Miller-Cotto Dana (Orcid ID: 0000-0003-0577-4868)

Sketching and Verbal Self-Explanation:

Do They Help Middle School Children Solve Science Problems?

Dana Miller-Cotto, Julie L. Booth, Nora S. Newcombe

Temple University

Running Head: SKETCHING AND VERBAL SELF-EXPLANATION

Corresponding author's current address:

Dana Miller-Cotto

damcotto@udel.edu

16 W Main St, Newark, DE 19716

ORCID: http://orcid.org/0000-0003-0577-4868

Acknowledgements: This research was supported by the National Science Foundation through grant #DRL-1252436 to Julie Booth at Temple University, grant #DRL-1560724 to Jennifer Cromley at the University of Illinois at Urbana-Champaign and grant #SBE-1041707 to Nora Newcombe at Temple University. The opinions expressed are those of the authors and do not represent views of the National Science Foundation.

Special thanks are extended to Briana Chang and Taylor Williams for their assistance in stimuli creation, data collection, and coding student work through the project.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acp.3980

Abstract

Both sketching and self-explanation are widely believed to be effective for problem-solving in science learning. However, it is unclear which aspects of these strategies promote learning and how they might interact. Compared to a read-only baseline, we examined the impact of instructing 11-year-old students to solve science problems to sketch, self-explain, or both. Problems were either high spatial or low spatial. We coded elements and relations in their sketches and think-alouds. Self-explanation led to greater accuracy on problems, but only when used alone. Sketching had no effect, although showing more elements and relations in sketches was associated with higher accuracy and occurred more often for high spatial problems. Students may be more comfortable with self-explanation than with sketching and require encouragement or specific instruction to use sketching routinely for the benefits hypothesized to be associated with sketching to appear.

Keywords: sketching, self-explanation, science, problem-solving

Sketching and Verbal Self-Explanation:

Do They Help Middle School Children Solve Science Problems?

Students use a variety of methods to process information and problem solve actively. Common strategies include reading, note-taking, and highlighting, but these are superficial ways of engaging with material (Peverly, Brobst, Graham, & Shaw, 2003). They may be especially ill-suited for understanding scientific concepts, where visual representations are ubiquitous, and comprehension of those representations is key to learning (Madden, Jones, & Rahm, 2011; Pantziara, Gagatsis & Elia, 2009). Students may need strategies to identify important information, generate inferences, and conceptualize spatial relationships. In the present study, we investigate the benefits afforded by sketching. This strategy externalizes spatial knowledge compared to self-explanation, a verbal approach that emphasizes deeper processing but does not easily ground spatial relationships. Sketching involves drawing an external representation of the information contained in the problem (Ainsworth, Prain, & Tytler, 2011). Self-explanation consists in reflecting on and describing the reasoning behind problem-solving decisions (Chi, 2000). In the following sections, we briefly review the literature on each approach and compare the potential benefits each may offer—alone and in combination—for learners' problem-solving performance in science.

Self-Explanation

One of the main benefits of self-explanation is that it requires students to elaborate on relationships between incoming information and their prior knowledge. This elaboration helps students develop a richer mental representation of concepts and principles in a specific domain (Atkinson & Renkl, 2003; Chi, 2000; Roy & Chi, 2005). Effective problem solvers often self-explain spontaneously (Chi, de Leeuw, Chiu, & LaVancher, 1994; Renkl & Atkinson, 2010), particularly with diagrams (Roy & Chi, 2005), likely because it helps them go beyond learning the names of structures to determine how the structures function together or focus on the key elements, or parts of a problem, and relationships, or processes between elements in a given problem (Hmelo-Silver & Pfeffer, 2004). The elaboration principle (Moreno & Mayer, 2010) suggests that learners benefit from responding to "why" questions. Profound processing benefits depend on linking information with prior knowledge, spending effort to find relationships, and creating rules and exceptions during their problem solving and reflecting on these processes. Indeed, some work suggests that self-explanation benefits differ by students' prior knowledge in science; some prior knowledge is still needed to successfully engage in self-explanation (Ionas, Cernusca, & Collier, 2012), suggesting that such self-explanations would be higher quality (Rittle-Johnson, Loehr, & Durkin, 2017). Prompting self-explanation is thus considered to be a productive instructional strategy because it encourages students to engage in deep processing as they generate explanations (Sweller, 2012); such prompts encourage the learner if they are not already doing so, to examine relationships between concepts, and reflect on these procedures (McNamara, 2004).

Despite robust findings of positive effects, there are instances where self-explanation is not helpful (see Berthold, Roder, Knorzer, Kessler, & Renkl, 2011; Kuhn & Katz, 2009). Berthold and colleagues (2011) found that using conceptually oriented prompts to encourage self-explanations encouraged tax law students to spend more time on self-explaining; however, these students also took longer to come up with accurate calculations than peers who were not prompted to self-explain. Thus, it may not be that self-explanation takes more time and is detrimental; instead, it is most useful when calculations or prior knowledge is at a certain baseline. Worse, when Kuhn and Katz (2009) asked pairs of fourth-graders to use information in vignettes to make predictions regarding the causes of earthquakes, children prompted to self-explain performed worse than those without prompting. One possible explanation for these counter findings could be that the efficacy of self-explanation prompts may vary depending on the prompt used (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Indeed, Kuhn and Katz (2009) asked students to engage in generated self-explanations that inferred causal effects of mechanisms compared to those who did not, suggesting a potential role of prior knowledge. Other factors may also influence the effectiveness of self-explanation, including the nature and quality of the learners' responses. Consistent with this assertion, high-quality self-explanations that are associated with strong learning outcomes often include principles (i.e., citing an underlying rule for why an answer must be accurate; Renkl, 2002); poor self-explanations—which do not include much detail, do not attempt to anticipate upcoming steps in a problem, or fail to demonstrate a clear understanding of the goal of a problem—are associated with worse learning (Matthews & Rittle-Johnson, 2009).

Sketching

Though science instruction is generally thought to benefit from the incorporation and use of external representations (e.g., National Research Council, 2012), having learners construct their representations seem to provide an additional benefit for learning (Rellensmann, Schukajlow, & Leopold, 2017; Scheiter, Schleinschok, & Ainsworth, 2017; Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014). Sketching may provide benefits by externalizing information needed to comprehend and solve a problem, activating prior knowledge, increasing students' attention to critical components of the problem, and potentially forcing them to slow down when interpreting a problem (for a review, see Ainsworth & Scheiter, 2021). The physical act of sketching these representations appears crucial; undergraduate students who generated sketches outperformed counterparts who were told to imagine what they were reading mentally or visualizing without sketching (Lin, Lee, Kalyuga, Wang, Guan, & Wu, 2017) or to summarize what they had read (Jaeger, Velazquez, Dawdanow, & Shipley, 2018).

One possible reason for these benefits is that sketching may make spatial relationships between elements more explicit, and may promote the comprehension of multiple representations, for middle school (Bobek & Tversky, 2016), high school (Leopold & Leutner, 2012), and undergraduate students (Jee, Gentner, Forbus, Sageman, & Uttal, 2009; Lin et al., 2017). Engaging the spatial structure of a problem situation by creating predictive sketches yields greater performance and learning than having learners copy representations without considering spatial relations (Gagnier, Atit, Omand, & Shipley, 2016). Further, sketching may provide an additional opportunity to visualize relationships between elements more concretely (i.e., learning through diagrams or illustrations; Schmidgall, Eitel, & Scheiter, 2019), promote inference making, externalize representations to reduce cognitive demands, and force students to be efficient in what they include in their sketch (i.e., only what will help solve the problem; Ainsworth & Scheiter, 2021). However, not all generative sketching is helpful; prompting undergraduate students to sketch relations between elements using standard, prescribed procedures rather than sketching to find the answer to a problem hurts performance (Heckler, 2010). In particular, prescribed sketching discouraged thinking about appropriate informal solution methods for the problem (Kuo, Hallinen, & Conlin, 2017). In addition, constructing inaccurate sketches—due to a failure to represent all relevant information and accurately organize and integrate that information—would be unlikely to support learning (Scheiter et al., 2017). Thus, it may be critical to include specific prompting and scaffolding to encourage students to sketch representations that are helpful to solve the problem (Ainsworth & Scheiter, 2021).

These studies suggest that thoughtful, non-routinized sketching can be a valuable learning strategy, but they do not specify which features of students' sketches are crucial for successful problem-solving. One essential point is that schematic conceptual representations are associated with successful problem solving, whereas pictorial sketches, or decorative sketches lacking relevant content, are related to poor performance (Edens & Potter, 2008; Hegarty & Kozhevnikov, 1999; Van Garderen & Montague, 2003). Additionally, whether individual elements vs. relations between elements are represented in sketches may matter; Sachse, Hacker, and Leinert (2004) found that for undergraduates solving pulley problems in physics, the benefits of sketching were tied to a better comprehension of relations between elements.

Within science diagrams, features of tasks and contexts may impact whether sketching occurs. For example, Uesaka and Manalo (2012) investigated competing hypotheses related to spontaneous sketching among 8th-grade students solving word problems—whether spatial features of the task vs. difficulty level of the task affect if a sketch will be made. They showed that sketches were more likely on one-step problems but no more likely on distance problems vs. less-spatial problems or problems where object orientation was not essential to solving the problem. The authors concluded that students weigh the costs of sketching and the benefits. With this in mind, we also manipulated the spatial features of the problems by including two problems with many spatial features, or high-spatial problems, and two problems with few spatial features, or low-spatial problems.

Generativity Theory

Rather than incorporate both self-explanation and sketching in a single study, most prior work has compared self-explaining *or* sketching to reading or note-taking; several studies of sketching-to-learn have also compared sketching to summarizing (Gobert & Clement, 1999; Leopold & Leutner, 2012; Leopold, Sumfleth, & Leutner, 2013). It is, perhaps, unsurprising that sketching or self-explanation, both of which involve selecting information, organizing it, and integrating it with prior knowledge (Fiorella & Mayer, 2015), are more fruitful than note-taking or summarizing, which at most involve selecting and organizing information (Schleinschok, Eitel, & Scheiter, 2017; Jaeger et al., 2018). These predictions follow from generativity theory (Fiorella & Mayer, 2015), which has been particularly productive in the sketching-to-learn literature, yielding significant effects on factual, inference/comprehension, and transfer learning alike. For example, implementing more generativity-based scaffolds with sketching yields a better transfer of learning than implementing fewer such scaffolds (e.g., draw+select/organize/integrate > draw+select; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010). Generativity theory (Fiorella & Mayer, 2015; Wittrock, 1974) asserts that learners make meaning of new knowledge by integrating it with prior knowledge. The learner must generate and transfer meaning from their background to build connections between concepts of the newly acquired information and existing knowledge. Generation processes, like sketching and self-explanation, encourage learners to select, organize, and associate concepts to infer newly acquired information and existing knowledge (Chi, 2009).

Self-Explanation vs. Sketching

Self-explanation and sketching are both regarded as highly effective, generative strategies for problem-solving. Still, sketching may have a unique role in promoting effective learning by assisting students in representing what may be abstract spatial relations, like solving a science problem regarding the pathway of light through the eye. While self-explanation makes the information in the text more concrete and explicit to the learner, sketching may help learners identify relevant information related to spatial relationships that might be difficult to do in a self-explanation and focus on relevant components of the problem. That is, learners may have difficulty verbalizing spatial relations because of the natural limitations of linguistic representations (Ainsworth & Scheiter, 2021; Scheiter, Schleinschok, & Ainsworth, 2017). It is possible that sketching and self-explanation work via the same mechanisms; increased inferential activity has been asserted to explain gains in both sketching interventions (Van Meter & Garner, 2005) and self-explanation (Chi, 2005), which may indicate a common underlying mechanism through increased meta-cognitive monitoring (van Meter et al., 2006). Beyond this, however, sketching may lead to additional improvements by encouraging students to elucidate abstract spatial relations (Jee et al., 2009), often typical in science education.

Only two studies to date have investigated both approaches simultaneously. Parnafes, Aderet-German, and Ward (2012) found improvement in performance in a classroom intervention in which 4th-8th grade students explained their sketches to others, thus combining the two approaches. More relevantly, Scheiter and colleagues (2017) taught middle school students how to sketch or self-explain while reading science texts and demonstrated how to retrieve essential information from the texts. Students in the sketch group were explicitly instructed to explain the main features of what helped them. In contrast, students in the self-explanation group were instructed to explain the main features to themselves and write down an explanation they could use later. The accuracy of each determined sketch and self-explanation quality. Results indicated that while differences in overall middle school students' performance were not significantly different, students who produced quality sketches outperformed students who produced quality written self-explanations. However, they did not operationally define "quality" in detail; quality was rated on a scale from zero to 20, a rubric indicating several correctly self-explained or sketched units from the problem.

Could sketching and self-explanation together produce greater gains than either alone? There are a few possibilities. First, sketching and self-explanation may work via the same mechanisms. Increased inferential activity has been postulated to explain both gains due to sketching interventions (Van Meter & Garner, 2005) and the effectiveness of self-explanation (Butcher, 2006; Roy & Chi, 2005), which suggests overlap in mechanism, i.e., a final common pathway. In addition, the effectiveness of self-explanation has been explained by increased metacognitive monitoring (Atkinson, Renkl, & Merrill, 2003), and this mechanism has sometimes also been suggested for sketching (Iiskala, Vaura, Lehtinen, & Salonen, 2011; Naug, Colson, & Donner, 2011; van Meter et al., 2006). However, sketching has also been specifically posited to lead to improvement by "assisting students' representations of abstract, spatial relationships" (Jee et al. 2009), which suggests a unique role for sketching (or other interventions that highlight spatial relationships).

Here, we focus on critical components of sketches and self-explanations that may concretely influence their quality: elements and relations. Elements and relations, or thinking about relationships between elements, have been investigated related to relational reasoning. In the next section, we discuss elements and relations as they've been situated in the relational reasoning literature.

Relational Reasoning in Science

Thinking about relationships between elements and then coming up with hypotheses about processes between elements is referred to as *relational reasoning*. Structure-mapping theory (Gentner, 1983) argues that relational reasoning requires mapping between elements of a domain or determining relations between concepts (Dumas, 2017) or elements. For this to be successful, a series of steps must be met. First, the reasoner must understand what aspects (i.e., elements) depend on other aspects to focus on goal-relevant information while ignoring other details (i.e., the functional structure). They must also make sense of hypothetical cause-effect relations to draw inferences from the text. Similarly, concept maps are often a useful tool to elucidate relationships amongst concepts for learners (Horton et al., 1993). It is possible that sketching and self-explanations may be helpful in ways like concept mapping. However, concept mapping often requires demonstrating the connections between prior knowledge concepts and newly acquired information (Novak, 1990). Like in the current study, Novak (1990) asserts that elements are parts of a problem, whereas relations are defined as relationships or processes *between* elements in a problem.

Research Questions

Compared to the control condition, we examined the unique and possible joint contributions of sketching and self-explanation to scientific problem-solving. We concentrated on the following research questions:

1. How do sketching, self-explanation, spatiality (the extent to which object orientation was essential to solving a problem; many objects meaning high spatial, few elements, or low spatial), and students' prior knowledge impact students' problem-solving accuracy?
2. How do sketching, self-explanation, spatiality, and students' prior knowledge impact students' inclusion of elements and relationships in their sketches or verbalizations?

Hypotheses

1. We hypothesized that conditions where students were explicitly told to self-explain or sketch would lead to higher performance than conditions where students were told to simply solve the problems (i.e., the read-only condition) since these strategies have previously been shown to improve problem-solving performance.
2. We hypothesized that performance would be higher on low spatial problems because high spatial problems included more elements often situated in a specific manner in space.
3. Sketching should lead to greater performance benefits on the high spatial problems compared with the low spatial problems because the affordances of sketching align with the spatial nature of these items.

Method

Participants

Participants were 199 sixth graders (mean age = 11 years) from eight middle schools in the Mid-Atlantic and Midwestern regions of the United States. The sample was 53% female and was ethnically/racially diverse (59% White, 17% Black, 5% Asian, 9% Latino, and 10% multiple races). Participants were also socioeconomically diverse, with 21% of students coming from families where neither parent had earned a Bachelor's degree, 26% from families where at least one parent had attended some college, and 29% from families where at least one parent had earned a Bachelor's degree or higher (24% of the sample did not know the level of education their parent/guardian obtained).

Students were randomly assigned across all participating schools to one of the five conditions: read-only (n = 45), self-explanation (n = 43), sketching (n = 43), self-explanation then sketching (n = 34), and sketching then self-explanation (n = 34). We organized each packet of study measures within a folder in the following order: read-only, self-explanation, sketch, sketch, then self-explain, self-explain, then sketch. The next student to participate in the study was assigned the next packet in the deck.

Materials

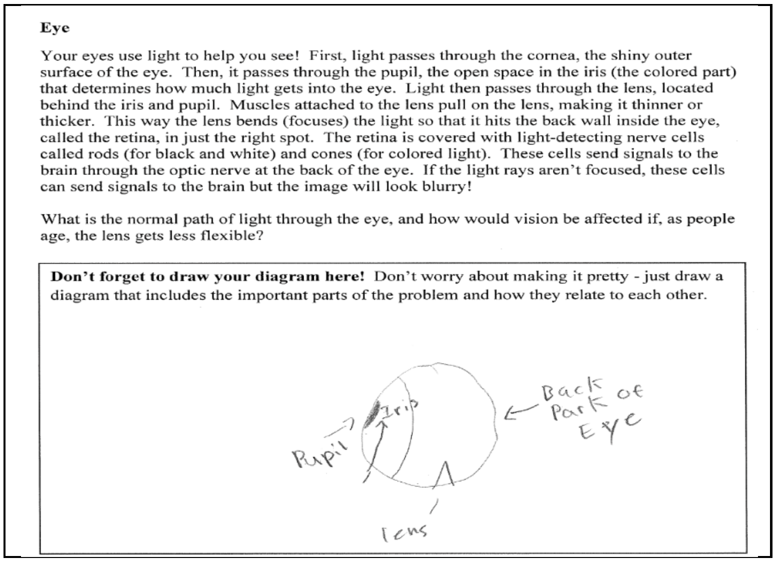
Stimuli

Four science stimuli were created based on middle school science curricula topics. Each problem included a short text on each topic (approximately 150 words each) and a short prompt. Responses to the prompt served as the measure of performance for each problem. The topics included parts of the eye and how light passes through the eye, parts of the heart and how blood passes through it, skin receptors and their functions, and blood cells and their functions. The experimental stimuli were developed through consultation with the curriculum expert and a school principal. The staff at the school verified that the stimuli were grade-appropriate in terms of difficulty, contained accessible vocabulary, and fit within the broader curriculum of the school. Problems were no more than ten sentences long and deemed appropriate for a sixth-grade reading level by our consultants.

The level of the spatiality – high versus low – of the stimuli was manipulated based on the number of spatial dimensions involved, how many spatial relationships the students had to consider, and whether the problem involved movement (i.e., the high-spatial problems involved more spatial dimensions, consideration of multiple spatial relationships, and a degree of movement). A team of science and spatiality experts were surveyed to determine what problems were high- versus low-spatial. Ultimately, one of the high-spatial problems focused on parts of the eye and the path of light through the eye, and the other focused on parts of the heart and the direction of blood flow. One low-spatial problem focused on the functions of skin receptors, and the other focused on blood cells and their functions. The spatial manipulation was also verified by ratings of the university's spatial intelligence psychology lab members. The stimuli were modified based on whether they were considered high spatial or low spatial problems. High spatial problems included elements and relationships with a specific orientation in space to understand the problem, whereas low spatial problems did not have a specific orientation.

The degree to which the science stimuli were amenable to sketching was tested through pilot stimuli administered to the university students in three undergraduate classes. Each item was scored for accuracy. Performance scores were computed as a proportion correct out of the total number of points allotted for all items (i.e., some items allowed for partial credit); these composite scores had an internal consistency of $\omega = .68$. It is perhaps not surprising that the combined scores did not have high internal consistency, as the problems were created to differ systematically in spatiality. Further examination of item reliability statistics indicated that dropping the "Eye Age" item, one out of the four items (see Figure 1), would improve our scale reliability, resulting in $\omega = .7$. However, since this would have resulted in only a marginal improvement in reliability, we opted to retain this item in our overall science performance measure. All problems are provided in the appendix. An example problem with student work is provided in Figure 1.

Figure 1: Sample stimulus for sketch conditions with student work.



Eye

Your eyes use light to help you see! First, light passes through the cornea, the shiny outer surface of the eye. Then, it passes through the pupil, the open space in the iris (the colored part) that determines how much light gets into the eye. Light then passes through the lens, located behind the iris and pupil. Muscles attached to the lens pull on the lens, making it thinner or thicker. This way the lens bends (focuses) the light so that it hits the back wall inside the eye, called the retina, in just the right spot. The retina is covered with light-detecting nerve cells called rods (for black and white) and cones (for colored light). These cells send signals to the brain through the optic nerve at the back of the eye. If the light rays aren't focused, these cells can send signals to the brain but the image will look blurry!

What is the normal path of light through the eye, and how would vision be affected if, as people age, the lens gets less flexible?

Don't forget to draw your diagram here! Don't worry about making it pretty - just draw a diagram that includes the important parts of the problem and how they relate to each other.

Prior Knowledge in Science

A 12-item multiple-choice researcher-constructed measure was used to assess prior knowledge of science concepts related to the stimuli in the study. Items were multiple choice and included questions about the direction of blood flow throughout the body, various components and functions of cells in the eye, features, and functions of cells in the blood involved in healing, and receptors in the skin. For example, for the blood flow question, a prompt was provided for a set of items: "State whether each of the following correctly represents the direction of the flow of blood among the right side of the heart, the left side of the heart, the body, and the lungs" followed by a series of states such as "Right side of the heart → Lungs," "Body → lungs" and "Right side of the heart → body." The path of light or path of blood through the heart, or orientation in space, made this problem "high spatial." Students were to indicate "Yes" or "No" to each statement. Scores across all problems were computed using the total correct out of the total number of items (i.e., proportion correct); student scores ranged from 1% to 100% (M = 62%, SD = 12%). Internal consistency was computed using Cronbach's alpha (α = .90).

Procedure

Students were recruited from their homeroom and typically completed the study in no more than three total sessions. In the first session, students were seen in groups of 4-5 individuals. They were given a brief demographics questionnaire to indicate their age, race, ethnicity, parent's occupation, and the prior knowledge measure. Before beginning these measures, students were reminded of the purpose of the study, provided their assent, and were told that they could stop throughout the study if they changed their mind about participating.

During the second session (and third if the student needed extra time), the procedure was conducted one-on-one with the experimenter and the student. The experimenter reminded the student of the purpose of the study and that they could stop or withdraw at any time. The experimenter then started the audio recording. All students were briefly shown the packet of problems they were going to solve and encouraged to narrate everything they were thinking and doing, including reading, writing, or asking themselves questions. Students were then allowed to practice thinking aloud by imagining where they lived and counting the windows in their home as a sample problem, distinct from the science problems they would be solving. Students were given one minute to narrate everything they were thinking about as they counted the number of windows in their homes. During the practice problem with the windows, if students remained quiet or only responded with the total number of windows, they were reminded to narrate everything they were thinking and speak loudly and clearly. Students were then given feedback on their verbalization, such as "You did a fine job!" or "remember to say *everything* you're thinking," or "don't forget to explain how you arrived at your final answer" if the student was in the self-explanation or combination condition. Then, the experimental manipulation began.

Experimental Manipulation

Students were randomly assigned to one of the five conditions: read-only (control), self-explanation, sketching, sketching, then self-explanation or self-explanation, then sketching. All students were told to narrate everything they were thinking and doing aloud as they worked. For purposes of the study, we operationally define verbalizations as students narrating their thinking and behavior as they worked through the problems; all students were asked to verbalize, regardless of condition. This is distinguished from our operational definition of self-explanation. Students identified the essential parts of the problem and justified why their answers must be accurate; only students in the self-explanation conditions were explicitly asked to do so. If any student remained silent for ten or more seconds, they were told, "Don't forget to say what you're thinking" or "Don't forget to say anything you're thinking or doing." All students, regardless of condition, were given 10 minutes on each problem before being prompted to move on to the next problem.

In the read-only condition, students were told that they would not receive any assistance on the problems except with the pronunciation of any words they found difficult. They were also reminded to say everything they were thinking and doing out loud and that they would receive reminders if they failed to do so. Finally, they were told that they had 10 minutes to complete each problem and did not have to write anything down unless they wanted to. Students were then shown the page setup and asked if they had any questions before starting the problem solving on their own.

In the self-explanation condition, students were told that a great way to solve the problem was to name the essential parts of the problem and explain how the parts work together. They were then told that once they had an answer, they should explain why their answer must be true. As an example, the experimenter then presented the student with a video of a student reading a sample problem about a farmer needing to get from one side of town to another while traveling along a road with an obstruction, prompting them to calculate the total number of miles the farmer would be required to travel, and then self-explaining the problem while saying everything they were thinking and doing. At the end of watching the video, students were asked, “So does explaining your answer make sense?”. Explaining these directions took approximately five minutes. The experimenter then recited the same directions in the read-only condition. If the student hadn’t self-explained after four minutes or implied that they were finished without self-explaining, the experimenter said, “Don’t forget to explain how the parts of the problem work together and why your answer must be true.” This was done while gesturing to a board with written instructions for that condition.

In the sketching condition, students were told that a great way to solve the problem was to draw a diagram that included the essential parts of the problem and how they relate to each other. They were then briefly presented with diagrams or drawings like what they might have seen in science textbooks, including the digestive system, the layers of skin, and functions of hair. The experimenter then presented the student with a video of a student reading a sample problem and then working through it by drawing while saying everything they were thinking and doing. At the end of watching the video, students were asked, “So does drawing your answer make sense?” Explaining these directions took approximately five minutes. The experimenter then recited the same instructions repeated in the read-only condition. Then, if the student hadn’t sketched after four minutes or implied that they were finished without sketching, the experimenter said, “Don’t forget to draw your diagram,” while gesturing to a small board with written instructions for that condition.

In the self-explanation and then sketching condition, students were first given the same directions as students in the self-explanation condition, followed by the additional directions from the sketching condition. In this condition, explaining the directions took approximately 10 minutes. Finally, in the sketching then self-explanation condition, the instructions were the same as in the self-explanation and sketching condition but repeated in the reverse order. Time spent on instruction was not matched given the nature of the instructions for sketching and self-explaining. Regardless of condition, children were given a maximum of 10 minutes to complete each problem, totaling 40 minutes across the four problems. Most frequently, children included in the “both” conditions required a minimum of two but no more than three sessions to complete the problems.

Performance

Students’ performance scores were created independent of coding elements and relations in the student-created sketches and self-explanations. Performance scores were scored by a different set of research assistants and based on the correctness of the answers the students gave to the problems. This allowed us to test the effects on performance distinctly from the effects of elements and relationship strategies. Students’ verbal answers for each stimulus item were transcribed and later rated as 0 (incorrect), 1 (partially correct), or 2 (correct). Each problem's scores were averaged across all problems the students completed to create our outcome measure. Research assistants who had not been a part of the sketch and self-explanation coding process rated students’ answers. The ratings are considered reliable, with intra-class correlations for each stimulus greater than .80. Our scoring rubric is demonstrated in Table 1.

Miller-Cotto Dana (Orcid ID: 0000-0003-0577-4868)

Table 1

Coding rubric of student responses to science stimuli

Stimulus/Question	0 – incorrect	1 – partially correct	2 - correct
Blood Cells			
Which cells would be involved in healing?	Student does not name any of the cells required for a correct answer or only names incorrect types as being involved	Student names only one or two of the types of cells specified for a correct answer, the student may also incorrectly name a different type of cell as being involved	Student names all three parts as Killer T cells, macrophages, and platelets
Eye			
What is the normal pathway of light through the eye?	The student does not address this part of the question or only names steps irrelevant to the pathway	The student names some portions of the pathway not but all or incorrectly specifies the steps of the pathway	Student specifies the pathway of light as cornea → pupil → lens → retina
How would vision be affected if, as people age, the lens gets less flexible?	Student does not address this part of the question or provides an incorrect/irrelevant answer	The student ties his/her response to the lens' inability to change shape but fails to implicate the role/function of the muscles attached to the lens	The student indicates that if the lens is less flexible then the muscles cannot as easily change the shape of the lens leading to poor vision
Heart			
What is the normal pathway of blood through the heart?	The student does not address this part of the question or only names steps irrelevant to the pathway	The student names some portions of the pathway not but all or incorrectly specifies the steps of the pathway	The student identifies the pathway correctly as lungs → LA → LV → Body → RA→ RV→ (back to lungs)
How would the body be affected if the pathway from the RA to the RV were locked?	Student does not address this part of the question or provides an incorrect/irrelevant answer	The student ties his/her response to the body's inability to pick up oxygen but fails to specify that it is because the blood won't be able to go RV → lungs	The student indicates that blood would not be able to pick up oxygen from the lungs
Skin			
Which parts would be involved in healing and which wouldn't	Student does not name any of the parts required for a correct answer or only names incorrect parts as being involved	Student names only names one or two of the possible elements required for a correct answer, the student may also incorrectly name a different part of the skin as being involved	Student names two parts and being involved – light touch receptors and bare nerve endings; students name deep pressure receptors as not being involved

Miller-Cotto Dana (Orcid ID: 0000-0003-0577-4868)

Coding Verbalizations

The coding scheme was developed to determine whether students could correctly identify the critical elements, their functions, and the relations between the elements in a problem to solve it (Hmelo-Silver & Pfeffer, 2004). We initially adapted Cromley and colleagues' reading-text and reading-diagrams coding scheme (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010). Eventually, we collapsed across codes to create a better parallel between self-explanations and sketch codes for comparison. Our final variables are shown in Table 2.

Table 2
Code, descriptions, and examples of the variables used to code think-aloud protocols

Code	Definition	Example
Element Invocation	Identifying or considering an element or elements without going beyond to consider relationships between elements	“Okay, right atrium, left atrium”
Relationship Invocation	Identifying or describing a connection between elements; identifying/describing an element's function; this code presumes element invocation	“The right atrium gets blood from the body and the left atrium gets blood from the lungs”

One of the authors coded all participant verbalizations specifically for any existing elements or relations. A second coder coded a third of the participant verbalizations used to calculate Cohen’s Kappa. We summed the number of elements and relationships identified by each coder for each problem and ran inter-rater reliability at the item level (e.g., each row in the database was one problem for one student and included the number of elements identified by the first rater, the number of relations identified by the first rater, the number of elements identified by the second rather, and the number of relations identified by the second-rater). The sums of participants' verbalizations by science domain and for each representation type (i.e., elements or relations) were averaged across all problems and divided by total verbalizations to control for students' verbosity or shyness. Inter-rater reliability was computed across the four problems indicating good reliability (k = .89). The resulting codes ranged from 1 to 7 (M = .73 to 4.07; SD = 1.00 to 4.88). Skewness ranged from .71 to 3.44 (SE = .06 to .07, with kurtosis ranging from -.29 to 21. Higher scores meant more elements or relations.

Distinguishing Verbalizations and Self-Explanations

To differentiate between a verbalization and a self-explanation, an example of a verbalization was "Platelets can be a circle here, and they send signals, so I draw an arrow here." An example of a self-explanation was, "When you get a cut from scraping your knee on a dirty playground, macrophages would be used because macrophages squirt out enzymes that would clean the scrape because they destroy infected body cells" (see more examples in Table 3).

Table 3

Distinction between Verbalizations and Self-Explanations

Problem type	Verbalization	Self-explanation
Blood cells	"Platelets can be a circle here, and they send signals, so I draw an arrow here".	"My answer is true because when you get a cut from scraping your knee on a dirty playground, macrophages would be used because macrophages squirt out enzymes that would clean the scape because they destroy infected body cells".

Eye	"To figure out the path of light through the eye I will draw some arrows through these circles and label each".	"The path of light through the eye first goes through cornea, then the pupil, then the lens and vision would be affected because the muscles aren't as flexible so you can't see as good".
Heart	"I'm going to draw a long line and then with little labels for each part of the heart to see the path".	"My answer must be true because first the blood goes from left atrium to the left ventricle, then goes to the body, then right atrium pumps to right ventricle and then picks up more oxygen. So if it's blocked you can't get oxygen".
Skin	"First, I'll list the receptors and what they do. So the first is skinny and spindly ones. Those are for pain and temperature".	"If someone were to pet a bunny softly, since the bunny is soft, let's see, we would need oval shaped light touch receptors. We wouldn't need ones for pain or temperature unless the bunny had a fever and bit you. We wouldn't need deep pressure ones unless it hugged you".

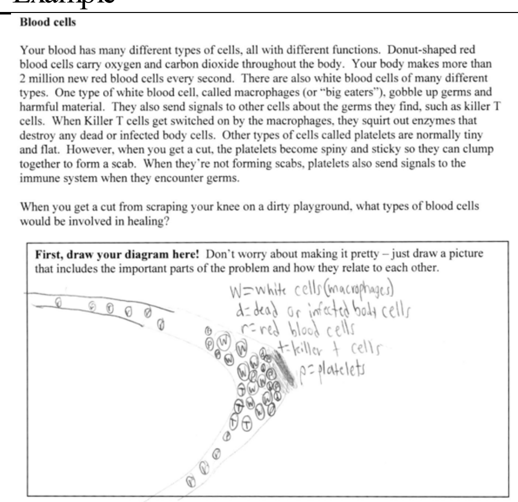
Coding Sketches

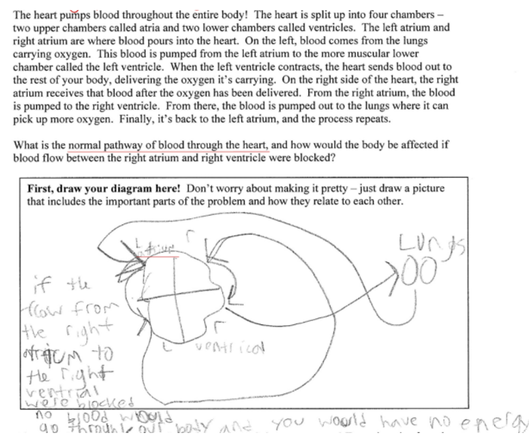
Scoring students' sketches first required that the problems be broken down into their essential elements. Problems were evaluated by a team of two science faculty advisors and two graduate students serving on the project with science education backgrounds. For each problem, the team members were asked to write out all their work to answer the question, and if they were hypothetically solving it, what would they be looking for to demonstrate comprehension. Their solution steps were broken down into each element and the necessary and sufficient relationships between elements.

Relevant elements and relationships were compiled as a coding rubric for each problem and used to score each participant's sketch. Initial coding for each sketch included processes (e.g., muscles adjust the shape of the lens) or relationships of the elements (e.g., rods and cones sit on the retina), whether each was represented in the sketch (e.g., rods and cones drawn with the retina), whether it was correctly identified, the number of irrelevant items, and the number of inaccuracies. Processes/relationships were collapsed across codes and re-coded as relationships to create better parallels between groups. Correctly identifying elements were re-coded as *elements*. To explore the relationship between sketches and successfully solve the problem, each element and relationship was scored by the type of sketch code presented in Table 4 below. Thus, texts that indicate processes between elements were coded as *relationships*, whereas labels indicating a word from the problem were coded as *elements*. Non-decorative and decorative codes were excluded from coding and subsequent analyses since they rarely contributed to correctly solving a problem.

Table 4

Sketch Coding Scheme

Problem Type	Type	Example
Blood Cells/Low Spatial	Elements and Relationships	 **Blood cells** Your blood has many different types of cells, all with different functions. Donut-shaped red blood cells carry oxygen and carbon dioxide throughout the body. Your body makes more than 2 million new red blood cells every second. There are also white blood cells of many different types. One type of white blood cell, called macrophages (or "big eaters"), gobble up germs and harmful material. They also send signals to other cells about the germs they find, such as killer T cells. When Killer T cells get switched on by the macrophages, they squirt out enzymes that destroy any dead or infected body cells. Other types of cells called platelets are normally tiny and flat. However, when you get a cut, the platelets become spiny and sticky so they can clump together to form a scab. When they're not forming scabs, platelets also send signals to the immune system when they encounter germs. When you get a cut from scraping your knee on a dirty playground, what types of blood cells would be involved in healing? **First, draw your diagram here!** Don't worry about making it pretty – just draw a picture that includes the important parts of the problem and how they relate to each other.

Heart Path/High Spatial	Elements and Relationships	

The resulting codes were sketch elements and relationships distinct from but parallel to self-explanation and verbalization relationship codes. A research assistant coded all participant sketches specifically for any existing elements or relations, and a second coder coded a third of the participant verbalizations used for calculating Cohen's Kappa. Inter-rater reliability was scored across the four problems indicating good reliability (k = .93). Inclusion of sketched elements or sketched relations ranged from 1 to 33 (M = 5.89 to 11.3; SD = 6.14 to 10.7). Skewness ranged from .38 to 1.02 (SE = .16 to .28, with kurtosis ranging from -.1.1 to .31. Higher scores suggested higher inclusion of elements or relationships between elements.

A Priori and Posthoc Analyses

A priori and post hoc power analyses were conducted to determine the minimum sample size to detect a significant effect of sketching, self-explanation, both, or neither and their mechanism on learning, α = .05. Before conducting the study, we conducted a power analysis based on Cohen's d = .50 as a standard for educationally relevant findings and power > .80 as is conventional in the social sciences. Our analysis suggested a minimum n = 20 per group (Cohen, 1977). To detect correlations $\geq$ .2 as reported in the literature, a sample of 193 was required. Post hoc analyses revealed that the observed power to detect a medium effect for the nine coefficients included in the multiple regression analyses for our first research question was high at .99.

Miller-Cotto Dana (Orcid ID: 0000-0003-0577-4868)

Table 5

Descriptive statistics of prior knowledge, sketch and think-aloud codes, and performance on paper measures.

	Prior Knowledge		Performance		Sketch Codes				Think Aloud Codes			
	PK Math	PK Sci	Math	Sci	Math SK-E	Math SK-R	Sci SK-E	Sci SK-R	Math V-E	Math V-R	Sci V-E	Sci V-R
M	12.42	15.15	0.66	0.74	11.26	8.64	11.32	5.89	0.93	1.29	0.73	4.06
SD	4.06	2.81	0.46	0.45	8.69	6.78	10.72	6.15	0.95	1.43	1	4.89
N	207	194	139	195	171	171	171	171	141	141	196	1.96

Note: PK = Prior knowledge, SK – E = Sketching Elements, SK-R = Sketching Relationships, V – E= Verbalization Elements, V-R= Verbalization Relationships

Miller-Cotto Dana (Orcid ID: 0000-0003-0577-4868)

Results

Data Analysis

We employed a linear mixed-effects (LME) model (conducted in R Studio using the lme4 and lmer.Test packages). First, LMEs do not require complete data from each participant as long as the restricted maximum likelihood estimator is used. Second, LMEs do not require normally distributed variables. Third, LMEs allow for the use of both fixed and random effects, taking into account the nested structure of the data. The data set included two levels of hierarchical data: individual items (Level 1) nested within students (Level 2). Level 1 included the accuracy score and sketch and verbalization codes for each item and whether the item was considered highly spatial or low-spatial. Level 2 included the condition and prior knowledge measures. To determine the amount of variance accounted for by Level 2 (students), the intra-class correlation (ICC) was calculated. A one-way ANOVA on prior knowledge scores showed no significant differences by condition at pretest (F [3, 195] = 1.804, p = .148, MSE = 12.951, η^2 = .027), so we considered the groups equivalent. Descriptives can be found in Table 5, with bivariate correlations in the Appendix.

Fidelity of Treatment

We wanted to determine whether students followed the instructions for specific conditions (i.e., self-explained or sketched when they were asked) and determine how much they engaged in self-explanation or sketching spontaneously (i.e., when their instructions did not ask them to do so). Self-explanations were defined as causal inferences verbally made by the student, consistent with prior the self-explanation literature (McNamara, 2004). Sketches included diagrammatic representations, labels (including text or numbers) in a diagram, and symbolic and non-symbolic representations (see Table 2).

Students often self-explained even when not requested: 60% of verbalizations in the read-only condition and 68% in the sketching condition included at least one self-explanation. However, levels were higher when students were instructed to self-explain: 80% of verbalizations in the self-explanation condition and 75% in the combination condition included at least one self-explanation. Spontaneous sketching was less common: 33% of students' work in the read-only condition and 17% of students' work in the self-explanation condition included at least one sketch. However, students sketched when asked: 89% of students' work in the sketch condition and 84% in the combination condition included a sketch. While trending in the expected direction, the difference in self-explanation frequency across conditions was not significantly different.

To examine these differences statistically, we created two variables. Students in the read-only and Self-Explanation conditions were coded a "0" for the Sketch variable, while students in the Sketch and combination conditions were coded a "1". For the Self-Explanation variable, students in the read-only and Sketch conditions were coded a "0". At the same time, students in the Self-Explanation and combination conditions were coded a "1". We then conducted a pair of independent-samples t-tests to determine if students in Sketch conditions sketched more than students in the non-sketching conditions and whether students in the Self-Explanation (SE) conditions self-explained more than students in non-SE conditions. These analyses revealed that sketching groups engaged in more sketching than non-sketching groups ($t(197) = -11.74$, $p < .001$, $d = 1.65$). There was also a trend towards SE groups engaging in more self-explanation than non-SE groups ($t(197) = -1.71$, $p = .089$, $d = 0.24$). One-sample t-tests comparing levels to a baseline of zero revealed that non-sketching groups also engaged in a significant amount of sketching ($t(87)= 5.87$, $p < .001$, $d = 0.63$), and non-SE groups also engaged in a significant amount of self-explanation ($t(98)=18.21$, $p < .001$), $d = 1.83$.

As described previously, the fine-grained qualitative codes of students' sketches and verbalizations were collapsed into two categories: elements and relationships. Because sketching and self-explanation occurred both in explicitly instructed groups and those that weren't, we included the variables measuring elements and relationships for all conditions.

Factors Influencing the Accuracy of Problem-Solving

RQ1: How do sketching, self-explanation, spatiality, and students' prior knowledge impact students' problem-solving accuracy?

To determine which factors significantly impacted students' problem solving, we conducted a single LME on each item's accuracy scores. At Level 2 (student), this model included fixed effects for Sketch, Self-Explanation, the interaction between Sketch and Self-Explanation, and prior knowledge. At Level 1 (item), we included fixed effects of the individual problem's high- vs. low-spatial nature and the number of elements and relationships identified in their sketches and verbalizations. The intraclass correlation (ICC) indicated that 18.1% of the variance occurred at Level 2 (student).

Table 6

Predictors of Problem-Solving Accuracy

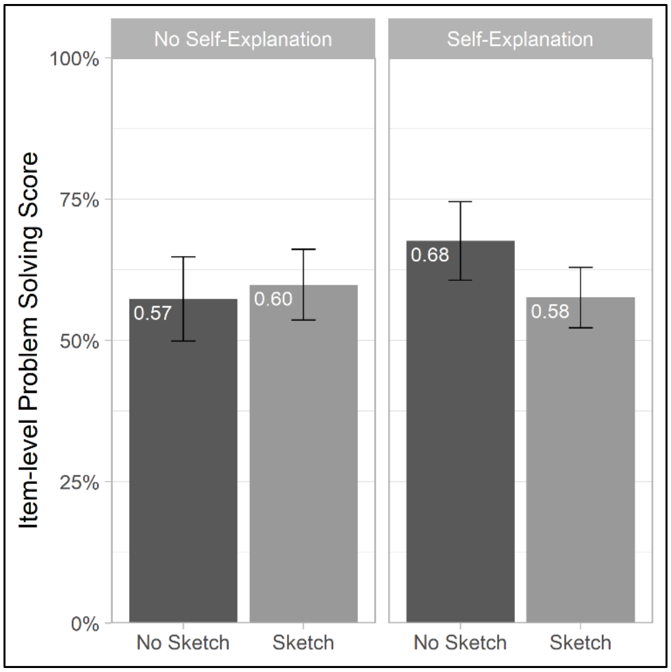
	B	*β*	*SE*
Sketch	.025	.07	.095
Self-Explanation	.103*	.29	.093
Sketch x Self-Explanation	-.125*	-.35	.121
Prior Knowledge	.428***	.17	.009
High/Low Spatial	-.350***	-.97	.045
Verbalized Elements	.034	.04	.028
Verbalized Relationships	.020*	.07	.010
Sketched Elements	.021***	.21	.004
Sketched Relationships	.017**	.11	.006

Variance Components	
Student-Level	.0686
Item-Level	.2990
*$p < .05$; **$p < .01$; ***$p < .001$	

As shown in Table 5, the model revealed a significant effect of Self-Explanation, indicating that students who were asked to self-explain had higher accuracy scores than those not in a self-explain group. There was also a significant interaction between sketching and self-explanation (see Figure 2). Follow-up pairwise tests indicated that when students were not asked to self-explain, sketching had no impact on their accuracy, $t(176.98) = 0.53$, $p = .598$. However, when students were asked to self-explain, also being asked to sketch was harmful to their accuracy, $t(211.99) = -2.31$, $p = .028$. At the same time, as long as students were not asked to sketch, self-explaining positively impacted their accuracy, $t(159.57) = 2.21$, $p = .028$. However, when students were asked to sketch, also being asked to self-explain did not impact their accuracy, $t(160.30) = -0.58$, $p = .566$.

Figure 2

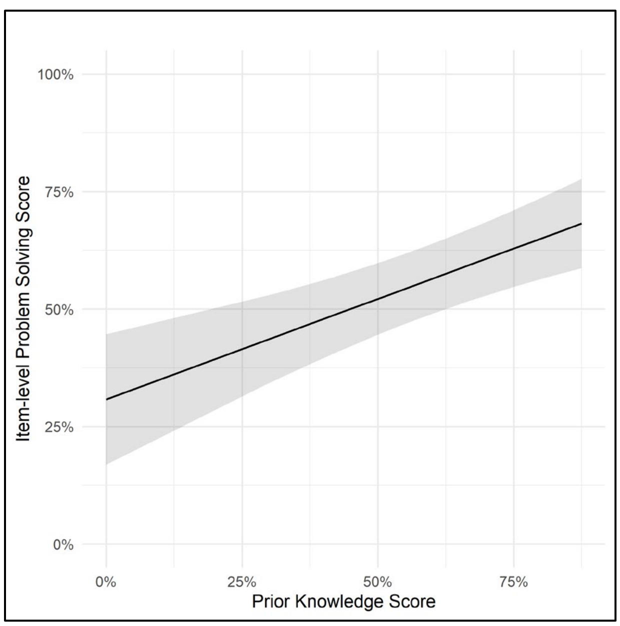
Interaction between self-explanation and sketching on average item-level problem-solving scores.



There was also a significant effect of prior knowledge, such that students with higher prior knowledge earned higher accuracy scores (See Figure 3).

Figure 3

Main effect of prior knowledge scores on average item-level problem-solving scores.



There was a significant effect of spatialization of the problem at the item level. Students scored higher on low-spatial items than high-spatial ones. In addition, there were significant effects on the number of relationships represented in verbalizations, the number of elements represented in sketches, and the number of relationships represented in sketches. In all cases, showing more predicted higher accuracy scores. Given the importance of these components for student performance, we wanted to examine further how conditions impacted what types of features students said or drew. This analysis would allow us to understand the mechanisms underlying the impacts of sketching and or self-explanation on problem-solving accuracy.

Factors Influencing Representation of Elements or Relationships

RQ2: How do sketching, self-explanation, spatiality, and students' prior knowledge impact students' inclusion of elements and relationships in their sketches or verbalizations?

To determine which factors significantly impacted students' inclusion of elements or relationships in their verbalizations or sketches, we conducted four LMEs on verbalized elements, verbalized relationships, sketched elements, and sketched relationships for each item. At Level 2 (student), this model included fixed effects for Sketch, Self-Explanation, the interaction between Sketch and Self-Explanation, and prior knowledge. At Level 2 (item), we included fixed effects of the individual problem's high- vs. low-spatial nature. Intraclass correlation (ICC) indicated that the percentage of the variance occurring at Level 2 (student) was 12% for verbalized elements, 58% for verbalized relationships, 37% for sketched elements, and 32% for sketched relationships.

Table 7
Predictors of Verbalization Codes

	Model 1: Verbalized Elements			*Model 2: Verbalized Relationships*		
	B	*β*	*SE*	*B*	*β*	*SE*
Sketch	-.092†	-.21	.053	-.218	-.15	.262

Self-Explanation	-.101†	-.23	.053	.051	.03	.263
Sketch x Self-Explanation	-.134†	.31	.072	.252	.17	.355
Prior Knowledge	-.140	-.05	.117	.513	.05	.577
High Spatial	-.038	-.09	.029	.016*	.11	.067
Variance Components						
Student-Level		.0208			1.2683	
Item-Level		.1627			.8873	

†$p < .10$; *$p < .05$;**$p < .01$; ***$p < .001$

As shown in Table 6, there was only one significant predictor of verbalization codes: spatialization of the problem on verbalized relationships. Students included more relationships in their verbalizations for high-spatial problems. For verbalized elements, there were non-significant trends for sketching, self-explanation, and the interaction between sketching and self-explanation, essentially showing that students who were not asked to self-explain or to sketch included more elements in their verbalizations.

Table 8

Predictors of Sketch Codes

	Model 1: Sketched Elements			*Model 2: Sketched Relationships*		
	B	*β*	*SE*	*B*	*β*	*SE*
Sketch	2.898***	.83	.432	1.574***	.70	.314
Self-Explanation	-0.583	-.17	.430	-0.184	-.08	.314
Sketch x Self-Explanation	1.185*	.34	.565	0.445	.20	.411
Prior Knowledge	1.343	.06	.962	0.155	.01	.701
High Spatial	1.433***	.41	.201	1.105***	.49	.134
Variance Components						
Student-Level		1.464			.9182	
Item-Level		6.906			3.101	

†$p < .10$; *$p < .05$;**$p < .01$; ***$p < .001$

As shown in Table 7, being in a sketch condition was a significant predictor of sketched elements and relationships; students who were asked to sketch included more elements and more relationships in their sketches. There was also a significant interaction between sketching and self-explanation for sketched relationships, indicating that students who were asked to both sketch and self-explain included the most relationships in their sketches. The spatialization of the problem impacted both sketched elements and relationships; students had more elements and more relationships in their sketches for high-spatial problems.

Discussion

Self-explanation and sketching are regarded as two generative problem-solving techniques with positive impacts on learning. However, there has been little investigation into the particular mechanisms by which each approach benefits learning and any unique contributions of one approach. The present study directly compared sketching vs. self-explanation strategies during problem-solving for middle school students' science problems that differed in spatial content (more vs. less spatial). Students asked to self-explain demonstrated higher accuracy in solving science problems than those not in a self-explanation group (i.e., sketching or read-only). Further, when students were not prompted to self-explain, sketching did not affect their science accuracy, though when they were prompted to self-explain, also being asked to sketch appeared to hurt their performance.

Sketching and Self-Explanation

Because both sketching and self-explanation are established strategies for improving problem-solving performance (e.g., Rittle-Johnson & Loehr, 2017; Schmeck et al., 2014) that invoke generative processes of selecting information, organizing it, and integrating it with prior knowledge (Fiorella & Mayer, 2015), we expected that students in both the self-explaining and the sketching conditions would out-perform students in the read-only (control) condition, our findings for self-explanations are consistent with generativity theory (Wittrock, 1989). That is, learners need to make connections between new information and existing knowledge, and self-explanations encourage students to identify the most important components of material they're learning, or in our case, elements and relationships, summarize how this information is related, and organize the information into a coherent model (Fiorella & Mayer, 2015). While sketching can also activate prior knowledge, the prompts used herein did not specifically prompt prior knowledge activation. Further, prior work demonstrates the difficulty of benefiting from sketching when engaging with information that includes complex systems such as anatomy education as students engaged here (Ainsworth & Scheiter, 2021). Future research might consider the complexity of information students are asked to engage with and what domains benefit most from sketching. When prompting students to sketch, prompts should also include reminders on how to activate prior knowledge when sketching.

Specific to performance, our results demonstrated lowered accuracy when sketching was requested in combination with self-explanation. One possibility is that asking students to engage in sketching and self-explanation increased their cognitive demands. Asking students to remember to self-explain *and* sketch while verbalizing everything they were thinking and doing likely interfered with their propensity to make connections between elements and relationships that helped them solve the problem. Verbalizing to self-explain likely felt more natural to them because they were already engaging in the process of explaining what they were doing, so articulating the reason behind their answers likely didn't require much more effort, whereas having to sketch on top of that may have increased the task difficulty need to learn from it. It is also possible that there might have been what is termed a *redundancy effect*, meaning that when there is more than the essential visuospatial information provided to learn, students must process both essential and non-essential information, reducing resources available for learning (Kalyuga

& Sweller, 2014; Castro-Alonso, Ayres, & Sweller, 2019). This is evidenced in Table 5, where the combination of sketching and self-explanation was a negative predictor of accuracy scores.

These contrasting patterns for self-explanation and sketching may suggest that students in the current study were already familiar with self-explanations and how to use them effectively to solve science problems and simply needed a reminder to engage attention, whereas sketching may have been a novel strategy that they had no training on using competently. This may be one possible explanation for why sketching alone did not boost performance overall and imposed a cost when combined with self-explanation. For sketching to be an effective strategy, instructions should specify the necessary information that students should include to benefit from this strategy (Ainsworth & Scheiter, 2021). Future work should establish *what* students should be sketching when prompted to sketch to gain the most benefit. Perhaps using strategies from the concept mapping literature (Horton, 1993), students should be encouraged to demonstrate key concepts and relations between concepts for sketching to be most beneficial.

Student Inclusion of Elements and Relationships

Interestingly, students who were instructed to self-explain were *not* more (or less) likely to include references to elements or relationships in their verbalizations than the students who were not instructed to self-explain. Encouragement to self-explain aided accuracy without significantly changing either the amount or nature of what students verbalized, with high base rates of self-explanation in the read-only group. On the other hand, sketching was at low base rates and was significantly more common with an instruction to sketch. Although sketching did not promote greater accuracy, it did lead to representing elements and relationships; students who were prompted to sketch and self-explain included the most connections in their sketches. Spatialization of the problem also mattered because students had more elements and relationships in high spatial problems. In addition, variation in sketch quality positively correlated with accuracy but, crucially, not with prior knowledge.

Table 9
Overview of Contrasting Findings

	Self-Explanation	Sketching
Base rates of approach use without prompting	High	Low
Change in frequency of approach use with prompting/instruction	Minimal	Large
Influence of prompting for approach on problem-solving accuracy	Yes	No
Influence of prompting for approach on quality of verbalizations	Less likely to verbalize elements; no change in verbalization of relationships	Less likely to verbalize elements; no change in verbalization of relationships
Influence of prompting for approach on quality of sketches	No change in sketching of elements or relationships	More likely to sketch elements and relationships
Influence of included content on problem-solving accuracy	Inclusion of relationships in verbalizations predicts greater accuracy	Inclusion of both elements and relationships in sketches predicts greater accuracy

Impact of Students' Prior Knowledge

Intriguingly, prior knowledge was not significant predictor of how often students used elements and relationships in verbalizations or sketches. Indeed, the only significant predictor of elements and relationships was the spatialization of a problem on verbalized relationships. Prior knowledge is frequently a significant predictor of later performance, as we observed when analyzing science accuracy, but this was not the case for the use of verbalization and articulating elements and relationships. However, it is essential to note that the prior knowledge measure we used was designed to tap broader knowledge of the underlying science concepts in the problem rather than specific knowledge of the elements and relationships in the problem. Therefore, it is possible that a different measure of prior knowledge which targeted these specific knowledge components would predict the degree to which students invoke specific elements and relationships in their sketches or verbalizations.

Strengths and Limitations

Among the strengths of this study were the process data that we collected in addition to the experimental design. The comparison between sketching, self-explanation, both, and neither allowed us to systematically compare if there were additive benefits in using both strategies to solve problems. Further, verbalization and sketch codes illuminated what mechanisms make sketching and self-explanation compelling. Balancing low spatial problems with high spatial problems gives us further insight into whether elements and relations elicit greater problem-solving accuracy and demonstrate critical components and connections. These features give us greater confidence that while sketching and self-explanation strategies have demonstrated effectiveness in prior studies (e.g., Scheiter et al., 2017), it is now clearer *what* features, such as elements and relations, make them effective.

Of course, the current study is not without limitations. One limitation is that the role of verbalization in the present study is unclear. Though students were told to do this across all conditions and thinking aloud typically does not influence performance (Fox, Ericcson, & Best, 2011), it is possible that verbalizing had differential impacts on problem-solving performance when students were sketching vs. explaining vs. just reading and solving the problem; the previously described scenarios in which verbalization and self-explanation are considered interchangeable are a crucial example. Capturing the comparative effectiveness of sketching and self-explanation may thus require having students engage in those processes without verbalization. In the present study, we included verbalization because we thought it would be necessary to analyze what students said while they were solving problems; however, it is helpful to note that what students *said* throughout their problem-solving process was not as important as what they *did*, as evidenced by the sketches that included elements and relationships being related to higher science problem-solving accuracy.

It is also worth noting that levels of self-explanation in conditions where students were not explicitly told to self-explain were still high, though not as high in the self-explanation conditions nor significantly different. There are a few reasons this could be the case. Encouraging students to justify their answers, like self-explanations, is a commonly used tool in classrooms; teachers often encourage students to justify their answers. So, students, excellent problem solvers, may be used to doing this even when unprompted (Renkl & Atkinson, 2010). This is unlikely to be the case for sketching, where there is no established professional development or materials that encourage this strategy.

Finally, the equivalence between high and low spatial tasks since the topics given throughout the tasks was different could not be established. Further, the difficulty and students' familiarity with these topics may have contributed to the findings. Of course, some issues in science are inherently more spatial than others, so this confound is difficult to circumvent. Future research should estimate the effects of prior knowledge more precisely than what was done here and complexity to explicate these effects further. Further, we directed students to discuss the spatial information (i.e., "describe the path") and asked them an additional application question. In the low spatial problems, we asked students to identify what elements are or are not relevant in a particular scenario. Asking a "how" question versus a "name the elements" question is a crucial confound to point out, and we acknowledge how this may alter our results.

A potential future direction could be looking at individual differences in spatial skills. Jaeger et al. (2018) show that spatial ability significantly predicted drawing quality. Thus, it's possible that sketching is only helpful for students with strong spatial ability. Students with strong spatial ability may be better at producing high-quality sketches that they can learn from. This awaits further confirmation.

Conclusion and Implications

Perhaps the most important lesson to be learned from the current study is that when determining the best generative process for problem-solving, some functions may require specific instructions to produce the most significant benefit. Students may need support on the best use of sketching and encouragement to do so spontaneously. While self-explaining in the current study led to greater science accuracy, sketching also helped students produce important features in external representations to help with understanding the problem. Quality of sketches, or including elements and relationships, may clue teachers in on students' knowledge and alert them to where they may need to intervene. Strategies that help students make their knowledge explicit and connect various problem features can help them solve problems they are most unfamiliar with.

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, *333*(6046), 1096-1097.

Ainsworth, S. E., & Scheiter, K. (2021) Learning by Drawing Visual Representations: Potential, Purposes, and Practical Implications. *Current Directions in Psychological Science*, 1-7.

Atkinson, R. K., & Renkl, A. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective.

Educational Psychologist, 38, 15-22.

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior*, *27*(1), 69-75.

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications*, *1*(1), 27.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. In *Visuospatial Processing for Education in Health and Natural Sciences* (pp. 111-143). Springer, Cham.

Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in instructional psychology*, *5*, 161-238.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences*, *14*(2), 161-199.

Chi, M. T. (2009). Active-constructive- interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, *35*(1), 59-74.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (revised ed.).

Dumas, D. (2017). Relational reasoning in science, medicine, and engineering. *Educational Psychology Review*, *29*(1), 73-95.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Edens, K., & Potter, E. (2008). How students "Unpack" the structure of a word problem: Graphic representations and problem solving. *School Science and Mathematics, 108*(5), 184-196.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity*. Cambridge University Press.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological bulletin*, *137*(2), 316.

Gagnier, K. M., Atit, K., Ormand, C. J., Shipley, T. F. (2016). Comprehending 3D diagrams: Sketching to support spatial reasoning. *Topics in Cognitive Science, 10,* 1-19.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155-170.

Gobert, J.D. & Clement, J.C. (1999): Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, *36*(1), 39-53.

Heckler, A. F. (2010). Some consequences of prompting novice physics students to construct force diagrams. *International Journal of Science Education*, *32*(14), 1829-1851.

Hegarty, M., & Kozhevnikov, M. (1999). Spatial abilities, working memory, and mechanical reasoning. In J. S. Gero & B. Tversky (Ed.), *Visual and spatial reasoning in design* (pp. 221-241). Cambridge, MA, June, 15-17.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*(1), 127-138.

Horton, P. B., McConney, A. A., Gallo, M., Woods, A. L., Senn, G. J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education*.

Ionas, I. G., Cernusca, D., & Collier, H. L. (2012). Prior Knowledge Influence on Self-Explanation Effectiveness When Solving Problems: An Exploratory Study in Science Learning. *International Journal of Teaching and Learning in Higher Education*, *24*(3), 349-358.

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction*, *21*(3), 379-393.

Jaeger, A. J., Velazquez, M. N., Dawdanow, A., & Shipley, T. F. (2018). Sketching and summarizing to reduce memory for seductive details in science text. *Journal of Educational Psychology*, *110*(7), 899.

Jee, B.D, Gentner, D., Forbus, K., Sageman, B., & Uttal, D. (2009). Drawing on experience: Use of sketching to evaluate knowledge of spatial scientific concepts. *Proceedings of the 31st Annual Conference of the Cognitive Science Society,* Amsterdam, The Netherlands.

Kalyuga, S., & Sweller, J. (2014). 10 The Redundancy Principle in Multimedia Learning. *The Cambridge Handbook of Mmultimedia Learning*, 247.

Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial?. *Journal of Experimental Child Psychology*, *103*(3), 386-394.

Kuo, E., Hallinen, N. R., & Conlin, L. D. (2017). When procedures discourage insight: epistemological consequences of prompting novice physics students to construct force diagrams. *International Journal of Science Education*, *39*(7), 814-839.

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction, 22*(1), 16-26.

Leopold, C., Sumfleth, E., & Leutner, D. (2013). Learning with summaries: Effects of representation mode and type of learning activity on comprehension and transfer. *Learning and Instruction, 27*, 40-49.

Lin, L., Lee, C. H., Kalyuga, S., Wang, Y., Guan, S., & Wu, H. (2017). The effect of learner-generated and imagination of comprehending a science text. *The Journal of Experimental Education, 85*(1), 142-154.

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology*, *104*(1), 1-21.

Madden, S. P., Jones, L. L., & Rahm, J. (2011). The role of multiple representations in the understanding of ideal gas problems. *Chemistry Education Research and Practice, 12*(3), 283-293.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1-30.

Moreno, R., & Mayer, R. E. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. Jan L. Plass, Roxana Moreno, Roland Brunken (Eds). Cognitive Load Theory, Cambridge: Cambridge University Press.

National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

Naug, H. L., Colson, N. J., & Donner, D. G. (2011). Promoting metacognition in first year anatomy laboratories using plasticine modeling and drawing activities: A pilot study of the "blank page" technique. *Anatomical Sciences Education*, *4*(4), 231-234.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research In Science Teaching*, *27*(10), 937-949.

Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72*(1), 39-60.

Parnafes, O., Aderet-German, T., & Ward, E. T. (2012). Drawing for understanding: an instructional approach for promoting learning and understanding. In *Annual Meeting of the American Educational Research Association, Vancouver, Canada* (pp. 1-36).

Peverly, S. T., Brobst, K. E, Graham, M., & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95(2)*, 335-346.

Rellensmann, J., Schukajlow, S., & Leopold, C. (2017). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educational Studies in Mathematics*, *95*(1), 53-78.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and instruction*, *12*(5), 529-556.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In *Cognitive load theory*. Cambridge University Press.

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic bulletin & review*, *24*(5), 1501-1510.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*(4), 599-611.

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. *The Cambridge handbook of multimedia learning*, 271-286.

Sachse, P., Hacker, W., & Leinert, S. (2004). External thought—does sketching assist problem analysis? *Applied Cognitive Psychology, 18*(4), 415-425.

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in Cognitive Science, 10,* 1-17.

Schmeck, A., Mayer, R., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology, 39*(4), 275-286.

Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text?. *Learning and Instruction*, *51*, 10-25.

Schmidgall, S. P., Eitel, A., & Scheiter, K. (2019). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, *60*, 138-153.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, *102*(4), 872.

Sweller, J. (2012). Human cognitive architecture: Why some instructional procedures work and others do not. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues* (pp. 295–325). American Psychological Association.

Uesaka, Y., & Manalo, E. (2012). Task-related factors that influence the spontaneous use of diagrams in math word problems. *Applied Cognitive Psychology*, *26*(2), 251-260.

Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice*, *18*(4), 246-254.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285-325.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142-166.

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, *11*(2), 87-95.