Diagram comprehension ability of college students in an introductory biology course

Alexa M. Kottmeyer, Peggy Van Meter, and Chelsea Cameron

The Pennsylvania State University, University Park, Pennsylvania

Submitted 21 August 2018; accepted in final form 20 December 2019

INTRODUCTION

Visual representations, such as systems diagrams, graphic data displays, and process flow charts, share the defining feature of using nonverbal representations to depict key knowledge elements in a spatially meaningful arrangement (30, 33). These representations are an important communication tool in the biological sciences, and test questions assess low-level comprehension (24). We use descriptive analyses of students’ performance on this measure to address research hypotheses related to students’ overall DCA, the influence of diagram characteristics on diagram comprehension, and differences between groups of students with stronger and weaker diagram comprehension abilities. The goal of these analyses is to provide instructors with information about how well the students will learn from them (14, 15, 26). Indeed, understanding and learning from visual representations is so critical that visual literacy, or representational competence, has been discussed as a core competency for biology students (24, 36). The evidence, however, suggests that many students have difficulty with diagram comprehension. Biology students, for example, score significantly lower on measures of visual knowledge compared with verbal knowledge of the same content (41), misinterpret parts of diagrams (42), and have difficulty comprehending and reasoning with diagrams (35), particularly representations that contain unfamiliar content (17).

While research to date points out some of students’ specific weaknesses, this research tends to focus on specific aspects of diagram comprehension, such as the ability to interpret a convention (e.g., arrows; Ref. 42), understand a type of diagram (e.g., process diagrams; Ref. 17), or reason about some content (e.g., biochemistry; Ref. 33). This research is informative, yet diagram comprehension ability (DCA) has not been explored as a general ability, which is applied across different types of diagrams that vary in the conventions used and the concepts displayed. Such a general ability is necessary for the biology student who needs to learn from the variety of diagrams embedded in their courses. In this respect, DCA functions similar to reading comprehension ability; that is, comprehension of a specific diagram is influenced by a student’s general representational competence, as well as task- and diagram-specific factors (e.g., prior knowledge). Examining this ability is necessary, however, because diagrams are not only prevalent in instructional materials, they are often used with the assumption that students can understand and learn from them (3, 19, 28). That is, we assume that students are able to accurately understand these representations and acquire content knowledge from them.

The present study addresses this assumption by examining biology students’ performance on a test of science diagram comprehension (10). This test, which was administered as part of a larger study (41), is composed primarily of diagrams from biological sciences, and test questions assess low-level comprehension. Although diagrams are only one type of visualization found in biology materials, we focus on this type here because diagrams are common (29) and vary in complexity (9) and level of abstraction (24). We use descriptive analyses of students’ performance on this measure to address research hypotheses related to students’ overall DCA, the influence of diagram characteristics on diagram comprehension, and differences between groups of students with stronger and weaker diagram comprehension abilities. The goal of these analyses is to provide instructors with information about how well the

HOW WE TEACH | Generalizable Education Research
students in their classrooms are able to understand the diagrams from which they are expected to learn.

The use of visualizations in biology education. Visual representations refer to external representations that depict biological concepts through spatial relationships and graphic or iconic markings (30, 33). Diagrams are a particular form of visualization that use conventions to communicate meaning (22, 40). These conventions are common markings used to communicate meaning, such as arrows, labels, coloring, and captions. A variety of diagram types are found in biology, including representational drawings or photographs of objects, schematic depictions, and process diagrams. These diagrams vary not only in type, but also along other characteristics, such as complexity (4) and abstractness (24). Figures 1, 2, and 3 show examples of the types of diagrams that are common in biology.

Visualizations, including diagrams, are prevalent in college-level biology courses. These representations are used to convey new concepts and can be found throughout the instructional material from which students must learn (14). In one examination of a popular introductory text, for example, Wright et al. (42) located more than 1,200 different figures. The use of diagrams is not limited to the instruction of new content. Biology students are also expected to work with visual representations of experimental data (3) and answer course exam questions that include visual displays (24).

The prevalence of diagrams is not surprising, because these representations are well-suited to convey biological principles. Diagrams are holistic, spatial arrangements; show both temporal and size scales; and depict structure-function relations (13). In this respect, diagrams are able to convey biological concepts and principles at varying levels of abstraction, both within and between systems (36). These diagrams often appear as part of a set that includes both verbal and visual representations of the content rather than a single representation (1, 9, 39). These combinations can be particularly valuable because of their ability to communicate relationships and promote the construction of high-quality knowledge representations (12, 16, 20).

Regardless of whether presented alone or in combination, diagrams are embedded in instructional material with the intent that students use them to construct internal mental models of the phenomenon (20, 30). This knowledge construction occurs

Fig. 1. Pollination diagram showing the following conventions: captions, arrows, cut outs, naming labels, color coding, and shading.
as the student selects and encodes key elements from diagrams and organizes these elements into an internal representation of the depicted content (9, 18). When more than one representation is present, students must not only organize their knowledge from each individual representation, but also generate connections to integrate the representations (1, 4, 12). These learning processes, whether with single or multiple representations, are active and lead to the generation of inferences that contribute to the construction of the students’ knowledge representation. This inference generation is the underlying process that explains why the type of representation studied affects what is learned from a representation (1, 9, 30). In a
study by Remmele and Martens (27), for example, eighth grade students used either stereoscopic or nonstereoscopic images to guide building clay models of the nasal cavity. Analysis of the final models showed that students who used stereoscopic images were better able to model the meatus of the nose. In short, differences in the type of image affected the students’ abilities to perceive anatomical structures and, consequently, affected the internal knowledge representation that was constructed.

Altogether, diagrams play an important role in the communication of biological principles, and students’ abilities to comprehend these representations is a key to their learning (13, 14, 35). Students who are able to accurately comprehend diagrams are better able to understand the depicted content and construct the desired knowledge representations. Alternatively, of course, poor or inaccurate diagram comprehension can result in poor or inaccurate knowledge of the content (32). Given the centrality of diagrams to student learning, understanding students’ DCA is necessary to understand how they interact with and learn from these representations.

Diagram comprehension ability. Though clearly important for successful learning in biology, accurate diagram comprehension is anything but certain. When Shah and Freedman (37) had college students complete a graph interpretation test, for example, the average score was only 64%; a fairly poor performance given that test items required only interpretation of explicitly depicted information. This average performance, however, is consistent with other research, which has shown that many novice students struggle to comprehend scientific visualizations (5, 6, 14). These weaknesses are demonstrated in a study by Schönborn and Anderson (33) in which biochemistry students were interviewed while studying three sets of visualizations on antibody-antigen interaction. Probes in the semistructured interviews began at a surface level and progressed to increasingly deeper comprehension. Students’ interview responses indicate that many had poor diagram comprehension and significant difficulty reasoning with the visualizations. Findings from these interviews are consistent with the report by Eilam (13) on various projects she has carried out in biology classrooms. Although these projects involved learning a variety of content and took place with students of different ages, a recurring theme was that students had difficulty comprehending and interpreting the visual representations used in instruction.

A variety of explanations have been offered to explain students’ weak diagram comprehension (9, 14, 26), and these explanations can be understood by considering the processes that are involved in this comprehension. Kriz and Hegarty (18) identified these processes in a model that describes the top-down and bottom-up influences on diagram comprehension. Top-down processes are those that are driven by prior knowledge, learner expectations, and higher-order processes, such as strategies. Bottom-up processes are those that are driven by the encoding of perceptual features. According to Kriz and Hegarty, diagram comprehension begins with bottom-up processes when the learner attends to and perceives visual features of the diagram. Features are selected for encoding and subsequently used to construct an internal representation of the diagram. At the same time, top-down processes are being applied to this internal representation. These top-down processes include the student’s prior topic knowledge, as well as general comprehension strategies (10). Together, bottom-up and top-down processes interact to influence the student’s final understanding of the diagram. A student, for example may read symbols or conventions from a novel diagram and apply prior knowledge to aid interpretation, even if this prior knowledge application is erroneous (35).

Bottom-up processes are particularly important in educational settings, because students are often using diagrams to learn about topics for which they have little prior knowledge. Students must not only be able to select and encode the main features of a diagram, but also rely on the ability to accurately interpret the meaning of these features to understand the meaning of the diagram. Diagram conventions are a particular class of these features that play a major role in diagram comprehension. Conventions are symbols, which can be either verbal or graphic, that are used to convey meaning. Examples include color coding to match objects to a legend, lines that label objects, and arrows to show movement. Conventions can also include graphic techniques, such as showing an object repeatedly to illustrate how it changes over time, cutting the surface of an object to expose a cross section, and scaling objects spatially to depict relative size. Students comprehend a diagram by “reading” these conventional elements (29). To understand the diagram in Fig. 1, for example, a student would need to read the lines and text that label objects and detect the cut surface depiction of a cross section. The student would also need to understand how color and arrow conventions indicate that a portion of the plant is shown under magnification and follow these signals to trace the sequence of steps.

As the example in Fig. 1 shows, diagram comprehension can require interpretation of multiple conventional elements. Unfortunately, understanding these conventions can pose significant difficulty for novices (2, 35, 42). One reason for these difficulties is that conventions can have inconsistent meanings. In some cases, the same convention can be used to mean different things. Arrows, for example, can be used to indicate direction in space or time, causality, or sequence (40), but diagrams rarely explicitly indicate which of these interpretations is correct. This point is illustrated in work by Wright et al. (42), who examined the arrows that were used in each of the figures included in an introductory biology textbook. Their analysis revealed 21 distinct types of arrows (e.g., thin curved arrow, merging block arrow, branching block arrow) that conveyed 12 different meanings (e.g., movement, change over time, input/output) with no consistent relationship between the type of arrow used and the intended meaning.

In addition, to variations in appearance and meaning, a convention can also vary according to its role in the diagram. Specifically, a convention that is meaningful in one diagram may not be so in another. The use of color provides a good illustration of this point. In some diagrams, color conveys critical information, such as showing true differences in structures (e.g., muscle and connective tissues are different colors) or a consistent depiction of different processes or concepts (e.g., positive numbers are black, negative numbers are red). At other times, however, color variations may be arbitrary or have no real meaning at all (e.g., different colors for different organs of the body). As a result of these inconsistencies, the “significance of color, if any, is often not self-evident” (Ref. 14, see pg. 174).

A third way in which conventions can pose difficulties is when students do not know how to interpret them. Many
college science students, for example, have difficulty understanding cut surfaces to show cross sections and call outs that magnify smaller portions (5, 19) (see Fig. 1 for an example). Cohen and Hegarty (7) demonstrate this point with college science students’ performance on a test of cross-section interpretation. Test items show participants a three-dimensional object, such as a cylinder or pyramid, being cut by a plane. Respondents select the multiple-choice answer depicting the resulting cut surface. Although there was a range of performance on this test, the average score was <70%, indicating that interpreting this convention is challenging for college science students. Similarly, students can have problems interpreting space and temporal conventions in diagrams. As Eilam (13) points out, for instance, the diagram of a cell shown on one part of a page could occupy the same amount of space as a diagram depicting the respiratory system. Differences in the spatial scales of these diagrams can make it difficult for students to understand relative sizes. This problem is particularly important for biology education because biological knowledge often requires understanding hierarchically organized levels of nested structures (39). The cell, that is, is nested within the respiratory system.

As this brief review demonstrates, diagrams play a critical role in biology education. Although we often assume that students are able to understand and learn from these visualizations (2, 3, 16, 19), the reality is that many students struggle to use them effectively (e.g., Refs. 6, 13, 37). Research summarized here shows that, even when presented diagrams contain all of the information necessary to answer experimental questions, student performance is relatively poor (35, 42). In fact, studies that have quantified this performance demonstrate that average scores are consistently <70% (7, 37), a threshold that would be barely passing in many college courses. There is also evidence, however, that students possess a range of diagram comprehension abilities. Not surprisingly, for instance, prior knowledge can support diagram comprehension so that students with more experience or more content-related knowledge are better able to reason with visual representations (8, 33). Likewise, both interviews and assessments of students’ reasoning with visualizations show a range of performances, indicating that at least some students do well with these representations, at least some of the time (7, 14, 26).

The present study. The main objective of the present research is to examine college biology students’ ability to comprehend scientific diagrams and the diagram characteristics that affect this comprehension. Our goal is to provide instructors with descriptive information about their students’ capabilities that can inform instructional decisions about the selection and use of visual representations (16). By better understanding their students’ abilities to work with diagrams, instructors will be better positioned to provide the support needed to improve their students’ comprehension of specific diagrams, as well as to build the general comprehension abilities necessary to support comprehension of novel diagrams.

Participants in this study were students from a human physiology course who completed a multiple-choice diagram comprehension test (DCA) (10) as part of a larger research project (41). Items on this test present a diagram, and students answer the accompanying comprehension question. Examples of questions include asking students to identify the best caption for a diagram, determine the meaning of a convention, and locate a particular element within a diagram. Three-fourths of the test comprised diagrams from biology disciplines, and the remaining diagrams examined transfer to another science discipline; i.e., geosciences. Research hypotheses test descriptive questions concerning students’ general diagram comprehension abilities and the factors that influence this comprehension.

Our first hypothesis is that students’ average performance will be generally low. Although there is not a specific cut point for determining low performance on this DCA test, we selected 75% as the threshold level to descriptively quantify low performance. This estimate is based on average scores reported in previous research on graph (37) and cross-section (7) interpretation tests, as well as typical grading scales applied to college exams. A 75% on a course exam, therefore, would earn a grade of “C”, the lowest grade possible for full course credit (e.g., required grades in major courses, required grades for transfer credit). Although a 75% on a course exam may be acceptable, this is a relatively poor performance on this test for two reasons. First, all test diagrams were taken from high school textbooks so that each one is at a level that college students are expected to handle. Second, all test items could be answered by using only the information given in the provided test diagram. Relevant prior knowledge could certainly aid diagram comprehension (18), but no test question required prior knowledge, and all could be answered by accurately reading the features of the provided diagram.

Our next two hypotheses are derived from the top-down processes of diagram comprehension (18); namely, general academic abilities and familiarity or prior experience. Specifically, hypothesis 2 predicts that DCA is related to other indicators of academic ability. We hypothesize that there will be significant correlations between scores on the DCA test and grade point average (GPA) and Scholastic Aptitude Test (SAT) scores. Significant correlations between these measures would demonstrate that diagram comprehension is influenced by the same abilities that contribute to other aspects of academic performance. Although not directly measured in this study, these shared abilities would include top-down processes, such as strategy use, metacognition, and inference generation.

Hypothesis 3 predicts that DCA will be influenced by students’ familiarity with the discipline. The diagram comprehension test includes 9 items that assess comprehension of geosciences diagrams, in addition to 25 items that use biology diagrams. We predict that the biology student participants in this study will score significantly higher on biology diagram test items compared with geosciences test items. As noted above, no test item required prior knowledge to interpret the diagram and answer the question. Although this prior topic knowledge is not necessary, we do expect familiarity with the content area to influence diagram comprehension, so that biology students are better able comprehend biology diagrams.

The remaining hypotheses focus on bottom-up processes by examining the effects of diagram characteristics on comprehension (18). Specifically, we predict that diagram comprehension will be influenced by diagram conventions (hypothesis 4) and complexity (hypothesis 5). To test these hypotheses, all test items were coded for the conventions used in the diagrams. Diagrams were classified according to convention type, and comparisons of performance across convention categories evaluate the influence of different types of conventions on comprehension. Complexity was determined by counting each
different type of convention used in a single diagram with more types of conventions, indicating a more complex diagram. The complexity hypothesis is evaluated by statistical comparisons of students’ scores on the test items at different levels of diagram complexity.

Finally, hypotheses 6 and 7 predict that there will be differences in how students with high (H-DCA) and low DCA (L-DCA) are affected by diagram conventions (hypothesis 6) and complexity (hypothesis 7). These hypotheses are tested by using scores on the DCA test to identify groups with stronger and weaker DCA. Performances across different diagram conventions and levels of complexity are compared with determination of how these diagram characteristics affect performance across the two student groups.

METHODS

Participants and materials. Participants were 318 undergraduate students from a human physiology course at a large, public university in the eastern United States. Most participants were in their first (n = 191, 60.1%) or second year of college (n = 74, 23.3%). The majority of participants were life sciences majors (e.g., biology, bio-behavioral health, kinesiology; n = 216, 67.9%), but the population also included nonbiology majors (e.g., nonbiological engineering, business, economics; n = 75, 23.6%), and undecided or unspecified (n = 27, 8.5%). A majority of participants identified as female (n = 205, 65.1%) and as Caucasian (n = 235, 73.9%). All participants were recruited from their course to take part in a larger study that tested manipulations designed to improve learning from text and visualizations (41). The diagram comprehension measure that comprises the data examined in this study was administered to test for possible interactions between these experimental manipulations and individual differences in DCA. Although the full experiment included five conditions, the diagram comprehension scores examined in the present study have collapsed across conditions because this test was administered at the start of the study, before any experimental manipulations related to this research took place. This study was reviewed by the Institutional Review Board, and students provided informed consent for participation and use of their data.

The DCA measure is a 34-item, multiple-choice test assessing comprehension of biological (n = 25) or geosciences (n = 9) diagrams (9). All test diagrams come from high school textbooks, and all test items could be answered given the information provided. That is, all questions could be answered using the bottom-up processes of reading diagram features without relying on top-down, prior knowledge-driven processes (18). A variety of diagram types were tested, including schematic, photographic, and representational. All test items required reading and understanding diagram features, such as labels, captions, color, and axes. Questions assessed low-level, literal comprehension. Examples of DCA test questions include asking what color shows in a stained micrograph, for the main idea that can be generated from a captioned diagram, and the differences between structures in two cross sections. Figures 1–3 show examples of the types of diagrams that were assessed, although these diagrams were not the exact diagrams included on the DCA test. Using Fig. 1, examples of comprehension questions consistent with those on the test are 1) asking where the macrospore ovule is located so that students must interpret both the cross section and the arrows. DCA scores reached an acceptable level of internal consistency; α = 0.68.

Table 1 shows the nine diagram conventions that were found across all test diagrams and points to examples of these conventions in the sample diagrams of Figs. 1–3. Two raters coded every diagram for the presence of these conventions, and exact agreement was 92%, with a Cohen’s κ of 0.82, indicating a high level of agreement between the two raters. Each diagram was then classified according to the conventions it contained, and test scores for each category were calculated to evaluate the effects of conventions on comprehension. Because more than two-thirds of the diagrams contained more than one convention, these diagrams were counted in more than one category.

The characteristic of diagram complexity was operationalized by counting the total number of unique types of conventions included in a single diagram. Coding showed that, across all diagrams, the number of conventions used ranged from one to seven, with an overall mean of 3.59 conventions per diagram. Geosciences diagrams were slightly more complex than biology diagrams, including around one more convention per diagram on average.

We used an extreme-groups split to identify groups of participants with L-DCA and H-DCA. In this method, participants’ scores on the DCA test are used to divide them into three equal groups of low, medium, and high DCA. Only participants scoring in the lower and upper thirds of this distribution are retained to form extreme low and high groups. Specifically, L-DCA participants scored between 5 and 22, and H-DCA participants scored between 26 and 33; participants scoring between these values were not included in the extreme groups’ analyses. This method ensures that groups are composed of participants who truly differ on the measured construct. Only data from participants in the L-DCA and H-DCA groups were analyzed.

| Convention Label | Definition | Example in Figs. 1–3 |
|------------------|------------|---------------------|
| Captions         | Text that provides context beyond the title. | Fig. 1 contains a caption beneath the diagram. |
| Naming labels    | Label used to provide the name of the object or concept. | Fig. 3 uses naming labels to indicate different parts (e.g., Axin, G-protein). |
| Explanatory labels | Label used to describe a process or further define a naming label. | Fig. 2 uses explanatory labels to elaborate some steps (e.g., step 3, step 6). |
| Arrows           | An object used to draw attention; show direction or a process. | Fig. 1 uses arrows to show the relationship of magnified parts to the whole and the sequence of steps. |
| Coloring coding  | Color emphasizes a component or differentiates multiple components of the diagram. | Fig. 3 uses color coding to represent proteins across different parts of the diagram. |
| Letter/number systems | A letter or number is used to represent phases, steps, or a missing object. | Fig. 2 uses numbers to represent the order of steps. |
| Shading          | Shading to emphasize a component or differentiate multiple components. | Fig. 3 uses letters to represent different pathways. |
| Symbols          | A symbol that carries meaning beyond what is present in the diagram (e.g., mathematical or chemical notation, degrees). | Fig. 1 uses different shades of green to depict the inside and outside of the plant. |
| Cut outs         | A cut surface or call out shows a detailed internal view. | Fig. 3 uses symbols (e.g., +) to represent different chemicals in Wnt signaling process. |

Advances in Physiology Education • doi:10.1152/advan.00146.2018 • http://advan.physiology.org
Table 2. Descriptive statistics for the diagram comprehension measure

| Section   | No. of Items | Minimum Score | Maximum Score | Mean Score | SD    | % |
|-----------|--------------|---------------|---------------|------------|-------|---|
| Biology   | 25           | 5             | 24            | 17.72      | 2.85  | 70.87 |
| Geosciences | 9          | 0             | 9             | 5.91       | 1.89  | 65.69 |
| Total     | 34           | 5             | 33            | 23.63      | 4.03  | 69.50 |

Findings from the tests of hypotheses 2 and 3 are consistent with predictions derived from the top-down processes of diagram comprehension (18). First, significant correlations between DCA test scores and GPA and SAT scores suggest that diagram comprehension shares some common processes with the abilities that influence students’ performance on other indicators of academic performance. Second, biology students’ stronger ability to comprehend biology diagrams relative to geosciences diagrams indicates that diagram comprehension is influenced by prior knowledge and experience, a finding consistent with previous research on diagram comprehension (17, 26).

### Diagram conventions and diagram comprehension ability

To test the fourth hypothesis, each test item was coded for the conventions included in the diagram for that item (Table 1). This coding did not consider whether the convention was necessary to answer the specific question associated with the diagram, since any and all conventions present in a given diagram may be used as the student worked toward a general understanding of the diagram. Items were then categorized according to convention type, and average correct scores were calculated separately for each convention category. The majority of individual test diagrams were included in more than one category, because each test diagram was included in the categories for every coded convention contained in that diagram. The second column of Table 3 shows the total number of items that were included each convention type. All conventions appeared in both biological and geosciences diagrams, and more than two-thirds of the diagrams included more than one convention.

We evaluated the influence of conventions on diagram comprehension by examining participants’ average scores across the categories. These averages, which are shown in Table 3, suggest that conventions do affect diagram comprehension. When the most difficult convention (letters/numbering system) is compared with the easiest convention (shading), there is a score difference of 10%. Average scores on diagrams

Additional tables and figures may be included to support the findings and conclusions.
containing the other seven conventions are distributed fairly evenly within that 10-percentage point range. This range and distribution indicate that performance on diagram comprehension test items is affected by the conventions that must be interpreted.

Closer inspection of item scores across the convention categories reveals some patterns in how conventions may affect comprehension. Although arrows have often been identified as difficult for students to comprehend (42), this convention was actually one of the easiest for students in this study. Comprehension items that used diagrams containing arrows and shading conventions were the easiest for participants. By contrast, student participants had the greatest difficulty with diagrams that included captions, cutouts, and letter/numbering systems. That these three conventions were the most challenging is not surprising. First, captions appeared so frequently, in over 90% of the diagrams, that participants’ difficulty with this convention is likely a reflection of weak overall performance rather than difficulty with captions specifically. Second, participants’ relatively low scores on items that used cutouts are consistent with other research, which has shown that these cut surfaces can be challenging to visualize (6, 7, 25).

Finally, the difficulty of diagrams that used letter/numbering systems may have been caused by high cognitive task demands. High demand is created when students must process multiple elements and their relations (38). In one DCA test item that used a letter system, for example, participants had to identify two muscles, labeled A and B, respectively. Answering the question required not only identification of the two muscles, but also remembering which was A and which was B. Maintaining and manipulating all of these elements creates high cognitive task demands, a characteristic that can negatively affect diagram comprehension (e.g., Refs. 12, 16, 17).

The sixth hypothesis predicted that there would be differences in how L-DCA and H-DCA participants would be influenced by the conventions in a diagram. Table 3 shows the percent correct scores for the diagrams in each convention category calculated for each group. These average scores indicate the relative difficulty order of different conventions and the magnitude of the score differences across the two ability groups. In addition to these averages, we also calculated deviation scores to show the magnitude of a convention effect within, rather than between, L- and H-DCA groups. These deviation scores were calculated by determining the average total test score separately for each of the two groups and then determining the difference between that total test average and the average score for items within each convention category. These deviation scores, which are shown in Fig. 4, show how far above or below the mean the average scores for each convention are for the L-DCA and H-DCA groups.

The descriptive data contained in Table 3 and Fig. 4 reveal that both DCA groups were affected by the same conventions in similar ways, but that the magnitude of the effect tended to be greater for the L-DCA group. In terms of similarities, these data show that both groups had their strongest performance on diagrams that included the same three conventions; i.e., shading, arrows, and explanatory labels. Likewise, the same three conventions proved to be the most difficult for both groups; i.e., cutouts, captions, letter/numbering systems.

The major differences between the two ability groups are in regard to the magnitude of the effect exerted by different conventions. It seems that L-DCA participants’ comprehension is more strongly affected by variations in the conventions found within diagrams. First, the range of scores across conventions is 50% larger for the L-DCA participants compared with H-DCA participants, which indicates more variance in performance across conventions for the lower ability group. Second, when conventions are ranked from the easiest (1) to the most difficult (9) for the two groups, the linear trend across these rankings reveals that the L-DCA participants show a marked decline in performance much sooner than H-DCA participants. For example, L-DCA participants’ scores decrease 4.4% between just the conventions ranked with the highest and second highest scores. For the H-DCA participants, a drop-off this large is not found until the easiest convention is compared with the sixth ranked convention.

In addition, L-DCA participants had relatively more difficulty with both the symbol and letter/numbering conventions compared with their H-DCA peers. As described earlier, both groups struggled with letter/numbering systems. L-DCA participants, however, had a marked disadvantage on diagram items that employed this convention. This disadvantage is even greater for the symbol convention. As depicted in Fig. 4, H-DCA participants scored above their average on diagrams.

![Fig. 4. Chart displaying deviation scores for low-(L-DCA) and high-diagram comprehension ability (H-DCA) groups across categories of conventions. The zero point in this graphic represents the group average, and the bar length represents the deviation from the mean for each convention.](image-url)
containing symbols, but scores on these diagrams were a little below average for L-DCA participants. Because both of these conventions require that information from one part of a diagram be related to another part, perhaps L-DCA participants are more strongly affected by the cognitive demands of these conventions.

**Diagram complexity and diagram comprehension ability.** **Hypothesis 5** predicted that diagram comprehension would be affected by complexity, and **hypothesis 7** predicted that students with H- and L-DCA would be affected by diagram complexity differently. To explore the complexity-comprehension relationship, test diagrams were divided into categories of low (1–2 conventions), medium (3–4 conventions), and high (5–7 conventions) complexity. A mixed-model ANOVA, with ability level as the between-groups variable and complexity as the repeated measure, tested hypotheses 5 and 7. Specifically, the main effect of complexity in the ANOVA tested the prediction that complexity would influence diagram comprehension, and the ability group by complexity interaction tested the prediction that the effects of complexity would not be the same for students at different levels of ability. Because a significant interaction qualifies interpretations of main effects in ANOVA techniques, we first examined the interaction and then proceeded to appropriate tests of the complexity main effect. Table 4 shows the percent correct scores at the three complexity levels. As these statistics show, there is a linear relationship between complexity and comprehension, where scores increase as diagrams become more complex.

Although this linear trend is visible for both groups, the ability by complexity interaction was statistically significant, which indicates that the effects of complexity were not uniform across ability group. [F(1.90, 435.80) = 3.66, P = 0.03, \( \eta^2_p = 0.02 \).] The degrees of freedom were adjusted using the Huynh-Feldt ε to minimize type I error due to a violation of the assumption of sphericity; all other conditions were met. As the means given in Table 4 show, H-DCA participants’ scoring advantage increased as the complexity of the diagrams increased. On average, H-DCA participants scored 21% higher on the least complex diagrams, but this advantage increased to 27% on the most complex diagrams.

We decomposed this interaction through follow-up simple-effects tests, which compare scores across levels of complexity separately for the two ability groups. These follow-up tests serve to both break down the interaction and test the prediction that complexity would exert a significant effect on comprehension (hypothesis 5). These tests show that there is a significant effect of complexity for both ability groups: H-DCA, [F(2, 234) = 56.27, P < 0.001, \( \eta^2_p = 0.33 \); L-DCA, F(1.72, 192.67) = 12.22, P < 0.001, \( \eta_p = 0.10 \)]. In both groups, low-complexity diagrams were significantly more difficult than medium-complexity diagrams; medium-complexity diagrams were significantly more difficult than high-complexity diagrams. Although the pattern is the same for the two groups, the effect size values from the simple-effects test are consistent with finding of an ability-complexity interaction. Namely, the complexity effect is stronger for the H-DCA group than for the L-DCA group. Although the finding that increased complexity aids comprehension may seem counter-intuitive, this pattern is consistent with research by Kragten et al. (17). Students benefit, it seems, when diagrams include a variety of conventions to guide comprehension. The significant interaction, however, suggests that H-DCA students are better able to take advantage of these supports than their L-DCA peers.

**DISCUSSION**

Visualizations play a central role in the communication of scientific knowledge (2, 8). College science students encounter these representations in instructional material, during both lectures and laboratories, and across tasks and assessments (13, 14). The present study was motivated by a concern that students may not be learning from these representations to the degree possible because they do not accurately comprehend the diagrams common in science classrooms (2, 42). If students cannot adequately comprehend these representations, then they also cannot fully develop knowledge of the depicted concepts.

Our results show that this concern is warranted. The physiology students who participated in this study evidenced significant weaknesses in diagram comprehension. First, overall performance on the diagram comprehension test was low; equating to a very low C or D grade from most instructors. Even H-DCA participants, who had the highest overall scores, were unable to comprehend one out of every five diagrams, on average. These scores are particularly striking due to the nature of the DCA test. Not only were all diagrams taken from high school textbooks, it was also the case that test questions required only low-level comprehension. In addition, although prior topic knowledge could certainly aid comprehension of a diagram, this knowledge was not required. All test questions, therefore, could be answered simply by interpreting the information provided. In short, test diagrams and comprehension tasks were ones that college science students would be expected to handle.

Study results also support the conclusion that both the top-down and bottom-up processes identified by Kriz and Hegarty (18) influence diagram comprehension. With respect to top-down processing, significant correlations between DCA test scores and both GPA and SAT scores indicate that college students who can accurately comprehend scientific diagrams also show evidence of greater academic success. We suspect this relationship can be attributed to shared top-down processes that influence both diagram comprehension and other academic performances; processes such as strategy use, inference generation, and metacognition. Further evidence for the role of top-down processes in diagram comprehension comes from the finding that biology students were better able to comprehend biology diagrams than geosciences diagrams. As previously noted, test items did not require prior knowledge, but this does not mean that either prior topic knowledge or familiarity with biology diagrams could not aid performance. Indeed, the pres-

---

Table 4. Descriptive statistics by complexity

| Complexity | No. of Items | Overall | L-DCA Group | H-DCA Group |
|------------|-------------|---------|-------------|-------------|
| Low        | 10          | 64.18 (14.75) | 52.66 (13.43) | 73.90 (10.71) |
| Mid        | 16          | 69.65 (13.54) | 57.25 (10.32) | 80.93 (8.56) |
| High       | 8           | 75.83 (18.28) | 61.39 (18.42) | 88.67 (10.07) |
| Total      | 34          | 69.50 (11.86) | 56.87 (8.86)  | 80.68 (4.43)  |

Values are means (SD) in %. L-DCA and H-DCA, low- and high-diagram comprehension ability, respectively.
ent study suggests that these top-down processes do boost students’ diagram comprehension.

The effects of bottom-up processes are revealed in tests of hypotheses related to diagram conventions and complexity. First, we found that diagram comprehension was affected by the conventions a diagram contains. The two conventions that posed the most difficulty were the use of a letter or numbering system where arbitrary letters/numbers were used to signify objects and cut-outs that showed cross sections and magnified interior views. Physiology students’ difficulties with these two conventions are not surprising. As previously noted, the letter/numbering systems likely increase the cognitive demands of the comprehension task by requiring students to keep track of the labeling system. A significant body of research demonstrates that learning decreases when characteristics of the task or materials create this type of extraneous, or unnecessary, cognitive load (e.g., Refs. 12, 15, 21). Our finding that students struggled with cut-outs that showed interior and magnified views is also consistent with previous research. Students have particular difficulty visualizing the plane through which an object is cut and mapping the interior view to the object (7, 25).

This discussion of relative difficulty should not be interpreted to mean that some conventions are easy or pose no difficulty for students. The average score for test diagrams that contained the easiest convention, shading, was still only 76%. Altogether, these data show that instructors should not assume that college science students know how to read the conventions that communicate meaning in diagrams.

Given students’ relatively weak DCA, it is tempting to believe that they should be provided with simplified diagrams. The data obtained in this study does not support this instructional decision. College students’ comprehension improves with an increase in the number of conventions that a diagram uses. Students, it seems, may have difficulty with a particular subset of conventions, but they are able to take advantage of the supports these communication devices provide.

These complexity effects must be interpreted in light of how complexity was defined in the present study. Our definition of complexity follows the description of complex biology diagrams of Cromley et al. (9), where more complex diagrams include more conventional features. Consistent with research by Kragten et al. (17), we found that comprehension improves as this type of complexity increases. This is not the only way to define complexity, however. Butcher (4), for example, defined complexity according to the amount of detail in the diagram and found that college students learned more from simpler diagrams. When the complexity of a visual representation is caused by the inclusion of unnecessary detail, students have difficulty distinguishing relevant from irrelevant information (4, 15, 26). In addition, Kragten et al. (17) argue that complexity can also be defined according to the comprehension or learning task. While the DCA test in this study assessed only low-level comprehension, tasks can also require higher-level comprehension. These high-level comprehension tasks require global inferences that draw across given information. Kragten et al. (17) report that outcome performances typically decrease as task complexity increases. In summary, the present study shows that diagram complexity can aid comprehension when complexity is defined according to the number of relevant, communicative features a diagram contains.

Although this interpretation of complexity effects is consistent with prior research, there are other possible explanations. Perhaps, for example, students spend more time with more complex diagrams because more time is needed to unpack all of the features. In this case, it is the time spent attending to and working to make sense of a diagram, rather than complexity itself, that explains this effect. Alternatively, it may not be that an increase in the use of conventions helps with diagram comprehension so much as a lack of conventions actually harms comprehension. This possibility suggests that diagrams that contain fewer conventions may simply be less clear in their meaning and thus more difficult to accurately comprehend. While the present study does not include the data necessary to determine which of these explanations is the underlying cause of the complexity effect, these data are consistent with the conclusion that complexity, as defined by the number of conventions used, aids diagram complexity.

Altogether, however, these conclusions must be qualified by both our definition of poor performance and the range of DCA found in this sample of physiology students. First, there is no objective standard, or cut score, that distinguishes good and poor performance on this or any other measure of visualization comprehension. We selected 75% as a descriptive value, with this selection guided only by scores on similar measures in previous research and our own expectations of academic performance. Despite this limitation, we interpret the obtained average score of 69.5% as poor because of the nature of the test itself. Simply put, this test is designed such that one with strong DCA could obtain a perfect score, even if the content of the diagrams was novel. In fact, some of the H-DCA participants did nearly achieve this perfect score.

A second, related point is that there was a range of performance among the students who participated in this study. The lowest test scores for this population were below chance, whereas the highest scoring students obtained nearly perfect scores. Comparisons across groups of students with the strongest and the weakest abilities revealed that these two groups are affected by the same diagram characteristics but differ in regard to the magnitude of the effect. For example, both L-DCA and H-DCA students did better on test items as the complexity of the diagram increased. H-DCA students, however, were able to take greater advantage of increased complexity. From the least to the most complex diagrams, scores for the H-DCA group increased by 15%, whereas L-DCA participants saw a much smaller gain of just under 9%. A similar pattern was found when examining diagram convention effects: both groups were influenced by the same conventions, but the influence was stronger for L-DCA participants.

Collectively, data from this study suggest that students with weaker diagram comprehension skills have more difficulty when the characteristics of the diagram increase the cognitive demands of the task. We have already noted how an increase in cognitive load could affect the ability to use certain conventions. Difficulty managing cognitive load could also explain why L-DCA students were less able to exploit the additional conventions in the more complex diagrams. Perhaps these students were less able to process and coordinate all of these communicative features.

The attribution that cognitive demands affected diagram comprehension in this study is only speculative. Additional research measuring cognitive load during diagram comprehen-
sion is necessary to fully understand this relationship. Nonetheless, our main findings are clear: college science students’ diagram comprehension abilities are not reliable. These findings, which are consistent with prior visualization research (5, 37), call into question the degree to which students can be expected to independently learn from instructional diagrams. A substantial portion of students in any particular course are likely to be struggling with the most basic diagram learning tasks.

Limitations. Three limitations to this study have already been noted; namely, the limits on our ability to identify the cause of the complexity effect, the lack of a measure of cognitive load, and the use of a subjective score to identify poor performance. An additional limitation is that the DCA test was designed to provide a general indicator of diagram comprehension and thus did not systematically vary or isolate diagram conventions. As a result, it is not possible to determine how specific conventions affect comprehension in isolation or exactly why some conventions were more difficult than others. Additional research is needed to further explore students’ knowledge of, and ability to use, conventions. Another limitation is tied to the fact that study participants were drawn from only a single human physiology course. Despite this limitation, there is reason to believe that our results generalize beyond the current research. In particular, the participants in this study represented a range of life sciences majors, suggesting that our results are relevant for the broad set of college students who are typically found in biology courses. This point, along with the consistency of findings between the current and previous studies, suggests that the diagram comprehension abilities uncovered in this research are pervasive among college science students.

Implications. Two main conclusions can be drawn from this research, with these conclusions holding significant implications for college science instructors. First, college science students are not able to comprehend many of the diagrams from which they are expected to learn. Second, these comprehension weaknesses highlight the need for students to receive instruction on how to understand and use the visual representations found in different science disciplines. With these conclusions, we join the chorus of voices that have pointed out the need for explicit instruction on the comprehension of visualizations (e.g., Refs. 3, 14, 22, 23, 36). Such instruction could take many forms, but there are two key recommendations that we can offer here.

First, we recommend that instructors be explicit about the use of visualizations. By explicit, we mean that instructors should directly and clearly discuss the use of visualizations with their students. This explicit instruction should point out how to read the visualization by explaining how to interpret the critical conventions it contains. Instructors should also inform students of a given visualization’s function. These explanations can include information about the learning objective of a visualization, as well as how that representation can be connected to other, related representations. Examples of these instructional efforts can be found in recent work on college biology education. Angra and Gardner (3), for instance, developed an instructional guide to support students’ understanding of how to use and interpret graphs that depict data. A second example is Crowther’s teaching of a graph drawing technique to support learning of electrochemical driving forces (11). The present study indicates the need for more efforts like these that explicitly teach students how to use visual representations.

A second recommendation is for instructors to be thoughtful about the visualizations that are selected and how these are used. Primarily, that is, we hope that an increased awareness of students’ difficulties with visualizations would affect how instructors use visualizations. Consider, for example, a course exam that uses visualizations for some test questions. Recognizing that students may have knowledge of the concept but have difficulty comprehending the visual representation of that concept, instructors can think about the degree to which a test item actually assesses the targeted knowledge. Does the test item assess course content knowledge, DCA, or both? Certainly, there are circumstances in which the ability to read a diagram is a critical part of content understanding. At other times, however, the visualization may not be necessary. Careful consideration of this point could lead instructors to design assessments that capture the desired abilities. In this respect, our suggestion is consistent with the recommendations of Khalil and Elkhider (16) that, when designing course material, instructors do not assume that students can understand the representations in that material. Instead, we agree with the suggestion that instructors first test the material to determine whether students will understand the intended meaning (16).

This thoughtful use of visualizations also extends to how these representations are used over the course of instruction. If a visualization is important for understanding a concept, for instance, then we would expect that visualization to be important enough for the instructor not only to explicitly explain it, but also to include it in meaningful learning activities. Examples of these activities include homework assignments in which students’ must explain visualizations, clicker questions embedded in lectures that require students to answer questions about visualizations, and assignments that require students to generate their own visual representations of target concepts. Finally, we would also encourage instructors to use “wait time” in which they allow students the opportunity to stop and think about the meaning of a visualization before proceeding to explanations of the depicted concepts. Students could be directed to use this time to read and comprehend the visualization so that the following explanation can be connected to the visual knowledge.

As these recommendations indicate, it is unlikely that a single lesson on diagram comprehension or a single approach to embedding these representations into instruction will be sufficient to address students’ abilities to understand and work with these representations. In this respect, teaching students to learn from visualizations is no different than teaching them to learn from other sources, such verbal text or experimental data. By increasing our own awareness of the value of these representations and students’ corresponding abilities to learn from them, however, we can improve their ability to acquire scientific knowledge and learn about the communication practices within these disciplines.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

P.V.M. conceived and designed research; P.V.M. and C.C. performed experiments; A.M.K. and C.C. analyzed data; A.M.K., P.V.M., and C.C. interpreted results of experiments; C.C. prepared figures; A.M.K., P.V.M., and
C.C. drafted manuscript; A.M.K., P.V.M., and C.C. edited and revised manuscript; A.M.K., P.V.M., and C.C. approved final version of manuscript.

REFERENCES

1. Ainsworth S. DeFT: a conceptual framework for considering learning with multiple representations. Learn Instruc 16: 183–198, 2006. doi: 10.1016/j.learninstruc.2006.03.001.
2. Anderson TR, Schönborn KJ, du Plessis L, Guptar AS, Hull TL. Identifying and developing students’ ability to reason with concepts and representations in biology. In: Multiple Representations in Biological Education, edited by Tregast DF and Tsui CY. Dordrecht, the Netherlands: Springer, 2013, p. 19–38. doi: 10.1007/978-94-007-4192-8.
3. Angra A, Gardner SM. Possible constraints of visualization in biology: challenges in learning with multiple representations. In: Multiple Representations in Biological Education, edited by Tregast DF and Tsui CY. Dordrecht, the Netherlands: Springer, 2013, p. 55–73. doi: 10.1007/978-94-007-4192-8.
4. Baker M. Relative effectiveness of animated and static diagrams: an effect of learner prior knowledge. Comput Hum Behav 24: 852–861, 2008. doi: 10.1016/j.chb.2007.02.018.
5. Balacheff N, Bruner JS, Coblenz S. How does one teach to think? Educational Studies in Mathematics 7: 191–213, 1976.
6. Bevc M, Golfinos JG, Caron A, Cullinane E, Winer B, et al. Cross-sectional anatomy of the human brain. Neuron 67: 3–44, 2010. doi: 10.1016/j.neuron.2010.05.007.
7. Baylis S, Murphy D. Using visualization to train neuroradiologists. J Radiol 94: 357–360, 2013. doi: 10.1051/jradiol/2012042.
8. Bloom BS. Taxonomy of Educational Objectives. Handbook I: The Cognitive Domain. New York: David McKay Co., 1956.
9. Bloom BS. A Taxonomy for Learning, Teaching and Assessing. New York: Longman, 1994.
10. Canham M, Hegarty M. Effects of knowledge and display design on comprehension of complex graphics. Learn Instruc 20: 155–166, 2010. doi: 10.1016/j.learninstruc.2009.02.014.
11. Cohen CA, Hegarty M. Individual differences in use of external visualizations to perform an internal visualization task. Appl Cogn Psychol 21: 701–711, 2007. doi: 10.1002/acp.1344.
12. Crowther GJ, Snyder-Hogan LE, Luciw-Dubas UA. Experts’ views on translation across multiple representations in textbooks and journal articles. J Coll Sci Teach 41: 1785–1800, 2013. doi: 10.1080/09500690701670535.
13. Crowther GJ, Thomson A. Top-down and bottom-up processes. Top Cogn Sci 3: 560–578, 2011. doi: 10.1111/j.1756-8765.2009.01066.x.
14. DeFrisco M, Turchi G, Zanini C. Diagrammatic literacy in college biology course. J Biol Educ 49: 188–198, 2015. doi: 10.1007/s10857-015-9603-5.
15. Derry SL. Cognitive activity theory: perspectives from a cross-disciplinary research team. Learn Instruc 24: 852–861, 2016. doi: 10.1016/j.learninstruc.2015.09.002.
16. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
17. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
18. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
19. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
20. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
21. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
22. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
23. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
24. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.
25. Derry SL, Fox R, Reigeluth CM. Multiple representations in learning and instruction: a review of current research. Educ Psychol 49: 1–22, 2014. doi: 10.1080/00461520.2013.830767.