Alternatively, readers may rely on cues such as vividness that are more predictive of their memory for discrete details rather than cues that are more predictive of comprehension of logical/causal relations that form the situation model. The Kress and van Leeuwen (1996) scheme was used here as a straightforward way to examine which objective features of graphics may be most likely to affect expectations of understanding.

**Experiment 1**

The purpose of this initial study was to provide an analysis of the kinds of graphics that appear in biology textbooks as a first step toward understanding when, how, and why they may lead to illusions of understanding. Because so much depends on the graphics themselves, the main goal for this study was to document what types of graphics are present in textbooks at middle school, high school, and college levels. Dimopoulos et al. (2003) reported a preponderance of realistic photographs/drawings and graphics that merely showed or labeled parts of objects (e.g., parts of a cell) within six general science texts. However, the dependence between grade-level and scientific discipline in their study limits the ability to make across-grade within-discipline comparisons and to understand how the nature of graphics might change as students advance in a particular content area. The present study held the disciplinary area constant (all texts are in biology) to examine whether the distribution of graphic types would change between texts written for middle school, high school, and undergraduate levels. Based on the results of Dimopoulos et al. (2003), it was predicted that a middle-school text would use primarily realistic, depictive graphics, whereas a college text might include fewer realistic graphics but more serving deconstruction, classification, and explanation functions.

**Methods**

Three biology textbooks were analyzed: one from middle schools in Chicago Public Schools (Padilla et al., 2001; 11–12 years), one from high schools in Chicago Public Schools (Miller & Levine, 2010, 13–17 years), and one at the undergraduate level (Freeman, 2010; 18–19 years) from Introductory Biology at the University of Illinois at Chicago. To hold content constant across grades, 10 topics were selected: cell division, photosynthesis, genetics, natural selection, bacteria, fungi, green plants, metabolism, digestion, and global cycles/ecosystems. At each grade level all numbered figures in the 10 chapters corresponding to these topics were analyzed. The use of only numbered figures provided a clear standard for inclusion of content but also excluded purely decorative graphics that were not referenced by the text. The content for each of the figures (N = 655) was categorized on two dimensions: form and function (Dimopoulos et al., 2003; Kress & van Leeuwen, 1996). Examples are shown in Figure 1.

There were three categories for form, representing realistic or conventional graphics, plus an intermediate category called hybrid for graphics that combined some elements of realism with abstractions or conventions. Photographs, microscopic images, and drawings were coded as realistic.

---

2 These texts were not intended to represent a random sampling of all science texts or all possible biology textbook graphics. This particular set of three texts was chosen to represent the actual developmental sequence of textbooks used by a substantial portion of undergraduates at University of Illinois at Chicago who attend Chicago Public Schools and to provide an authentic sample of biology textbook graphics for a descriptive content analysis.
Figure 1. Example graphics for different categories from college textbook (FREEMAN, SCOTT, BIOLOGICAL SCIENCE, 4th ed.) © 2011 Pearson Education, Inc. Reprinted by permission of rightsholder. For an example of a Realistic, Depictive graphic included in the textbook, see “Juvenile crocodile eating a small frog” Getty Image #520473858. (http://www.gettyimages.com/detail/photo/juvenile-crocodile-eating-a-small-frog-high-res-stock-photograph/520473858). Realistic, Classifications were juxtapositions of two realistic images, such as males and females of a species, or two different species. Examples of these image types could not be included in this figure due to copyright restrictions.
All graphics that presented information in a conventional, codified, or abstracted way were considered conventional (i.e., graphs, maps, flowcharts, molecular structures, and diagrams constructed according to conventions). Realistic graphics that had conventional notations (superimposed diagrams or text, labeled parts, insets, or balloons showing cross-sections or zoomed-in to show details, or scale markers) were coded as hybrid, as were abstract displays (tables or graphs) that were illustrated with realistic examples.

There were four categories for function: depictive, deconstructive, classification, and explanative. Explanative graphics depicted causal or logical sequences, or processes of change, with action often visualized by arrows in order to illustrate technical (e.g., experimental procedures) or natural processes (e.g., nitrogen cycle). In lieu of the “narrative” label proposed by Kress and van Leeuwen (1996), we adopted the “explanative” label from Mayer and Gallini (1990) and Harp and Mayer (1997) as a better descriptor for these graphics, which largely consist of process models that are so important in biology (Griffard, 2013).

Classification graphics displayed taxonomic relationships between entities. This category was used when the objects of the natural world were represented in terms of a hierarchy or taxonomy. Some linguistic equivalents of these diagrams are expressions like “this belongs to” or “this is a kind of.” This category also included comparisons where two subordinates of a category or two examples from different categories were compared, with linguistic equivalents “this is similar to/different from.” Other classification graphics involved composites showing where a particular exemplar could be found.

Both Dimopoulos et al. (2003) and Kress and van Leeuwen (1996) had a category called “analytic” that included graphics that simply depicted an object or entity without elaboration as well as graphics with explicit labeling or other devices showing parts. Following Koulaidis et al. (2005), the “analytic” category was subdivided into two subcategories that we refer to as Deconstructive and Depictive. Deconstructive graphics were those that focused on part-whole relations. The meaning of these graphics corresponds to “this consists of” or “this is made of.” (If a figure compared parts of two or more entities, it was coded as classification.) Graphics that were used to show a thing without...
Two independent coders categorized all of the graphics, with Intraclass correlation coefficients (ICCs) > .92 for both form and function. In accord with the common expression “form follows function,” these two dimensions were not expected to be orthogonal. Not only are some instructional functions better served by particular forms, but it is difficult to imagine how some functions could be served at all by some forms. For example, the inherent inferential nature of ideas of causality (Hume, 1896) seems to preclude an explanatory function being served by a fully realistic graphic that only shows what is available to the senses. Likewise, a purely conventional abstraction is unlikely to serve the function of depicting the actual properties of objects. Thus, it was expected that the dimensions would correlate, and some combinations might not occur.

### Results and discussion

Using a likelihood ratio chi-square test, differences were found across grade levels in both form, $G^2(4) = 135.00, p < .0001$, and function, $G^2(6) = 83.73, p < .0001$ (Table 1). The gamma correlations with grade for both form and function were $\gamma = .63$ and $\gamma = .41$, respectively, $ps < .001$. As shown in Figure 2, as grade level increased the use of realistic graphics decreased, whereas hybrid and conventional increased. As shown in Figure 3, as grade level increased, the use of depictive graphics decreased, whereas the use of explanatory increased. In the middle-school text, the most common category of graphic was realistic/depictive, whereas hybrid/explanative was the most common in the college text.

As expected, no graphics at any grade level were realistic/explanatory or conventional/depictive. In addition, all graphics that served a deconstructive function had a hybrid form. Conventional/classification graphics were relatively uncommon. If the two dimensions are viewed as ordinal categories in the order shown in Table 1, then as the function moved from depiction toward explanation, form moved from realistic toward conventional. This was supported by strong significant gamma correlations between the form and function dimensions at each grade level, $\gamma = .79$ (middle school), $\gamma = .85$ (high school), and $\gamma = .66$ (college); $ps < .0001$.

The categorization results replicate the bias found by Dimopoulos et al. (2003) toward realistic graphics in middle-school science textbooks. In addition, conventional and explanatory graphics became more frequent in college. One possible consequence of early exposure to primarily realistic and depictive graphics is that students may not learn how to use more conventional forms. Students may also fail to appreciate the varying purposes of visualizations, which could also be reflected in their expectations of understanding.

### Table 1. Results of analyses of textbook graphics for form and function showing percentages at each grade level.

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Form</th>
<th>Depictive</th>
<th>Deconstructive</th>
<th>Classification</th>
<th>Explanative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle school</td>
<td>Realistic</td>
<td>39.10%</td>
<td>-</td>
<td>14.90%</td>
<td>-</td>
</tr>
<tr>
<td>N = 215</td>
<td>Hybrid</td>
<td>-</td>
<td>14.90%</td>
<td>14.90%</td>
<td>12.60%</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>-</td>
<td>-</td>
<td>2.30%</td>
<td>1.40%</td>
</tr>
<tr>
<td>High school</td>
<td>Realistic</td>
<td>21.70%</td>
<td>-</td>
<td>4.40%</td>
<td>-</td>
</tr>
<tr>
<td>N = 203</td>
<td>Hybrid</td>
<td>3.40%</td>
<td>14.80%</td>
<td>26.10%</td>
<td>25.60%</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>-</td>
<td>-</td>
<td>1.50%</td>
<td>2.50%</td>
</tr>
<tr>
<td>College</td>
<td>Realistic</td>
<td>6.30%</td>
<td>-</td>
<td>3.00%</td>
<td>-</td>
</tr>
<tr>
<td>N = 237</td>
<td>Hybrid</td>
<td>9.70%</td>
<td>6.30%</td>
<td>19.00%</td>
<td>36.70%</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>-</td>
<td>-</td>
<td>5.50%</td>
<td>13.50%</td>
</tr>
</tbody>
</table>

*Note. Combinations resulting in no observations are indicated by dashes.*

3The metaphoric category (Kress & van Leeuwen, 1996) was not used because graphics rarely fell into that category (as in Dimopoulos et al., 2003; Lemoni et al., 2011). In our sample, all could be coded as an analogical explanation of a process (i.e., explanatory graphics).
Experiment 1 demonstrated that students are exposed to a wide range of graphics in their biology texts and that the relative prevalence of different kinds of graphics can change dramatically across grade levels. Previous work has demonstrated that illustrating science texts with decorative graphics can alter judgments of understanding because students expect that multimedia materials will benefit learning (Serra & Dunlosky, 2010). The present work examines whether students possess different expectations for the different types of graphics categorized in Experiment 1.

To test for differences in expectations of understanding, students were asked to judge whether the different types of graphics found in Experiment 1 would improve their understanding. Judgments were made in the absence of seeing a text. Isolating the graphics allowed for the examination of expectations that might be specific to different categories based on their objective features and independent of accompanying text. Experiment 2 tested whether the form and function classifications assigned in Experiment 1 would impact the magnitude of students’ ratings of how helpful graphics would be for increasing their understanding.

One group of students was asked to rate how much each graphic would increase their understanding of the topic. If readers use a generalized multimedia heuristic (Serra & Dunlosky, 2010), no differences should be observed among graphic conditions, and all graphics should be perceived as helpful for increasing understanding. Alternatively, the relation of graphic types to a variety of

---

**Figure 2.** Distribution of graphic forms across grade levels.

**Figure 3.** Distribution of graphic functions across grade levels.

---

**Experiment 2**

Experiment 1 demonstrated that students are exposed to a wide range of graphics in their biology texts and that the relative prevalence of different kinds of graphics can change dramatically across grade levels. Previous work has demonstrated that illustrating science texts with decorative graphics can alter judgments of understanding because students expect that multimedia materials will benefit learning (Serra & Dunlosky, 2010). The present work examines whether students possess different expectations for the different types of graphics categorized in Experiment 1.

To test for differences in expectations of understanding, students were asked to judge whether the different types of graphics found in Experiment 1 would improve their understanding. Judgments were made in the absence of seeing a text. Isolating the graphics allowed for the examination of expectations that might be specific to different categories based on their objective features and independent of accompanying text. Experiment 2 tested whether the form and function classifications assigned in Experiment 1 would impact the magnitude of students’ ratings of how helpful graphics would be for increasing their understanding.

One group of students was asked to rate how much each graphic would increase their understanding of the topic. If readers use a generalized multimedia heuristic (Serra & Dunlosky, 2010), no differences should be observed among graphic conditions, and all graphics should be perceived as helpful for increasing understanding. Alternatively, the relation of graphic types to a variety of
possible cues predicts variations in expectations. If realism triggers superficial cues like interest, then the greater realism of hybrid graphics over purely conventional graphics could lead to higher judgments of interest and higher expectations of understanding. As for differences between functions, students may only have higher expectations for explanatory and classification graphics over deconstructive or depictive graphics if they recognize the main goal of reading in biology as developing understanding of processes rather than memorizing names of organisms and their parts. Given prior research showing students’ insensitivity to this distinction (Thiede et al., 2011), differences between these functions might not be predicted. However, prior research also shows individual differences in the types of cues students rely on to inform their judgments of understanding (Theide et al., 2010). A self-report measure of cue basis was also included to explore these differences for expectations.

In addition, a second group of students rated how much each of the graphics increased their interest in the topics. Interest ratings were included because students often use perceived interest, difficulty, or fluency as a basis for their judgments of understanding (Ball, Klein, & Brewer, 2014). Harp and Mayer (1997) suggested that interest could be driving some of the problematic effects associated with decorative graphics, because students may “like” seeing them but do not actually learn from them (Sung & Mayer, 2012). Finally, if patterns of understanding and interest ratings mimic each other, within or across categories, this would suggest that readers are using feelings of interest to drive their predictions of how much graphics might increase their understanding.4

Methods

Participants. Forty (26 women, 14 men, M_age = 18.97, SD = 1.68) undergraduates from the Introductory Psychology subject pool at the University of Illinois at Chicago participated in partial fulfillment of a course requirement. Half the participants completed understanding ratings and half completed interest ratings. The ratings conditions did not differ in science background in terms of college courses or self-reported ACT Science scores, t < 1.

Graphics stimuli. A subset of 70 graphics was taken from the college biology set used in Experiment 1. For each of the seven categories that had 10 or more graphics in them in Experiment 1, 10 graphics were randomly selected from each category, with constraints that only graphics that received 100% agreement during category coding were used, and with a goal of using a similar number of graphics from each of the 10 chapter topics to the extent that was possible. The graphic categories were Realistic/Depictive, Conventional/Explanative, Conventional/Classification, Hybrid/Explanative, Hybrid/Classification, Hybrid/Deconstructive, and Hybrid/Depictive.

Judgment prompts. Participants were asked to rate how the graphics would change their understanding or interest for a given topic relative to reading a plain text without that graphic. Participants used a slider to select values from −100 meaning “decrease interest/understanding” to 100 meaning “increase interest/understanding.” The graphics were presented at the top of each screen, the topic was printed below each graphic, and the rating scale appeared below the topic. The topics were the keywords used in description of the chapters in Experiment 1 (digestion, genetics, etc.).

Procedure. Participants completed the study individually using Qualtrics. After the instruction screen, participants viewed the 70 graphics in randomized order and made their ratings. After completing ratings, participants were given a cue-use question, which asked, “What things did you consider as you made your ratings of UNDERSTANDING/INTEREST for each image? Describe

4Understanding and interest ratings were performed by separate groups because pilot testing showed sizable contamination (order effects) across the two types of judgments when performed within subject.
what information you used when you decided to give an image a high or low rating.” Finally, participants were asked to complete a background survey that collected their age, gender, and science background. The study took 30 minutes or less to complete.

In addition, a prescreening survey administered at the beginning of the semester included this multimedia heuristic question: “Which do you think produces better learning in earth sciences or biology, a science textbook chapter without any images or a chapter with images?” A large portion of the sample (N = 32) completed the prescreening survey. All (100%) endorsed that a textbook chapter with images would produce better learning, showing the same strong support for a generalized multimedia heuristic as shown in Serra and Dunlosky (2010).

**Results**

The main focus for this study was testing whether students have different expectations about various types of instructional graphics, operationalized by examining differences in ratings about how each graphic would affect understanding of a topic. As shown in Figure 4, the magnitude of ratings varied across categories differently for understanding versus interest. All analyses were performed as linear mixed-effects models that included fixed effects for graphic category and rating type and random intercepts for participants and items. Both random intercepts accounted for significant variance. The analysis revealed no main effect for rating type (F(1, 38) = 3.35, p = .07) but a significant main effect for graphic category on the magnitude of ratings (F(6, 63) = 5.71, p < .0001). However, these results were qualified by a significant category × rating type interaction (F(6, 2685) = 14.17, p < .0001).

Follow-up analyses on understanding ratings revealed that expectations varied across graphic categories (F(6, 63) = 4.11, p < .01). Examination of Least Significance Difference (LSD) pair-wise comparisons among graphic categories and testing whether 95% confidence intervals included zero revealed that the categories fell into three clusters of magnitude. The hybrid graphics (other than hybrid/depictive) represented the top cluster whose ratings were significantly higher than all other categories and higher than zero. Graphics that served a merely depictive function (regardless of form) were not rated as significantly higher than zero for aiding understanding. In fact, 39% of these ratings were below zero, suggesting that many students were aware that depictive graphics can

**Figure 4.** Ratings of understanding and interest by graphic category (error bars represent 95% confidence intervals).
undermine understanding. Conventional graphics formed the third cluster that fell in the middle (regardless of function). They were also rated significantly lower than the hybrid (other than hybrid/depictive), but despite not being significantly different from the realistic graphics, they were rated as significantly higher than zero. Thus, students appear to hold expectations that some graphic forms and functions are more conducive to understanding.

Follow-up analyses on interest ratings revealed that the magnitude of judgments of interest also varied across graphic categories ($F(6, 63) = 12.34, p < .0001$) but in a pattern that was partially similar to and partially the inverse of the pattern for expectations of understanding. Only two clusters of graphics were found. Regardless of function, the categories that had some realism (realistic and hybrid) were viewed as increasing interest to a similar degree and above zero. On the other hand, conventional graphics were seen as significantly less interesting than all other graphics and not different from zero. Over half of the students rated the graphics in the explanatory/conventional (55%) and classification/conventional (64%) categories as decreasing interest.

Comparing the patterns among ratings for understanding and interest reveals there was not complete alignment. Expectations of an increase in interest seemed to depend on the graphic having at least some realism. Students expected purely depictive graphics to increase interest but not understanding. However, among the graphics that served some function beyond mere depiction, interest and understanding ratings were both higher for those with some realism (the hybrids). In fact, the between-category correlation among mean interest and understanding ratings across the five nondepictive categories was $r = .95, p < .02$. A between-graphic correlation was also computed between the interest and understanding ratings of those 50 nondepictive graphics. The correlation was $r = .56, p < .0001$, and after controlling for mean interest ratings within each category, the partial correlation was still $r = .47, p < .001$. This partial correlation reflects the interest–understanding relationship among graphics independent of variance in interest due to category. College students doubted the learning benefits of interesting but purely depictive graphics. However, among other graphics, having some realism still increased both interest and expectations of understanding across functional types.

**Differences in expectations of understanding due to cue-use responses.** Following the analyses performed by Thiede et al. (2010), responses to the cue-use prompt were coded to suggest what students may be using as a basis for their ratings for understanding. The responses fell into three categories (two independent raters agreed 100% on this coding). Approximately one-third ($N = 7$) of the participants used the ease of processing of the graphics, how simple or interesting they were, or how obvious their meaning was as a cue. A second group of participants ($N = 5$) used perceived topic relevance of the graphic. A third group ($N = 5$) considered the function of the graphic, whether it explained a process or included arrows, as the basis for their ratings.

The patterns of understanding ratings for each cue group can be seen in Figure 5. A linear mixed-effects model that included fixed effects for graphic category and cue group and random intercepts for participants and items was used to test for differences among these groups. Both random intercepts accounted for significant variance. The analysis revealed no main effect for cue group ($F < 1$) and a significant main effect for graphic category on the ratings for understanding ($F(6, 63) = 4.58, p < .0001$). However, these results were qualified by a significant interaction ($F(12, 1090) = 2.59, p < .01$). Follow-up analyses for each cue group used LSD pair-wise comparisons among graphic categories and tested whether 95% confidence intervals included zero. Participants relying on ease-of-processing rated realistic/depictive graphics as helpful for understanding (ratings above 0) and no lower than any other graphic category. Participants who considered relevance gave significantly lower ratings to realistic/depictive than to hybrid graphics that were explanatory or classificatory. The group that considered function gave realistic/depictive graphics mostly negative understanding ratings (but the confidence interval included 0) and lower ratings than all other nondepictive graphics.

---

5Three gave no response.
Discussion

The ratings data collected in Experiment 2 provide evidence for differences in expectations about whether different kinds of graphics are likely to improve understanding. Although most participants viewed the realistic/depictive graphics as not helpful or even harmful, one-third of participants who relied on ease-of-processing cues viewed these graphics as being as helpful as any other. This coheres with prior research showing that readers who report relying on superficial cues fail to discriminate among different levels of understanding for different texts, resulting in poor metacomprehension accuracy (Thiede et al., 2010). For all cue-use groups, the pattern among other categories was similar to the overall pattern, with hybrid graphics generally being viewed as more helpful than conventional and with minimal differentiation among the hybrid graphics that served deconstruction, classification, or explanatory functions. Also, although most learners thought that realistic/depictive graphics would increase their interest but fail to help their understanding, interest predicted understanding ratings for nondepictive graphics.

General discussion

The purpose of this research was to test for expectations among college students about the value of different kinds of textbook graphics for learning about biology. First, this required documentation of the prevalence of different kinds of graphics that appear in biology textbooks. Experiment 1 showed reliable coding of graphics using a scheme that considered both form and function, and changes in prevalence of different graphic types at different grade levels. This scheme was used in Experiment 2 to test college students’ expectations about how different graphic types affect understanding.

Realistic/depictive graphics were the most common in middle school texts but became much less frequent in the college texts. College students did report strong interest in realistic/depictive
graphics, but most were able to differentiate this interest from an expectation of understanding. However, some students who reported relying on superficial cues to make their ratings presumed these interesting graphics would increase understanding. These results have implications for research on seductive images that has used graphics similar to the realistic/depictive graphics used in this study. They suggest that observed effects of seductive images on college students’ judgments of understanding may not be primarily due to spontaneous use of heuristics about the utility of such graphics. It is possible that because Serra and Dunlosky (2010) prompted readers to report their general beliefs about graphics at the start of their study that likely primed students to apply a general multimedia heuristic even to graphics serving only a depictive function.

However, beyond the distinction for depictive graphics, students did not seem to further differentiate graphics based on their deconstructive, classificatory, or explanatory functions. This was despite being asked to rate how the graphic would impact their understanding of topics that primarily implied explanatory processes (e.g., photosynthesis, digestion, natural selection). One possible reason for this insensitivity is that even people who relied on relevant features (e.g., “arrows,” “process”) for their ratings were still impacted by superficial cues. Evidence for this emerged in the strong correlation between ratings of interest and understanding for all nondepictive graphics. Looking forward, several manipulations could help students to better differentiate among graphic types. For example, prompting students to explain while reading illustrated text can help them become aware of comprehension-relevant cues (Jaeger & Wiley, 2014), and such an intervention could help students differentiate these functional categories. Similarly, providing students with clearer goals for learning and informing them that they should consider whether graphics are helping them to understand how or why phenomena occur could also alter both prereading expectations and postreading judgments of understanding (Wiley, Theide, & Griffin, 2016).

Although students did not differentiate among nondepictive graphics due to their function, they did show sensitivity to their form, with a preference for hybrid over conventional graphics. Hybrid graphics were prevalent in both high-school and college texts, with purely conventional graphics appearing mainly in college. The lack of early experience with abstractions limits opportunities for students to become familiar with visual discourse conventions in science (Latour, 1987). Students were more interested in hybrid graphics and reported higher expectations of understanding from them. This could be due to lack of experience or instruction with the conventions (Hinze et al., 2013) or because the realism present in hybrid graphics increased interest that cannot be easily discounted (Ikeda et al., 2013).

If the substantial overlap between ratings of interest and understanding among nondepictive graphics was due to students using interest to inform their expectations of understanding, then this could explain the preference for hybrid over conventional graphics and the similarities across functional categories. Of course, it is possible that the direction of the interest–understanding relationship is reversed. Students may rate graphics as increasing their interest in part because they believe it helps them to understand. This would be consistent with the construct of cognitive interest, which can facilitate learning (Kintsch, 1980), in contrast to emotional interest, which can distract (Harp & Mayer, 1997; Park et al., 2005). To better understand the relations between emotional interest, cognitive interest, and metacomprehension, future work is needed that includes measures for each of these constructs.

The fact that no texts were presented alongside the graphics was both an asset and limitation of the current study. On the one hand, this was necessary to test whether people hold different expectations about different types of graphics, based on their objective features and independent of accompanying text. In addition, not requiring readers to actually read text with each graphic enabled the testing of a large number of images in order to obtain more valid estimates of expectations for each kind of graphic, rather than estimates that could be overly dependent on features of particular graphics or texts. On the other hand, understanding how readers apply such beliefs to making judgments of understanding when they are actually attempting to read and learn from text is a critical next step. Further, the accuracy of judgments of understanding can only be
assessed once comprehension outcomes are known. Learning from multimedia requires making connections between the text and graphics, which can require a number of relational reasoning processes not addressed here (Danielson & Sinatra, 2017). Considering the role of the relation between the graphic and text is another important future direction. Nevertheless, these findings set the stage for future studies that test which graphics do actually confer advantages in comprehension outcomes, which may affect judgments of understanding, and which may ultimately affect metacomprehension accuracy.

In conclusion, the differences (or lack thereof) in expectations about different types of graphics set the stage for finding possible differences in metacomprehension accuracy. When expectations of understanding exist about features of a learning context, they have the potential to alter monitoring processes and self-regulated study if those beliefs are used as the basis for judgments of understanding. Poor metacomprehension accuracy can be predicted when the inclusion of graphics has different effects on understanding than students expect. Poor metacomprehension could result from a student believing that different types of graphics have variable effects on understanding when they do not. Alternatively, poor metacomprehension could also result from a student believing that different types of graphics do not differentially affect understanding when they actually do. Effective use of explanatory graphics is a characteristic of successful biology learners (Griffard, 2013), and such graphics appear to aid learning more than deconstructive graphics showing labeled parts without their causal links (e.g., Mayer & Gallini, 1990). Thus, it could be problematic that students rate both explanatory and deconstructive graphics as similarly beneficial for understanding science topics that are based in explaining causal processes (e.g., photosynthesis, digestion, natural selection). Also, given the lack of an empirical or theoretical basis to presume hybrid graphics are much more helpful than conventional graphics (Imhof et al., 2011; Mason et al., 2013) and some evidence that conventional graphics are more conducive for understanding than realistic (Brucker et al., 2014; Butcher, 2006; Scheiter et al., 2009), expectations that hybrid graphics support the best understanding may be problematic. More optimistically, merely depictive graphics have been shown to be of minimal value compared with other instructional graphics, and most students in this study appeared to hold beliefs consistent with this view, despite finding those graphics interesting.

**Funding**

This material is based on work supported by the National Science Foundation under DUE grant 1535299. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Marta Mielicki, Allison Jaeger, and Tim George for their comments on this project.

**References**


