Special Series: Scientific Literacy

JMBE Journal of Microbiology & Biology Education

Visual Literacy of Molecular Biology Revealed through a Card-Sorting Task

Dina L. Newman,[a] Hannah Spector,[a] Anna Neuenschwander,[a] Anna J. Miller,[a] Lauren Trumpore,[a] and L. Kate Wright[a]

[a]*Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA*

Visual literacy, which is the ability to effectively identify, interpret, evaluate, use, and create images and visual media, is an important aspect of science literacy. As molecular processes are not directly observable, researchers and educators rely on visual representations (e.g., drawings) to communicate ideas in biology. How learners interpret and organize those numerous diagrams is related to their underlying knowledge about biology and their skills in visual literacy. Furthermore, it is not always obvious how and why learners interpret diagrams in the way they do (especially if their interpretations are unexpected), as it is not possible to "see" inside the minds of learners and directly observe the inner workings of their brains. Hence, tools that allow for the investigation of visual literacy are needed. Here, we present a novel card-sorting task based on visual literacy skills to investigate how learners interpret and think about DNA-based concepts. We quantified differences in performance between groups of varying expertise and in pre- and postcourse settings using percentages of expected card pairings and edit distance to a perfect sort. Overall, we found that biology experts organized the visual representations based on deep conceptual features, while biology learners (novices) more often organized based on surface features, such as color and style. We also found that students performed better on the task after a course in which molecular biology concepts were taught, suggesting the activity is a useful and valid tool for measuring knowledge. We have provided the cards to the community for use as a classroom activity, as an assessment instrument, and/or as a useful research tool to probe student ideas about molecular biology.

KEYWORDS molecular biology, card sorting, conceptual understanding, visual literacy, visual representations

INTRODUCTION

Many subdisciplines of biology, including molecular biology, cell biology, and genetics, are invisible to the naked eye. Experts in these fields rely on the creation and the interpretation of visual representations to communicate and investigate concepts and processes (e.g., gene regulation, recombination) central to these domains (1–3). Experts in these fields also use visual representations to communicate with biology learners (novices) to help them grasp and learn foundational concepts. Much research has shown, however, that students (at all levels, including graduate students) are not proficient in comprehending and gaining correct meaning from visual representations (4–11). Another challenge when using visual representations as teaching and learning tools is with the visuals themselves. Most textbook illustrations of molecular processes, for example, are quite colorful and complex, but do learners know what to look at and how to interpret the shapes and illustrations that they see? For example, less than 1% of arrow-containing figures in biology textbooks included a key to help learners (11). Illustrations and representations are, out of necessity, oversimplifications of the actual phenomenon, but illustrators have to make choices about what is highlighted and what is omitted in each representation. Learners are supposed to productively use these visual tools to help fill in the gaps of their own mental models of scientific processes, but if the symbols used are more confusing than helpful, learning may be hindered (12). For instance, the process of DNA replication may be showcased as a simplified replication fork that omits an image of the DNA polymerase enzyme and nucleotide substrates. Alternatively, a different figure about DNA replication may show the entire DNA polymerase holoenzyme and other coordinating proteins at a replication fork. Do learners "see" these figures as representations of the same process? Can learners differentiate between unimportant (stylistic) and important (conceptual) features of visual representations in biology? Can learners correctly interpret symbols (e.g., lines and boxes to represent gene structure) and connect their own mental models of the phenomena to the image they are viewing?

Scientific literacy is not only about the comprehension and correct application of scientific concepts (13) but also "foundational ways of reading" (14–16), which includes comprehension of scientific information in the forms of text and images (17).

Editor Michael J. J. Wolyniak, Hampden-Sydney College
Address correspondence to Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA. E-mail: lkwsbi@rit.edu.
The authors declare no conflict of interest.
Received: 28 October 2022, Accepted: 13 January 2023, Published: 31 January 2023

Visual literacy, which is the ability to effectively identify, interpret, evaluate, use, and create images and visual media (18), is a part of science literacy. It would seem as if extended exposure to visual representations throughout an undergraduate biology curriculum should result in the development of scientific visual literacy skills, but Schönborn and Anderson (1) pointed out that, "Students do not necessarily automatically acquire visual literacy during general instruction." In other words, students need opportunities to practice and develop their visual literacy skills, and instructors and researchers need tools to better understand where learners are in the development process (19–21).

Decades ago, Chi and colleagues (22) conducted a landmark experiment in which they demonstrated the different ways in which novices and experts in physics organized conceptual physics problems. Novices sorted problems based on surface features (e.g., the problem involved a ramp or a projectile), while experts sorted based on the underlying conceptual theory of the problem (e.g., the problem was about Newton's first law). This sorting task provided a novel way to measure how novices and experts organized their discipline-specific knowledge; experts had the ability to look past surface details and find the root meaning of the problems, while novices struggled to do this. Card-sorting tasks, as these types of knowledge-probing methods are called, require participants to group or sort items (such as index cards containing various word problems) into groups based on their own ideas. How participants categorize and name groups is then used a proxy to their underlying knowledge about a particular subject. This work in physics has inspired a number of other card-sorting tasks to measure conceptual expertise in other science, technology, engineering, and math (STEM) disciplines, such as the Biology Card-Sorting Task (23, 24), in which subjects were asked to analyze and sort biology word problems into logical groups. In this work, novices were anchored on the particular model organism presented in each problem (e.g., fruit fly or yeast) instead of the underlying biological concept (e.g., evolution). As the experience level of the participants increased, their ability to look beyond surface features and find the foundational concepts increased. Similarly, the Chemistry Card-Sorting Task (25) has been shown to differentiate conceptual understanding of novices and experts on concepts related to thermodynamics, equilibrium, kinetics, and structure-function relationships. The cards in the Chemistry task include word problems with an accompanying chemical formula for participants to analyze and then sort. In addition to the qualitative analysis, the researchers used the measure-of-edit distance (26) to quantify the differences in how individuals and groups scored on the tasks.

Based on classroom artifacts, teaching observations, and prior research studies, we wanted to investigate visual literacy skills of biology students. Specifically, we wondered whether students were capable of seeing past surface details to identify underlying concepts in common visual representations in molecular biology, particularly ideas that center around DNA (e.g., replication, mutation, etc.). We built upon our prior experiences investigating visual representations in molecular biology (27, 28) to create a variety of images to use and test with students and experts. Similar to the previously described card-sorting tasks, we also implemented the strategy of combining a "surface" feature with an underlying concept. Instead of using text or chemical formulae, we created and tested visual representations of concepts in molecular biology. We attempted to characterize the level of visual literacy that students had regarding DNA-based representations by specifically investigating the following questions:

- Can students recognize underlying DNA-based concepts when different visuals are used?
- How does visual literacy about DNA change with experience?

Here, we present a novel, research-backed tool to help biology researchers and instructors measure visual literacy skills in students learning concepts related to molecular biology. The tool, called the DNA Visualization Card-Sorting Task, is comprised of image-containing cards covering the topics of DNA replication, DNA repair, gene expression, and mutation, which we refer to as "deep features." Each of the 4 topics is illustrated using one of five particular surface features (chemical structure, sequence, helix, boxes and lines, and chromosome structure) for a total of 20 cards. The cards were tested with introductory and intermediate biology students as well as biology experts. We found that the ability to sort based on deep features increased as participants gained more experience in biology.

METHODS

An overview of the process of development and testing of the cards is shown in Fig. 1. Additionally, a synopsis of each of the three main experiments is included.

Development of the cards

In the first stage (beta card set), we developed images for cards based on a matrix of 6 surface features (type of drawing) by 5 deep features (concepts) that reflected 30 images to be used in this card-sorting activity. The hypothesized surface feature categories were chemical structure, sequence, ladder, helix, box and line, and chromosomal. The surface levels were based on previous work in which we explored the various ways in which DNA could be represented in introductory and advanced biology textbooks (28). Our hypothesized deep conceptual categories were DNA replication, mutation, gene expression, meiosis, and DNA repair, which are topics covered in almost all introductory biology, molecular biology, and cell biology textbooks. The construction of the images was reflected from common figures found in biology textbooks, but none was taken directly from published resources. Once we had the initial beta card set, we recruited local experts in biology and biochemistry ($n = 10$) for in-person consultations. Experts were defined as individuals who had a Ph.D. in molecular biology or a related field (e.g., genetics,

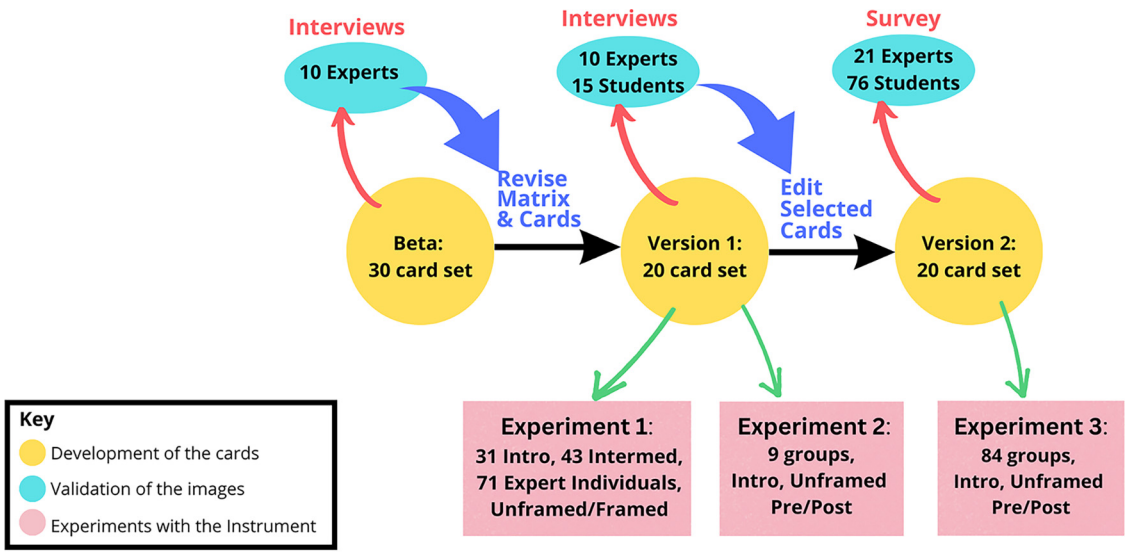

FIG 1. Flow chart illustrating the major steps of the project. A 30-card beta set was created based on experience and prior research. Revision based on interviews with experts resulted in version 1, a 20-card set. Version 1 of the card-sorting task was validated through interviews and then used in several experiments with introductory students, intermediate students, and biology experts. Further refinement of the images led to version 2 of the cards, which were validated through surveys with experts and novices and then used in an experiment with introductory biology students. Black arrows show changes over time in card design. Red arrows show which version was used for each type of validation. Green arrows show which version was used for each experiment. Blue arrows show how the results of each type of validation were used.

biochemistry). Experts were asked to describe what they saw in each card and to point out anything that was confusing or misleading to them. We then revised the cards to remove the helix level and the meiosis concepts, resulting in version 1, a set of 20 cards.

Ethics statement

All studies involving human subjects had institutional approval (HRSO 02050819).

Unframed and framed card-sorting tasks

Within the unframed sorting task condition, participants were given the images of the cards either as physical cards, in a PowerPoint with each card on a different slide, or on a Google Jamboard with movable images. Deployment of the cards depended on the situation in which testers were recruited (i.e., in-person beta testers, online course at outside institution, online course at home institution, or in-person classes). As shown in Fig. 2, each card was labeled with a letter for identification purposes. Participants were instructed to sort the cards into as many groups as they liked using whatever criteria they desired. They were also prompted to assign a descriptive name to each group. When completing the activity in-person, participants could rearrange the physical cards into groups; when PowerPoint was employed, participants used the slide sorter to move cards around and added extra slides for group names; when Jamboard was used, participants could drag and drop images, enlarge them if needed, and write on the board to label groups. Research

		Surface Level Categories				
Deep Level Categories		Chemical Structure	Nucleotide Sequence	Double Helix	Box and Line	Chromosome
	DNA Replication	B	L	M	T	R
	Gene Expression	S	G	F	E	C
	Mutation/ Evolution	N	I	D	J	O
	DNA Repair	A	P	K	H	Q

FIG 2. Card-sorting task matrix. Hypothesized deep features include concepts, while surface features include types of drawings. Each card labeled A to T depicts a single deep feature using a single surface-level feature.

subjects or lab assistants entered the responses into the Collection and Analysis of Research Data for Sorting (CARDS) web portal (https://atom.calpoly.edu/cardsort/about). For trials within the framed sorting task condition, the same cards and physical or digital formats were used; however, for this condition, the names of the hypothesized deep categories were provided. In other words, the participants were instructed to sort the cards into four groups: DNA replication, mutation, gene expression, and DNA repair. These responses were also entered into CARDS. The framed task was generally implemented immediately after the unframed task was completed.

Pre- and postcourse testing

Pre- and postcourse testing was performed by assigning the unframed task (as described above) on the first day of class and then again toward the end of the semester (weeks 12 to 14). In these trials, students were placed in groups of 3 to 5 individuals, and the same groups were used for both pre- and postcourse testing.

Percent pairings

Percent card pairings (following the method of Smith et al. [24]) measured the frequency of two cards being placed together within a group. These could be surface-level pairings, deep pairings, or unexpected pairings. Surface and deep pairings were two cards that belonged to the same surface or deep category, respectively. Unexpected pairings were comprised of two cards that did not share a surface feature or a deep feature. In each of the four 5-card surface groups, there were 10 possible pairs ($4 \times 10 = 40$ surface pairs). In each of the five 4-card deep groups, there were 6 possible pairs ($5 \times 6 = 30$ deep pairs). All remaining pairs were considered unexpected (190 total pairs -40 surface -30 deep pairs = 120).

Edit distance

Edit distance, as introduced by Deibel et al. in their 2005 paper (26), is a measurement of the minimum number of cards that would need to be moved in order to complete either an exact deep or surface sort (Smith et al. [24]). An exact sort includes the participant-created groups with only the cards contained within each row or column of the matrix for a surface sort (ED-Surface) or deep sort (ED-Deep), respectively. With an exact hypothesized surface sort, ED-Surface = 0 and ED-Deep = 40. Similarly, with an exact hypothesized deep sort, ED-Surface = 30 and ED-Deep = 0. Edit distance was calculated for each participant in every trial of the card-sorting task.

Validation interviews using version 1 of the cards

In this context, we considered experts in the field to be the gold standard for interpretation of the card images. To validate that experts interpreted the images as we intended, we conducted semistructured interviews with 10 experts (defined as individuals with a Ph.D. in biology or fields related to molecular, developmental, genetics, or cell biology). In addition, we predicted that less experienced individuals would differ from experts in their interpretation. Thus, we interviewed 15 undergraduate biology students (defined as individuals who had taken at least 1 year of a college introductory biology course) as a comparison group Participants were recruited by contacting professional and social networks. As interviews were conducted over Zoom, we first created a PowerPoint presentation of the cards with one card image per slide (random order). Participants were asked to sort the cards in a way that made sense to them (unframed task) using the slide sorter function on PowerPoint and to name their groups. They then explained their reasoning for why groups were created and how each card fit within the group. The interviews were recorded and transcribed. For each individual card, the team recorded the group designation (e.g., "DNA replication") and reasoning provided by experts and students. Two researchers (not the interviewer) then read through all the descriptions and categories and coded each as correct or incorrect. The "correct" designation meant that the card was either placed in the expected deep sort category or the participant gave an accurate description of the process on the card. The "incorrect" designation, for our study, did not necessarily mean the subject had no knowledge about the process being depicted on the card but, rather, that the subject performed an unexpected interpretation or provided an insufficient explanation about the image.

Synopsis of experiment 1

Experiment 1 consisted of a framed and unframed approach to testing card version 1 with individuals of various levels of expertise who analyzed the cards using edit distance and percent pairing. Biology and biology-related majors were recruited from a large, private institution in the northeastern United States to participate in the in-person experiment. Students were either in their first (introductory; $n = 31$) or second year (intermediate; $n = 43$) of study in biology or a related program. Individuals with a Ph.D. in biology or fields related to molecular, developmental, genetics, or cell biology (experts; $n = 71$) were recruited from a listserv of an organization devoted to biology education research to participate in the online version of the sorting task.

Synopsis of experiments 2 and 3

For experiment 2, an unframed approach was used to test card version 1 in groups as a pre- and postcourse assessment. Pre- and postcourse testing was performed by assigning the unframed task on the first day of class and then again toward the end of the semester. Honors-level biology students ($n = 36$; placed in 9 groups of 4) completed the in-class group activity in week 1 and week 12 of their semester. The same groups were used for both pre- and postcourse testing. The same protocol was followed in experiment 3 using Card Version 2, with a

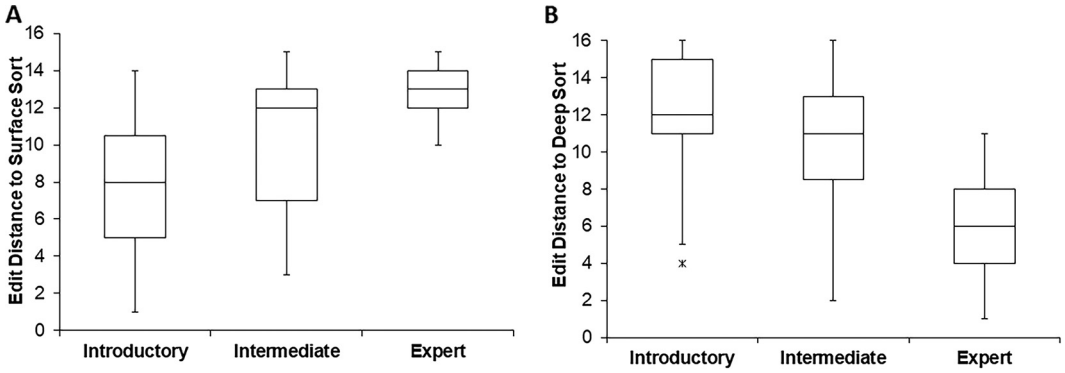

FIG 3. In experiment 1, an unframed card-sorting task revealed that more experienced subjects moved away from the surface level toward deep sorting. (A) Edit distance to surface sort for introductory students ($n = 31$), intermediate students ($n = 43$), and experts ($n = 71$). (B) Edit distance to deep sort for the same populations. Analysis of variance (ANOVA) results: $P < 0.00001$.

different group of students. These students were Introductory Biology students at a large University in the Northwest US.

Validation surveys using version 2 of the cards

Results from the interview on version 1 of the cards resulted in a final round of revision. Some revisions were minor (e.g., making the letters in a strand of DNA sequence larger), and others were more substantial (e.g., redesign of an image of gene expression at the chromosomal level). To validate the major changes, we developed an online Qualtrics survey to investigate how experts and biology students deciphered the images on cards C, F, J, L, P, and Q, which underwent revision (see Fig. S1 in the supplemental material for images of version 2 of the cards). Card R, which did not undergo revision, was also included in the survey to double-check that the image was a clear representation of replication at the chromosome level. We recruited survey participants by online campaigns through social media and professional organizations of biologists and biology educators. Our survey included demographic questions for experts (i.e., Ph.D. in biology fields related to molecular, developmental, genetics, or cell biology) and students (undergraduates who had taken at least 1 year of college biology) to ensure individuals with adequate backgrounds were included. Survey participants were presented with one card at a time and asked, in an open-response format, to explain what they thought was being represented by the illustration. To reduce time spent taking the survey, each subject was presented with only 4 of the cards (in a randomized order). More than half of the experts ($n = 21$) and more than half of the students ($n = 76$) provided a written explanation of each card. Survey data were exported, and written explanations were coded as correct or incorrect by two members of the research team.

RESULTS

The DNA visualization card-sorting task matrix is shown in Fig. 2, where each card (A to T) lies at the intersection of one surface and one deep category. Final versions of all cards can be found in Fig. S1. Challenges related to creating the cards included (i) limiting each image to only one hypothesized surface and deep feature, (ii) maintaining uniformity of surface features throughout the deep categories, and (iii) minimizing word use to avoid students focusing on the text rather than the images.

Implementation of version 1 of the cards

To investigate how students (introductory and intermediate) and experts sorted the cards, we gave the unframed and framed card-sorting tasks to different populations. We calculated ED-Surface (Fig. 3A) and ED-Deep (Fig. 3B) for all three groups. Our analyses demonstrated two things: distance from the surface sort increased with experience, and distance to the deep sort decreased with experience. In other words, introductory students often used the appearance of images in their sorting, while experts primarily used the underlying concepts.

When participants were given the framed sorting task, i.e., they were given the category names and then asked to sort into 4 groups, not surprisingly, all groups got closer to the deep sort compared to the unframed sort. Furthermore, Fig. 4 shows the same pattern as the unframed sort: ED-Deep decreased with experience, with experts sorting the cards almost exactly as expected.

Edit distance provided a measure of the strategies used by participants to sort cards: did they categorize images based on superficial characteristics or by underlying deeper concepts? We also explored pairing frequencies, i.e., how often two cards were put together in the same group by various populations (introductory students, intermediate students, and experts). We found, in general, that experts tended to pair cards together based on conceptual features and students were more likely to pair cards based on superficial features. In Fig. 5 we showcase the pairing frequencies for cards about DNA replication (B, L, T, M and R; see Fig. S1 for the images). Ideally, any of the cards in this set should have a high pairing frequency with any other card within this set. As expected, the pairing frequencies for

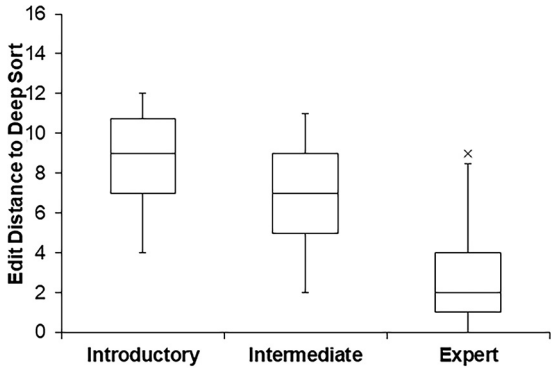

FIG 4. In experiment 1, ability to sort by deep categories in the framed condition improved with experience. Edit distance to deep sort was analyzed for introductory students ($n = 31$), intermediate students ($n = 43$), and experts ($n = 71$). ANOVA results: $P < 0.00001$.

any of the pairs within this 5-card set ranged from 0.8 to 0.95 for experts. Introductory students paired DNA replication cards together in a wider range, from 0.233 to 0.7. Intermediate students had a slightly narrower range, compared with introductory students, for DNA replication cards, 0.39 to 0.731. Introductory students had particular trouble correctly pairing card B (chemical structure of DNA replication) with anything else (range, 0.233 to 0.30), but experts correctly paired card B with the 4 other cards more than 80% of the time.

We also explored all of the pairs made in the data set. As we had anticipated, subjects with more biology experience made fewer surface pairs and a greater number of deep pairs

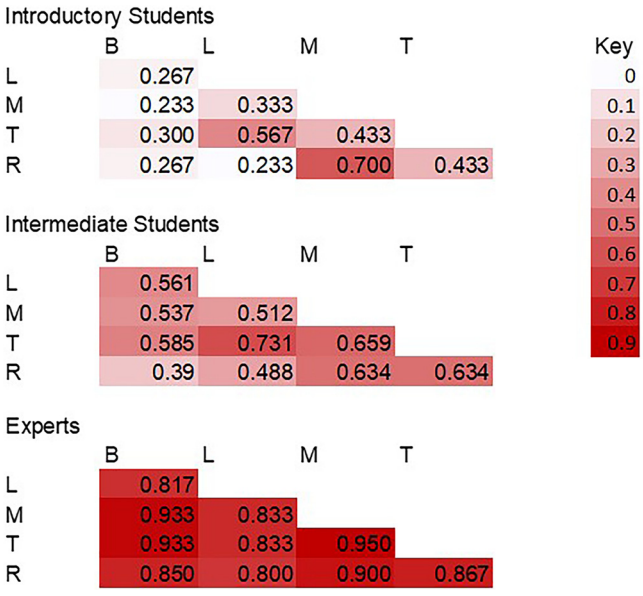

FIG 5. For experiment 1, pairing frequencies for cards about DNA replication were analyzed. Heat maps show frequencies of pairings for the DNA replication deep category for introductory students ($n = 31$), intermediate students ($n = 43$), and experts ($n = 71$) with the framed condition using the version 1 of the cards. The lightest colors represent the lowest frequencies, and the darkest shades correspond to the highest pairing frequencies.

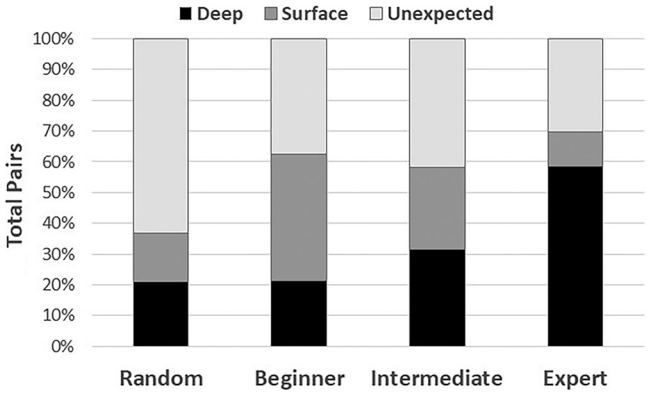

FIG 6. For experiment 1, percentages of unexpected, surface, and deep pairings made by introductory students ($n = 31$), intermediate students ($n = 43$), and experts ($n = 71$) were analyzed. The percentages in the random category show the proportion of pairs that would be expected by chance to fall into each category. Each group's distribution was significantly different from random and from each other by χ^2 test.

(Fig. 6). We calculated the proportion of pairs that would be expected by chance to fall into each category and found the introductory students made the same number of deep pairs as would be expected by chance but more surface pairs than would be expected by chance. Intermediate students had increased deep pairings and decreased surface pairings, while the majority of pairs made by experts were classified as deep pairs. These data support our previous observations that introductory students seemed to identify surface similarities while experts identified the conceptual similarities.

Pre- and postcourse testing

Until this point in our study, all data were obtained from individual participants. With increased attention being paid to active learning and group activities in STEM classrooms, we wanted to explore the card-sorting activity in a group setting. In week 1 of the semester, students in a highly structured, active learning-based, honors version of Introductory Biology were randomly assigned into groups of 4 and were asked to sort cards (unframed setting). Later, at the 12-week point, the students were put back into their original groups and given the unframed task again. The instructor recorded the sort data after each time point. Figure 7 shows that by the end of the course, students were moving away from using surface features and toward using deep features. We also found our introductory students who worked in groups started with significantly more deep pairs than expected by chance, which was higher than either the introductory or intermediate students who worked alone in the previous experiment (38% in groups compared to 20 to 32% in individuals [Fig. 6]).

Validation of version 2

To validate our revised cards (C, F, J, L, P, and Q) plus card R, we developed an online Qualtrics survey to investigate how experts and biology students deciphered the card images.

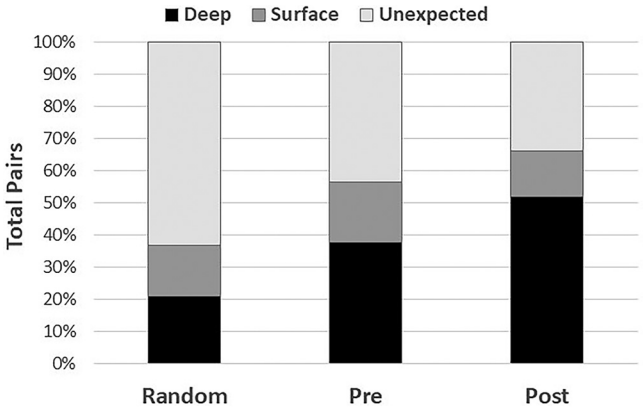

FIG 7. For experiment 2, percentages of unexpected, surface, and deep pairings made by introductory students (n = 9 groups) during weeks 1 (precourse) and 12 (postcourse) of their introductory biology course. The percentages in the random category show the proportion of pairs that would be expected by chance to fall into each category.

Survey participants were presented with one card at a time and asked, in an open-response format, to explain what they thought was being represented by the illustration. The vast majority of experts correctly identified the concept of the card (Fig. S2). Students, on the other hand, described the concepts correctly much less frequently. Card Q was the one exception for which we did not see a large difference between experts and learners. The image on card Q shows DNA repair at the chromosomal level by showing a "broken" chromosome undergoing a repair event. Participants who were incorrect in their descriptions often described telomere shortening or DNA damage as the underlying concept. We hypothesize that participants who incorrectly described this image as damage (instead of repair) did not consider the direction of the arrow between the two chromosome images.

Finally, we tested version 2 of the cards in the unframed condition in a pre- versus postcourse scenario (Experiment 3). Students in an introductory biology course completed the card-sorting activity at the beginning and end of the semester. Analysis of edit distance (Fig. 8) revealed dramatic results; students increased their edit distances to the surface sort and decreased edit distances to the deep sort. In other words, after instruction, students were better able to recognize deep conceptual features represented in the card images.

DISCUSSION

The DNA visualization card-sorting task is a novel, research-backed tool that was designed to investigate visual literacy skills of students when interpreting visuals about DNA-based representations. The images for the cards were developed by first examining common representations of DNA (in processes such as replication and gene expression) found in undergraduate biology textbooks. We categorized figures based on the style in which the DNA was drawn: chemical structure, DNA sequence, box and line, helix, and chromosome; we then designed figures that represented particular concepts drawn in each of the five styles. The cards were tested and revised with various populations of participants and are now in a format that can be used by the broader biology education teaching and research community. In this study, we sought to answer the following two research questions (RQs): (RQ1) Can students recognize underlying DNA-based concepts when different visuals are used? (RQ2) How does visual literacy about DNA change with experience?

Regarding RQ1, we found that students could recognize DNA-based concepts, but they did so inconsistently. Students are often misled or distracted by the style of representation used, such as chemical or chromosome drawings, and focus on those superficial details as they interpret and sort various images. For example, in our implementation of version 1 of our cards, introductory students rarely placed cards B (DNA replication at the chemical structure level) and L (DNA replication at the sequence level) in the same group, even both cards illustrated the same concept. These same students, however, placed cards L with P (DNA repair at the sequence level) 65% of the time, most likely because both images featured chemical structure representations of DNA. Card E (gene expression at the box and line level) did not have a high frequency of pairing with any of the other cards about gene expression. Instead, students paired it the most frequently with card J (mutation at the box and line level), likely because

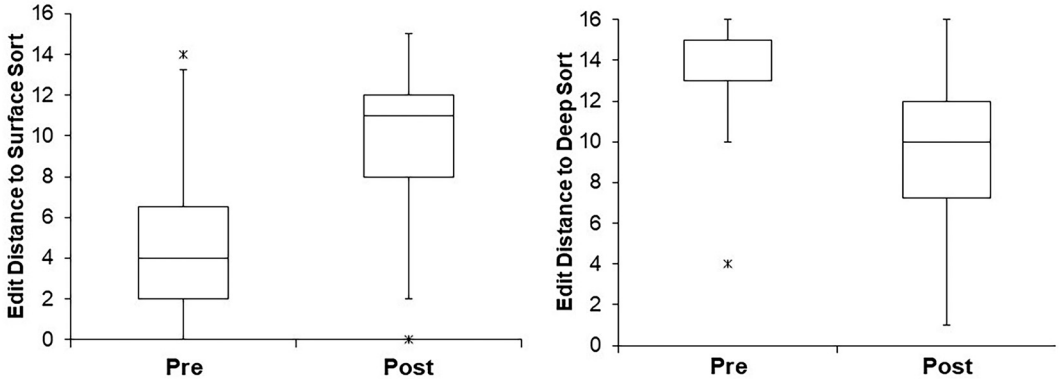

FIG 8. In our validation of experiment 3, pre- to postcourse changes in edit distances in the unframed condition (with version 2 of the cards) demonstrated learning for introductory biology students (n = 84). Results of t tests: $P < 0.00001$ for edit distance to surface and to deep sorts.

of the similarities of gene structure (box and line) used in both diagrams. Experts, on the other hand, were much better at recognizing DNA-based concepts in the visual form and almost always interpreted the images as we had anticipated. Experts did make some unexpected pairings and did not always complete the sort exactly as we anticipated, but interviews helped reveal some of the reasoning behind their choices. For example, experts put cards J (mutation at the box and line level) and C (gene expression at the chromosome level) together at a higher frequency than anticipated. In this case, experts focused more on the "expression of a product" concept in J, even though they also recognized a mutation was also being illustrated, and so chose to put the card with C. An earlier version of Q (DNA repair at the chromosome level) was correctly described as a repair event by an expert, but it was placed in a category of genetic engineering, as a genetic engineering approach could be used to repair the damage. The incidence of small variations from the "perfect" deep sort was not unexpected and has been observed in other STEM card-sorting tasks (24, 25).

Our answer to RQ2 is tied in with the types of inferences we can make about performance on this instrument. In the context of assessments, the term valid means that "the concept or characteristic that a test is designed to measure" strongly relates to the inferences and interpretations one can make about performance on that assessment (29). In other words, a valid instrument measures the concept and skill that it was designed to measure. Our evidence strongly suggests the validity of our card-sorting instrument is sound. We found that biology experts routinely outperformed biology students, advanced biology students outperformed introductory students, and students enrolled in a biology course about molecular biology topics improved for postclass compared to preclass performance. Edit distance calculations, card-pairing data, and interpretations of card images supported the relationship between instrument performance and knowledge about the underlying topics. Our results demonstrate that this activity can be used to generate valid inferences about a population of students; more underlying knowledge and experience with molecular-based concepts yield better performance on the task.

Implications for teaching and learning

The number of figures and illustrations encountered by biology learners is enormous. Instructors should routinely ask, "Can my students correctly interpret what they are looking at?"; most probably do not. Instead, instructors (including ourselves) make many assumptions about what their students "see" when viewing scientific representations. We suggest the DNA visualization card-sorting task be a good first step for instructors wishing to better understand how adept their students are at recognizing and interpreting illustrations of DNA-based concepts. Instructors may be surprised to find that their students mix up amino acids and nucleotides, have trouble distinguishing DNA replication from transcription, and cannot "see" gene expression embedded in canonical diagrams of the *lac* operon. Thus, using the card-sorting task as a preclass assessment may help an instructor think about how to best structure classroom activities and discussions to best promote learning. The card-sorting task may also be used as a pre- and postcourse assessment to measure learning gains after a newly designed activity. Other instructors, after reading this study, may be inspired to review the illustrations and representations used in their own teaching and textbook materials, to refamiliarize themselves with the kinds of images their students are using. Instructors may be inspired to ask their students, "How are you interpreting this diagram? Should we walk through this together?" to help build their students' understanding and interpretation skills.

We have also found that the card-sorting activity makes a good first day of class icebreaker activity and an end of semester reflective activity. On day 1, it gets students talking and sharing ideas, and based on the improved performance of groups compared to individuals, we can also say that it promotes peer learning. This sets the stage for students entering into an active learning environment. It can also be used at the end of the term to show students how much they have learned.

Future directions

The card-sorting task could also be used to explore visual literacy about molecular-based concepts in different ways. The cards could be used to investigate group interactions around a visual literacy task, for example. A current limitation of our study is that we did not record what was said and which participant did what during the group card-sorting activities, but future work could focus more deeply on how students worked together during the sorting tasks. In a different future project, an investigator could create groupings of cards and ask participants to identify the card that does not belong to the group and ask for an explanation about why. Or, an investigator could present all of the cards in a particular conceptual group and ask participants to explain the connections between each figure. Other studies might include exploration of how different groups of students interpret the images and how their background experiences may shape their ideas.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

ACKNOWLEDGMENTS

We thank Nicola Barber for testing the cards in her class. Many students and experts also donated their time by trying the activity, answering surveys, and/or participating in interviews. This work was partially supported by NSF DUE-1757477.

REFERENCES

1. Schonborn KJ, Anderson TR. 2006. The importance of visual literacy in the education of biochemists. Biochem Mol Biol Educ 34:94–102. https://doi.org/10.1002/bmb.2006.49403402094.
2. Towns MH, Raker JR, Becker N, Harle M, Sutcliffe J. 2012. The biochemistry tetrahedron and the development of the taxonomy of biochemistry external representations (TOBER). Chem Educ Res Pract 13:296–306. https://doi.org/10.1039/C2RP00014H.
3. Trumbo J. 1999. Visual literacy and science communication. Sci Commun 20:409–425. https://doi.org/10.1177/1075547099020004004.
4. Kozma RB, Russell J. 1997. Multimedia and understanding: expert and novice responses to different representations of chemical phenomena. J Res Sci Teach 34:949–968. https://doi.org/10.1002/(SICI)1098-2736(199711)34:9%3C949::AID-TEA7%3E3.0.CO;2-U.
5. Kozma R, Chin E, Russell J, Marx N. 2000. The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. J Learn Sci 9:105–143. https://doi.org/10.1207/s15327809jls0902_1.
6. Schönborn KJ, Anderson TR, Grayson DJ. 2002. Student difficulties with the interpretation of a textbook diagram of immunoglobulin G (IgG). Biochem Mol Biol Educ 30:93–97. https://doi.org/10.1002/bmb.2002.494030020036.
7. Stieff M, Lira M, DeSutter D. 2014. Representational competence and spatial thinking in STEM. ICLS 2014 Proc 2:987–991.
8. Stull AT, Hegarty M. 2016. Model manipulation and learning: fostering representational competence with virtual and concrete models. J Educ Psychol 108:509–527. https://doi.org/10.1037/edu0000077.
9. Timmerman BC, Feldon D, Maher M, Strickland D, Gilmore J. 2013. Performance-based assessment of graduate student research skills: timing, trajectory, and potential thresholds. Stud Higher Educ 38:693–710. https://doi.org/10.1080/03075079.2011.590971.
10. Wright LK, Fisk JN, Newman DL. 2014. DNA → RNA: what do students think the arrow means? CBE Life Sci Educ 13:338–348. https://doi.org/10.1187/cbe.cbe-13-09-0188.
11. Wright LK, Cardenas J, Liang P, Newman DL. 2018. Arrows in biology: lack of clarity and consistency points to confusion for learners. CBE Life Sci Educ 17:ar6. https://doi.org/10.1187/cbe.17-04-0069.
12. Roth W-M, Pozzer-Ardenghi L. 2013. Pictures in biology education, p 39–53. *In* Treagust DF, Tsui C-Y (ed), Multiple representations in biological education. Springer, Heidelberg, Germany.
13. National Research Council. 1996. National science education standards. https://nap.nationalacademies.org/catalog/4962/national-science-education-standards. Accessed 5 August 2022.
14. Norris SP, Falk H, Federico-Agraso M, Jiménez-Aleixandre MP, Phillips LM, Yarden A. 2009. Reading science texts. Epistemology, inquiry, authenticity: a rejoinder to Jonathan Osborne. Res Sci Educ 39:405–410. https://doi.org/10.1007/s11165-008-9118-5.
15. Krajcik JS, Sutherland LM. 2010. Supporting students in developing literacy in science. Science 328:456–459. https://doi.org/10.1126/science.1182593.
16. Norris SP, Macnab JS, Wonham M, de Vries G. 2009. West Nile virus: using adapted primary literature in mathematical biology to teach scientific and mathematical reasoning in high school. Res Sci Educ 39:321–329. https://doi.org/10.1007/s11165-008-9112-y.
17. Rybarczyk B. 2011. Visual literacy in biology: a comparison of visual representations in textbooks and journal articles. J College Sci Teach 41:106–114.
18. Lundy AD, Stephens AE. 2015. Beyond the literal: teaching visual literacy in the 21st century classroom. Proc Soc Behav Sci 174:1057–1060. https://doi.org/10.1016/j.sbspro.2015.01.794.
19. Airey J, Linder C. 2009. A disciplinary discourse perspective on university science learning: achieving fluency in a critical constellation of modes. J Res Sci Teach 46:27–49. https://doi.org/10.1002/tea.20265.
20. Offerdahl EG, Arneson JB, Byrne N. 2017. Lighten the load: scaffolding visual literacy in biochemistry and molecular biology. CBE Life Sci Educ 16:es1. https://doi.org/10.1187/cbe.16-06-0193.
21. Anderson TR, Schönborn KJ, Du Plessis L, Gupthar AS, Hull TL. 2013. Identifying and developing students' ability to reason with concepts and representations in biology, p 19–38. In Treagust DF, Tsui C-Y (ed), Multiple representations in biological education. Springer, Heidelberg, Germany.
22. Chi M, Feltovich P, Glaser R. 1981. Categorization and representation physics problems by experts and novices. Cogn Sci 5:121–152. https://doi.org/10.1207/s15516709cog0502_2.
23. Bissonnette SA, Combs ED, Nagami PH, Byers V, Fernandez J, Le D, Realin J, Woodham S, Smith JI, Tanner KD. 2017. Using the biology card sorting task to measure changes in conceptual expertise during postsecondary biology education. CBE Life Sci Educ 16:ar14. https://doi.org/10.1187/cbe.16-09-0273.
24. Smith JI, Combs ED, Nagami PH, Alto VM, Goh HG, Gourdet MAA, Hough CM, Nickell AE, Peer AG, Coley JD, Tanner KD. 2013. Development of the biology card sorting task to measure conceptual expertise in biology. CBE Life Sci Educ 12:628–644. https://doi.org/10.1187/cbe.13-05-0096.
25. Krieter FE, Julius RW, Tanner KD, Bush SD, Scott GE. 2016. Thinking like a chemist: development of a chemistry card-sorting task to probe conceptual expertise. J Chem Educ 93:811–820. https://doi.org/10.1021/acs.jchemed.5b00992.
26. Deibel K, Anderson R, Anderson R. 2005. Using edit distance to analyze card sorts. Expert Systems 22:129–138. https://doi.org/10.1111/j.1468-0394.2005.00304.x.
27. Newman DL, Coakley A, Link A, Mills K, Wright LK. 2021. Punnett squares or protein production? The expert-novice divide for conceptions of genes and gene expression. CBE Life Sci Educ 20:ar53. https://doi.org/10.1187/cbe.21-01-0004.
28. Wright LK, Wrightstone E, Trumpore L, Abid D, Newman DL. 2022. The DNA landscape: development and application of a new framework for visual communication about DNA. CBE Life Sci Educ 21:ar47. https://doi.org/10.1187/cbe.22-01-0007.
29. American Educational Research Association. 1999. Standards for educational and psychological testing. AERA, Washington, DC.