Learning and Instruction 55 (2018) 93–104

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Effectiveness and efficiency of adding drawing prompts to an interactive educational technology when learning with visual representations

Sally P.W. Wu*, Martina A. Rau

Department of Educational Psychology, University of Wisconsin-Madison, Madison, United States

ARTICLE INFO

Article history:
Received 8 September 2016
Received in revised form
22 September 2017
Accepted 29 September 2017
Available online 7 October 2017

Keywords:
Drawing
Visual representations
Sense making
Mental models
Chemistry

ABSTRACT

This paper investigates whether prompting students to draw their own visual representations enhances students' learning from technology-based instructional activities with visual representations. Seventy-two undergraduate students were randomly assigned to receive an educational technology with (1) drawing prompts throughout instruction, (2) drawing prompts before and after instruction, or (3) no drawing prompts. We assessed learning outcomes with respect to instructional effectiveness and efficiency using immediate and delayed posttests. Results on instructional efficiency showed a significant advantage for drawing prompts. Results on instructional effectiveness showed an advantage at the delayed posttest for drawing prompts provided throughout instruction, compared to prompts before and after. Qualitative analyses suggest that adding drawing prompts throughout instruction promotes drawing quality. In sum, our findings expand theory by suggesting that drawing prompts facilitate visual sense making of concepts shown in visual representations. Furthermore, we provide practical recommendations on how best to implement drawing prompts with technology-based instructional activities.

Many concepts in science, technology, engineering, and mathematics (STEM) are visual-spatial in nature. Therefore, students' learning of domain knowledge in STEM critically depends on their ability to make sense of *visual representations* (Gilbert, 2005; Mathewson, 1999). For example, students learn about atoms in chemistry with the visual representations shown in Fig. 1. These visual representations are typically used in instructional materials such as textbooks, worksheets, and webpages. We refer to these visual representations as *traditional* because they are designed and used by STEM professionals, not *generated* by students.

Prior research shows that students have tremendous difficulties in making sense of traditional representations (Ainsworth, 2006; Rau, 2016). Therefore, an important educational goal in STEM is to support students' learning with these representations. Much prior research has investigated how to design instructional activities that support students in *verbally* making sense of traditional representations (Rau, 2016; Rau & Wu, 2015a). For example, adding self-explanation prompts to instructional activities has been shown to be particularly effective (Berthold & Renkl, 2009; van der Meij & de Jong, 2011). Such prompts can ask students to self-explain while they construct, manipulate, and reason with representations (Rau, Aleven, & Rummel, 2015b). However, a new line of research suggests that visual-spatial concepts are difficult to explain verbally (Bobek & Tversky, 2014; Vosniadou, 1994). Instead, instructional activities that prompt students to engage in *visual* sense-making processes may be more effective in supporting students' learning with representations (Leopold & Leutner, 2012; Scheiter, Schleinschok, & Ainsworth, 2017). For example, *prompting students to draw* their own visual representations has been shown to be effective (Prain & Tytler, 2012; Van Meter & Garner, 2005). Drawing prompts can simply ask students to draw on paper and thus are easy to integrate within instructional activities that support verbal sense-making processes with traditional representations. Such prompts may be effective for two reasons. First, prompts to generate drawings can help students organize visual-spatial concepts from traditional representations and activate their own mental models (Brooks, 2009; Van Meter & Garner, 2005). Second, prompts to revise their drawings can help students revise their mental models after comparing their drawings to traditional representations (Prain & Tytler, 2012; Valanides, Efthymiou, & Angeli, 2013). Yet, prior research has not investigated whether providing

---

* Corresponding author.
E-mail address: pwwu@wisc.edu (S.P.W. Wu).

https://doi.org/10.1016/j.learninstruc.2017.09.010
0959-4752/

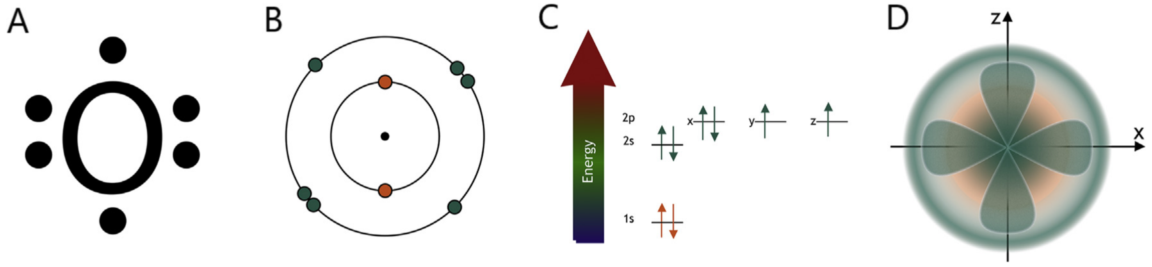


Fig. 1. Four traditional visual representations of an oxygen atom (from left): Lewis structure, Bohr model, energy diagram, and orbital diagram.

prompts to generate and revise drawings are effective when combined with typical instructional activities.

To this end, we present a controlled experiment that investigates whether adding prompts to generate and revise drawings to an educational technology enhances undergraduate students' learning of domain knowledge. We situate this experiment in undergraduate chemistry learning because success in chemistry requires learning with traditional representations and often involves drawing (Kozma & Russell, 2005; Talanquer, 2013).

1. Theoretical background

In the following, we first review prior research on how students learn with traditional visual representations and typical instructional activities that support *verbal* sense-making processes as well as recent findings on instructional activities that support *visual* sense-making processes. Then, we highlight gaps in prior research on visual sense-making supports, which we investigate in our experiment.

1.1. *Learning with traditional visual representations in STEM*

Prior research shows that students have difficulties in making sense of how visual representations depict domain-relevant concepts (Ainsworth, 2006; Rau, 2016). Students often focus on irrelevant surface features and fail to make connections among representations (Cook, Wiebe, & Carter, 2008; Kozma & Russell, 2005; Rau, Aleven, Rummel, & Pardos, 2014). For example, when students use Lewis structures and Bohr models (Fig. 1a and b) to learn about electrons in atoms, they may focus on irrelevant features such as color while failing to attend to relevant features such as the number and location of dots. Making such connections is particularly difficult for students with low spatial skills (Höffler, 2010) because it requires students to mentally rotate representations (Stieff, 2007).

A large body of research has investigated how best to help students overcome difficulties with visual representations. This research shows that effective instructional activities support students in *making sense* of how representations depict concepts (for an overview, see Ainsworth, 2006; Rau, 2016). Cognitive learning theories (Koedinger, Corbett, & Perfetti, 2012) suggest that instructional activities should engage *verbally mediated* sense-making processes, for instance self-explanation prompts.

1.1.1. *Self-explanation prompts that support verbal sense-making processes*

Self-explanation prompts have proven effective in helping students engage in sense-making processes (Roelle, Lehmkuhl, Beyer, & Berthold, 2015; Wylie & Chi, 2014). For instance, self-explanation prompts can ask students to explain how the spatial arrangement of electrons around the nucleus explains an atom's properties and bonding behavior. Research shows that such self-explanation prompts are especially effective when implemented in *educational technologies* that provide adaptive feedback on students' self-explanations (Rittle-Johnson, Loehr, & Durkin, 2017; Wylie & Chi, 2014). Self-explanation prompts with feedback can help students focus on relevant visual features shown in representations and connect features among multiple representations (Berthold & Renkl, 2009; Rau et al., 2015b).

Self-explanation prompts engage students in *verbally mediated* sense-making processes (Koedinger et al., 2012). Such processes involve verbal explanations of principles that describe how representations depict concepts (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). However, a new line of research suggests that verbal explanations may not adequately help students make sense of visual-spatial concepts shown in representations (Bobek & Tversky, 2014; Vosniadou, 1994). Specifically, studies show that self-explanation prompts that support verbal sense-making processes may be less effective than drawing prompts that support *visually mediated* sense-making processes (Leopold & Leutner, 2012; Scheiter et al., 2017).

1.1.2. *Drawing prompts that support visual sense-making processes*

Recent research shows that prompting students to draw their own representations is an effective means to support visual sense-making processes (Brooks, 2009; Van Meter & Firetto, 2013). Drawing prompts have been shown to enhance students' learning of domain knowledge in STEM (Leutner & Schmeck, 2014; Van Meter & Garner, 2005) by helping students learn how visual representations depict concepts (Prain & Tytler, 2012; Valanides et al., 2013). In addition, drawing has been shown to enhance long-term retention of concepts shown in visual representations (Mason, Lowe, & Tornatora, 2013).

How can drawing help students visually make sense of concepts? According to Van Meter's Cognitive Model of Drawing Construction (CMDC), students' generation of drawings involves three iterative phases (Van Meter & Firetto, 2013). In the first phase, students must understand the drawing task at hand. For instance, a prompt that instructs students to "draw what comes to mind" when they think of an atom will direct students to focus on their mental models, not traditional representations. Prior research suggests that students do not spontaneously draw (Leutner & Schmeck, 2014; Van Meter, Aleksic, Schwartz, & Garner, 2006). Therefore, simply providing paper and pens is insufficient. Students must receive *drawing prompts* to help them engage in visual sense-making processes, discussed in the following two phases.

In the second phase, students generate the drawing. To this end, students must identify, organize, and integrate relevant information about the to-be-learned concepts into a coherent mental model and then translate it into a visual representation. For example, to draw atoms, students first determine what concepts are relevant (e.g., nucleus and electrons), organize information about the atom by determining how different concepts relate to one another (e.g., spatial arrangement of electrons relative to the nucleus) and integrate it with their prior knowledge (e.g., electron repulsion). They then translate concepts using pen and paper, for instance by representing electron movement with arrows and lines. This process helps students internalize and integrate concepts at a deeper level (Brooks, 2009; Valanides et al., 2013). *Prompts to generate drawings* can facilitate this process. Indeed, generate prompts have been shown to enhance learning outcomes (Van Meter & Garner, 2005; Zhang & Linn, 2011), particularly if students generate high quality drawings in response to drawing prompts (Scheiter et al., 2017; Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner, 2014).

In the third phase, students monitor their progress by comparing and revising their drawings to align with domain-relevant concepts. For example, a student may notice that his drawing does not evenly space electrons around the nucleus and then revise the drawing based on electron repulsion. Because students externalize their mental models through generating drawings in the second phase, students can compare drawings with traditional representations to identify inaccuracies and inconsistencies within their mental models (Vosniadou, 1994). Through this process, students can reflect on what concepts they do not yet understand and revise their mental models (Valanides et al., 2013). *Prompts to revise drawings* can facilitate this process. Research has shown that revise prompts can enhance students' learning of domain-relevant concepts (Prain & Tytler, 2012; Valanides et al., 2013).

1.2. *Open questions about drawing prompts*

In the following, we highlight several open questions regarding how best to implement drawing prompts that support learning with traditional visual representations.

1.2.1. *Combining drawing prompts with traditional representations*

Most prior research on drawing prompts has focused on learning visual-spatial concepts described in text. Particularly, the effectiveness of *generate* prompts for learning from texts is well established (Leutner & Schmeck, 2014; Van Meter & Garner, 2005). However, only a few recent studies show that generate prompts can help students learn from traditional representations (Mason et al., 2013; Zhang & Linn, 2011).

The research on *revise* prompts has focused more extensively on learning with traditional representations. However, in these studies, revise prompts were often confounded with other instructional activities (Brooks, 2009; Valanides et al., 2013). For instance, revise prompts could be provided after students watch a video, role play the movement of atoms, and discuss observations with their classmates (Prain & Tytler, 2012), which makes it difficult to isolate the effects of revise prompts.

This prior research leaves several questions open. First, research on generating and revising drawings have mostly been separate lines of research. To the best of our knowledge, no prior study has systematically tested the effects of combining generate *and* revise prompts. Because both generate and revise prompts support phases of the CMDC, combining both prompts may be effective in helping students engage in visual sense-making processes.

Second, prior research has not investigated the nature of visual sense-making processes for traditional representations. CMDC suggests that drawing helps students visually make sense of concepts described in scientific text through a process of activating and reflecting on their mental models. Hence, we expect students to generate and revise drawings that increasingly align with traditional visual representations as students align their mental models with traditional representations. However, one might argue that repeated drawing prompts can reinforce inaccuracies in students' mental models or increase their tendency to focus on irrelevant concepts. Investigating changes in the quality of students' drawings can provide insight into how generate and revise prompts facilitate visual sense-making processes.

Third, research on drawings has not considered how drawing prompts complement typical support for verbal sense-making processes. To the best of our knowledge, there is only one study that systematically tested whether drawing prompts enhance learning from other instructional activities. Zhang and Linn (2011) found that students who were prompted to generate drawings were more likely to discuss domain-relevant concepts with peers and use traditional visual representations to make sense of concepts. These students generated more coherent drawings and showed higher learning outcomes, compared to students who only received extra time with traditional representations. Based on this finding, we hypothesize that drawing engages students in visual sense-making processes that make it easier to translate visual-spatial concepts into verbal explanations, thereby enhancing students' learning from instructional activities that engage verbal sense-making processes, such as self-explanation.

1.2.2. *Timing of drawing prompts*

Little research has investigated *how frequently* students should be prompted to draw. When implementing generate prompts, most prior studies provided them only once or twice, typically *before and/or after* instruction (Bobek & Tversky, 2014; Gadgil, Nokes-Malach, & Chi, 2012; Mason et al., 2013). A few studies prompted students to generate drawing *throughout* the entire instructional period (Leopold & Leutner, 2012; Schmeck et al., 2014; Van Meter et al., 2006). Studies on revise prompts have not examined how the frequency of revise prompts affects students' learning. However, correlational analyses within drawing studies show that students who draw more frequently show higher learning outcomes (Davis, 2000; Schank & Kozma, 2002). Therefore, we hypothesize that providing drawing prompts throughout instruction will yield higher learning outcomes than providing drawing prompts only before and after instruction.

Investigating the frequency of drawing prompts poses a methodological challenge for research that tests the effectiveness of drawing prompts. Some studies that compared drawing to no-drawing conditions provided additional instructional time for the drawing conditions to complete drawing activities (e.g., Schmeck et al., 2014). In such cases, higher learning outcomes in the drawing conditions may be attributed to longer time-on-task. Other studies controlled for time-on-task by comparing drawing conditions to other time-intensive instructional activities, such as writing verbal explanations (e.g., Bobek & Tversky, 2014; Scheiter et al., 2017). However, to the best of our knowledge, no prior study has controlled for time-on-task by *reducing* the number of instructional activities in drawing conditions. Investigating the effects of reducing other instructional activities will yield recommendations about whether instructors should incorporate drawing prompts when time is limited—a situation common in classrooms and other learning environments.

2. Research questions and hypotheses

To address the open questions from prior research, we conducted a controlled experiment that compared three conditions: (1) drawing prompts throughout instruction, (2) drawing prompts before and after instruction, and (3) a control condition without drawing prompts. We assess learning of domain knowledge with an immediate posttest and a delayed posttest given one week after instruction. We measure *instructional effectiveness* by learning gains and *instructional efficiency* by learning gains while accounting for time-on-task. We account for time-on-task in two ways. First, at the level of condition, we control for time-on-task by adjusting the number of instructional activities such that all conditions spent about the same amount of time on average across all activities (i.e., instructional activities and drawing). To this end, students who received drawing prompts were given fewer instructional activities to account for the time they would spend drawing. Second, at the individual level, we account for time-efficiency with instructional efficiency measures as proposed by van Gog and Paas (2008) because students worked on activities at their own pace. We consider drawing prompts to be *effective* if they enhance students' learning gains and *efficient* if they enhance students' learning gains in less instructional time.

Our experiment allows us to investigate the following questions:

Research Question 1: Do drawing prompts enhance learning when added to self-explanation prompts that support verbal sense-making processes with traditional visual representations? One may argue that self-explanations prompts that typically support verbal sense-making processes are sufficient. Yet, the research reviewed above suggests that drawing prompts complement verbal sense-making supports because they help students externalize their mental models and visually make sense of concepts shown in representations. As a result, students may be more likely to integrate new concepts from representations into their mental models. Hence, we test:

Hypothesis 1. Students who received drawing prompts show higher learning gains than students who received no drawing prompts with respect to instructional effectiveness (**H1a**) and instructional efficiency measures (**H1b**).

Research Question 2: *How frequently* should drawing prompts be provided? In prior research, drawing prompts were provided either throughout instruction or only before and after instruction (e.g., Gadgil et al., 2012; Schmeck et al., 2014). Because the frequency of drawings correlates with learning outcomes (Davis, 2000; Schank & Kozma, 2002), additional drawing prompts throughout instruction may lead to greater learning outcomes. Hence, we test:

Hypothesis 2. Students who received drawing prompts throughout instruction show higher learning gains than students who received prompts only before and after instruction with respect to instructional effectiveness (**H2a**) and instructional efficiency measures (**H2b**).

Finally, to examine how drawing prompts facilitate visual sense-making processes, we assess changes in the quality of students' drawings, as measured by alignment with traditional representations. Because we cannot compare drawings across time for each condition (e.g., the control condition did not receive drawing prompts during instruction), our analysis of drawing quality is exploratory. Specifically, to gain insights into the mechanisms underlying hypothesis 1, we examine whether students who received drawing prompts generate higher quality drawings than students who received no drawing prompts during instruction. Moreover, to gain insights into the mechanism underlying hypothesis 2, we examine whether students who received drawing prompts throughout instruction generate higher quality drawings than students who received prompts only before and after instruction.

3. Method

3.1. Participants

Seventy-two undergraduate students participated in this study for extra credit in an educational psychology course at a large Midwestern United States university. None of the students majored in chemistry. Some students have taken at least one introductory-level (68.1%) or one intermediate-level college chemistry course (22.2%).

3.2. Experimental design

Students were randomly assigned to one of three conditions. Students in the *no-prompt* condition ($n = 24$) received no drawing prompts while working on instructional activities. Students in the *before-after* condition ($n = 23$) received prompts only before and after they worked on instructional activities. Students in the *throughout* condition ($n = 25$) received prompts before and after as well as throughout instructional activities. All students sat at a computer with paper and pens to the right of the computer throughout the experiment and were allowed to draw spontaneously.

3.3. Materials

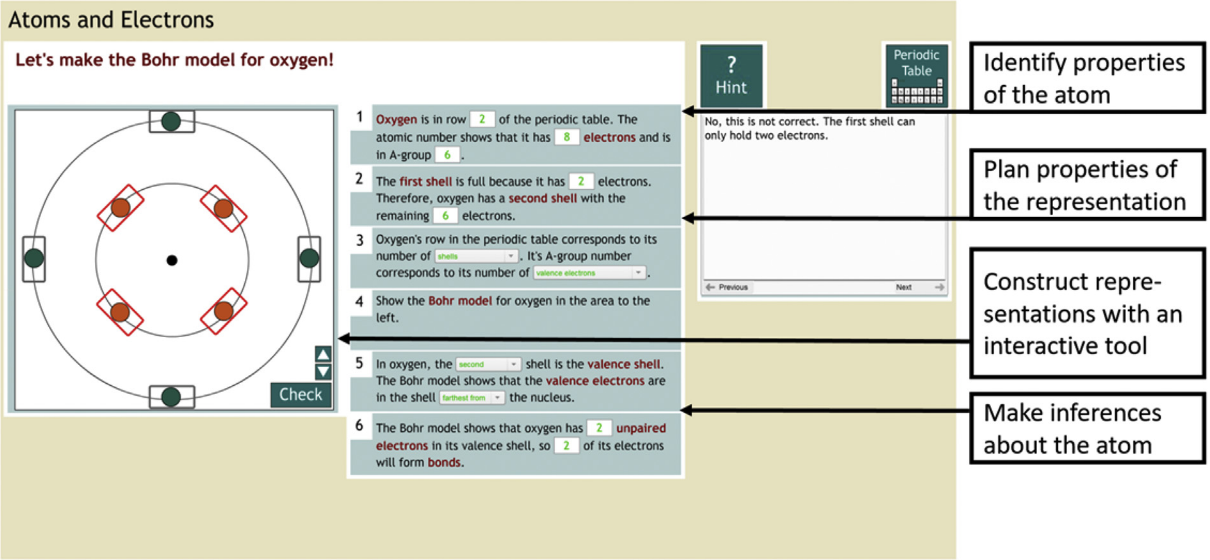
3.3.1. Educational technology: Chem Tutor

Instructional activities were provided via Chem Tutor, an educational technology for undergraduate chemistry that has been proven effective for science and non-science majors (Rau & Wu, 2015a). Chem Tutor provided students with interactive instructional activities on four visual representations of atoms: Lewis structures, Bohr models, energy diagrams, and orbital diagrams (see Fig. 1). First, students received an introduction to the representations. Then, they worked on two problem sets in which they used representations to learn about atomic structure. An example Chem Tutor problem is shown in Fig. 2. First, students identified properties of the atom and planned the representation by completing fill-in-the-blank explanations. Next, they used an interactive tool to construct the representation by selecting the appropriate number and position of electrons, shells, and orbitals for the atom. Finally, students made inferences about the atom based on the representation. For all interactions, Chem Tutor provided error-specific feedback and on-demand hints (Rau, Michaelis, & Fay, 2015c).

3.3.2. Drawing prompts

For students in the prompted conditions (before-after, throughout), Chem Tutor provided *generate* and *revise* prompts. Generate prompts asked students: “Draw what comes to mind when you think about the concept: ‘atom’.” Generate prompts were provided at the beginning of Chem Tutor. Revise prompts asked students: “Review your drawing, labels, and captions. Revise them as needed.” The before-after condition received one revise prompt after completing instructional activities in Chem Tutor. The throughout condition received three revise prompts interspersed with instructional activities in Chem Tutor. Table 1 shows examples of drawings students made in response to generate and revise prompts.

To control time-on-task across conditions, we calculated expected time-on-task using data from prior studies with Chem Tutor and pilot studies with eight undergraduates of various majors. We determined two minutes were needed on average per drawing prompt to allow students to read the prompt and draw an atom. For



Atoms and Electrons

Let's make the Bohr model for oxygen!

1 Oxygen is in row 2 of the periodic table. The atomic number shows that it has 8 electrons and is in A-group 6.

2 The first shell is full because it has 2 electrons. Therefore, oxygen has a second shell with the remaining 6 electrons.

3 Oxygen's row in the periodic table corresponds to its number of shells. It's A-group number corresponds to its number of valence electrons.

4 Show the Bohr model for oxygen in the area to the left.

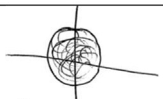
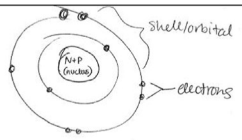
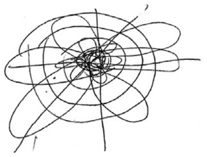
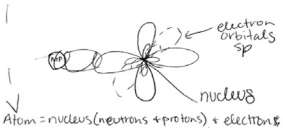
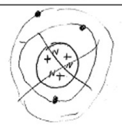
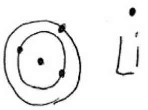
5 In oxygen, the second shell is the valence shell. The Bohr model shows that the valence electrons are in the shell farthest from the nucleus.

6 The Bohr model shows that oxygen has 2 unpaired electrons in its valence shell, so 2 of its electrons will form bonds.

No, this is not correct. The first shell can only hold two electrons.

Fig. 2. Example Chem Tutor activity about the Bohr model of oxygen.

Table 1
Example drawings from three students in response to the generate and revise prompts.

	Generate Prompt	Revise Prompt
Student A	 An atom is a small part of a cell that is represented as a sphere.	 An atom is a small part of a cell that is represented as a sphere
Student B	 The atom consists of a nucleus w/ a certain number of neutrally charged neutrons and positively charged protons. & Around the nucleus are the electron orbitals that contain negatively charged electrons. They stay near the nucleus because ~~they are~~ the protons & electrons are attracted to each other.	*Student added*:   Atom = nucleus (neutrons + protons) + electrons, shells where electrons energy state is at, and electron orbitals, where electrons are most likely to be.
Student C	 This atom has 3 protons ? 3 electrons - unstable because outer shell is not completely full	*Student crossed out previous drawing and added*: 

each condition, we then reduced the number of Chem Tutor problems in accordance with the number of drawing prompts they received. The no-prompt condition received 92 Chem Tutor problems, which took students an average of 31 min,[1] as conducted in a prior study (Rau & Wu, 2015a). To accommodate two drawing prompts, the before-after condition received 48 Chem Tutor problems (corresponding about 26 min). To accommodate four drawing prompts, the throughout condition received 44 Chem Tutor problems (corresponding to about 22 min).

3.3.3. Assessments

To assess *spatial skills*, we used the Vandenberg & Kuse mental rotation test (Peters et al., 1995). To assess students' learning gains, we administered three *chemistry tests* about the structure and properties of atoms (pre-test, post-test, and delayed-test). We

---

[1] The no-prompt condition received a set of 36 problems that each requires one step and, therefore, an additional minute to complete. We provide the average time to completion because each instructional problem requires a different number of problem-solving steps, and thus, a different amount of time to complete. We reduced instructional problems in even numbered sets to balance the pairs of representations presented.

created three isomorphic tests that were counterbalanced across the three test times. Cronbach's alpha showed acceptable reliability for all three test versions (0.78, 0.65, and 0.64). Each test included nine reproduction and transfer items. Appendix A provides example test items. Seven reproduction items assessed concepts discussed in Chem Tutor using multiple-choice responses. Two transfer items assessed concepts not explicitly taught in Chem Tutor using free-response. Two independent coders graded 10% of the transfer items on a scale of 0 (shows no understanding) to 3 (shows substantial understanding). Grading was highly reliable ($\text{ICC}(2, 2) = 0.87$ and 0.81, Shrout & Fleiss, 1979). Reproduction and transfer items were combined into one scale because a factor analysis showed that items did not load onto separate factors.

3.4. Procedure

Table 2 provides an overview of the procedure for each condition. The experiment involved two sessions, one week apart. Session 1 took approximately 90 min. Students first took the pre-test and spatial test. Then, they worked with the version of Chem Tutor that corresponded to their condition. Finally, students completed the post-test. In Session 2, all students received a generate prompt and completed the delayed-test.

3.5. Analyses

Instructional effectiveness corresponds to a student's test score, computed as the proportion of correct answers out of the nine possible correct answers. We computed *instructional efficiency* using Z-standardized test scores and Z-standardized time-on-task for learning time, as discussed by van Gog and Paas (2008):

$$\text{Instructional efficiency} = \frac{z_{\text{test score}} - z_{\text{learning time}}}{\sqrt{2}}$$

We computed total *time-on-task* as the sum of the time students spent on all Chem Tutor and drawing activities.

To assess the quality of student drawings, we used video recordings to identify when student generated and revised drawings throughout session 1 and session 2. Because the type of representations students chose to draw affected which features they drew, we coded drawings using a two-step process (see Appendix B for the full coding scheme). First, we categorized the *type of drawing* by counting features in the drawing that aligned with the representations presented in Chem Tutor (Bohr model, energy diagram, Lewis structure, orbital diagram). Drawings that did not align with any of the four representations were categorized as "Other." Second, we graded the *accuracy of drawing* by rating features shown in the drawing on a scale of 0 (inaccurate) to 4 (accurate). Two independent coders graded 11% of the drawings using the two-step process; grading was highly reliable ($\text{ICC}(2, 2) = 0.98$ and 0.91, Shrout & Fleiss, 1979).

Table 2
Overview of the procedure by session and condition.

		Condition		
		Throughout	Before-after	No-prompt
Session 1	Pre-test	X	X	X
	Spatial test	X	X	X
	Generate Prompt	X	X	
	Introduction	X	X	X
	Revise Prompt 1	X		
	Problem Set	X	X	X
	Revise Prompt 2	X		
	Problem Set	X	X	X
	Revise Prompt Final	X	X	
	Post-test	X	X	X
One week delay				
Session 2	Generate Prompt	X	X	X
	Delayed-test	X	X	X

4. Results

Table 3 shows means and standard deviations of test scores by condition.

We report effect sizes as Cohen's d and η_p^2 (Cohen, 1988). An effect size d of 0.20 corresponds to a small effect, 0.50 to a medium effect, and 0.80 to a large effect. An effect size η_p^2 of 0.01 corresponds to a small effect, 0.06 to a medium effect, and 0.14 to a large effect. For post-hoc analyses, we used Bonferroni correction of multiple comparisons and report adjusted p-values.

4.1. Prior checks

First, we checked for differences between conditions on the pre-test. A one-way ANOVA with condition as the between-subjects factor and pre-test scores as the dependent measure showed no main effect of condition on pre-test, $F(2, 69) = 2.926$, $p = 0.060$. Pre-test scores significantly correlated with post-test ($r = 0.633$, $p < 0.001$) and delayed-test scores ($r = 0.472$, $p < 0.001$). Therefore, we included pre-test scores as a covariate in the analyses of condition effects below. Second, we checked for differences between conditions on the spatial test using the same one-way ANOVA with spatial test scores as the dependent measure. There were no main effect of conditions on spatial test, $F(2, 69) = 0.747$, $p = 0.478$. Spatial test scores significantly correlated with the pre-test ($r = 0.237$, $p = 0.045$), post-test ($r = 0.256$, $p = 0.030$), and delayed-test scores ($r = 0.421$, $p < 0.001$). Therefore, we included spatial test scores as covariates in the analyses of condition effects below.

Third, we checked whether students' chemistry knowledge improved from pre-test to posttests. A repeated-measures ANOVA with test-time (pre-test, post-test, delayed-test) as the independent factor and test scores as dependent measures revealed large-sized significant learning gains, $F(2, 142) = 57.229$, $p < 0.001$, $\eta_p^2 = 0.446$. Post-hoc comparisons showed that students performed significantly better on the post-test, $t(71) = 8.611$, $p(adj) < 0.001$, $d = 0.875$, and delayed-test, $t(71) = 9.190$, $p(adj) < 0.001$, $d = 1.109$, compared to the pre-test. Post-test scores did not significantly differ from delayed-test scores, $t(71) = 2.176$, $p(adj) = 0.108$.

Fourth, we checked whether time-on-task differed between conditions using a one-way ANOVA with condition as the between-subjects factor and time-on-task as the dependent measure. Table 4 shows the mean and standard deviations for total duration of time-on-task as well as duration for each drawing task within the tutor by condition. There was a significant main effect of condition, $F(2, 69) = 6.566$, $p = 0.002$. Students in the no-prompt condition spent significantly more time-on-task than students in the before-after condition, $t(46) = 3.585$, $p(adj) = 0.002$, $d = 1.009$. This difference in time-on-task resulted from students in the throughout and

Table 3
Means and standard deviations (in parentheses) of students' test scores by condition.

Condition	Spatial test	Chemistry test		
		Pre-test	Post-test	Delayed-Test
Throughout	0.64 (0.27)	0.26 (0.17)	0.44 (0.15)	0.54 (0.16)
Before-after	0.70 (0.16)	0.39 (0.16)	0.51 (0.16)	0.51 (0.15)
No-prompt	0.70 (0.18)	0.33 (0.20)	0.49 (0.22)	0.51 (0.21)

Table 4
Mean and standard deviations (in parentheses) of duration for time-on-task (in minutes) and prompts (in seconds).

Condition	Time-on-task (*minutes*)	Prompt duration (*seconds*)			
		Generate	Revise 1	Revise 2	Revise Final
Throughout	53.32 (7.66)	85.54 (45.24)	19.80 (34.45)	18.64 (27.13)	28.32 (31.77)
Before-after	49.95 (7.42)	70.88 (29.66)	–	–	12.17 (9.99)
No-prompt	58.65 (9.67)	–	–	–	–

before-after conditions taking less time on average for drawing prompts than expected from pilot tests. Moreover, students in the before-after condition spent significantly less time on the final revise prompt than students in the throughout condition, $t(45) = 2.488$, $p = 0.019$, $d = 0.720$.

4.2. *Effects of drawing prompts*

To test the effects of drawing prompts on *instructional effectiveness*, we used a repeated-measures ANCOVA with test-time (post-test, delayed-test) as the repeated within-subjects factor, condition as the between-subjects factor, pre-test and spatial-test scores as covariates, and scores on the post-test and delayed-test as dependent measures. To test H1a (prompting students to draw enhances learning gains), we compared the two drawing prompt conditions to the no-prompt condition. To test H2a (prompting students to draw throughout instruction more so enhances learning gains), we compared the throughout condition to the before-after condition. Fig. 3 provides a summary of the findings.

There was no main effect of condition, $F(2, 67) = 1.256$, $p = 0.291$. However, there was a medium-sized significant interaction of test-time with condition, $F(2, 67) = 4.201$, $p = 0.019$, $\eta_p^2 = 0.111$. Post-hoc analyses showed no differences between conditions at post-test, $F(2, 67) = 0.078$, $p = 0.925$. However, there was a medium-sized significant difference at the delayed-test, $F(2, 67) = 3.881$, $p = 0.025$, $\eta_p^2 = 0.104$. Contrary to H1a, the throughout condition, $t(48) = 2.100$, $p(adj) = 0.121$, $d = 0.614$, and the before-after condition, $t(46) = 0.659$, $p(adj) = 1.000$, did not outperform the no-prompt condition. In support of H2a, the throughout condition outperformed the before-after condition, $t(47) = 2.643$, $p(adj) = 0.031$, $d = 0.782$. In sum, findings on instructional effectiveness partially support H2a, but not H1a: drawing prompts *throughout* instruction are more effective than drawing prompts before and after instruction, but only at the delayed-test.

To test the effects of drawing prompts on *instructional efficiency* (H1b and H2b), we used the same ANCOVA model with efficiency scores as dependent measures. Fig. 4 provides a summary of the findings.

There was a medium-sized significant main effect of condition, $F(2, 67) = 5.051$, $p = 0.009$, $\eta_p^2 = 0.131$. In support of H1b, the throughout condition, $t(48) = 3.465$, $p(adj) = 0.003$, $d = 0.992$, and the before-after condition, $t(46) = 2.599$, $p(adj) = 0.035$, $d = 0.757$, were significantly more efficient than the no-prompt condition. Mean instructional efficiency scores for the throughout ($M = 0.228$, $SD = 0.769$) and before-after conditions ($M = 0.213$, $SD = 0.769$) were significantly higher than the no-prompt condition ($M = -0.367$, $SD = 0.755$). Further, there was a medium-sized significant interaction of test-time with condition, $F(2, 69) = 4.324$, $p = 0.017$, $\eta_p^2 = 0.114$. Post-hoc analyses showed medium-sized differences between conditions at post-test, $F(2, 67) = 3.556$, $p = 0.034$, $\eta_p^2 = 0.096$, and large-sized differences at delayed-test, $F(2, 67) = 6.372$, $p = 0.003$, $\eta_p^2 = 0.160$. The before-after condition was significantly more efficient than the no-prompt condition at post-test, $t(46) = 2.599$, $p(adj) = 0.035$, $d = 0.757$, but not at delayed-test, $t(46) = 2.401$, $p(adj) = 0.058$. The throughout condition was significantly more efficient than the no-prompt condition at delayed-test, $t(48) = 3.465$, $p(adj) = 0.003$, $d = 0.992$, but not at post-test, $t(48) = 1.758$, $p(adj) = 0.250$. Contrary to H2b, we found no significant differences between mean instructional efficiency scores for the throughout and before-after conditions, $t(47) = 0.852$, $p(adj) = 1.000$. We also found no significant interactions between these conditions at post-test, $t(47) = 0.852$, $p(adj) = 1.000$, or at delayed-test, $t(47) = 0.983$, $p(adj) = 0.989$. In sum, these findings on instructional efficiency support H1b, but not H2b: drawing prompts enhance instructional efficiency if prompts were provided throughout instruction (especially at the delayed-test) and before and after instruction (especially at the post-test).

4.3. *Exploration of drawing quality*

To explore changes in drawing quality over time, we analyzed the *types of drawings* students generated (Appendix B, Step 1) and *accuracy* of their drawings (Appendix B, Step 2). Because the number of drawings by type and condition was low ($n < 5$; see Table 5), we cannot statistically test for effects of condition on

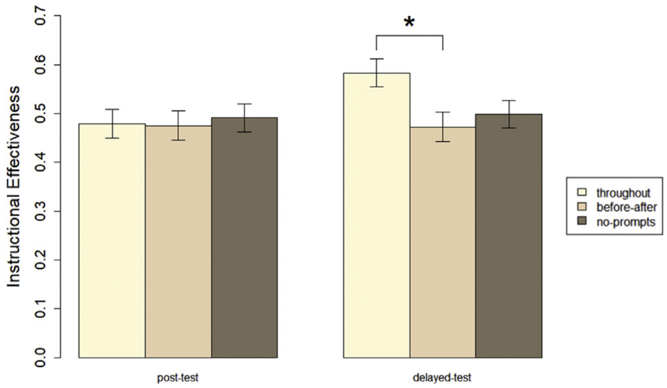


Fig. 3. Estimated marginal means for instructional effectiveness at post-test and delayed-test by condition. Error bars show standard errors. * = $p < 0.05$.

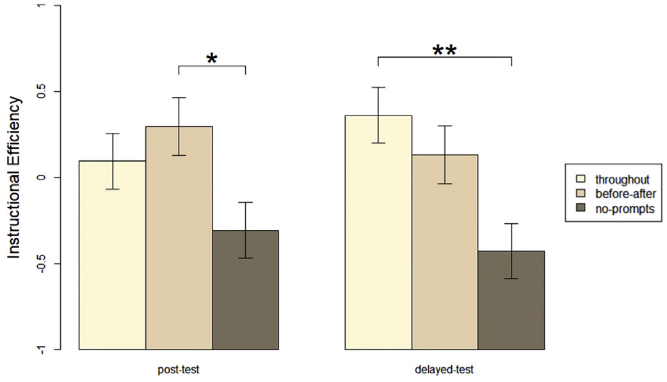


Fig. 4. Estimated mean instructional efficiency at post-test and delayed-test by condition. Estimated means are shown on a standardized scale from −1 to 1. Error bars depict standard errors. * = $p < 0.05$; ** = $p < 0.01$.

Table 5
Frequency of the five types of representations in student drawings by time of prompt and condition.

	Type of Representation					Total
	Lewis	Bohr	Energy	Orbital	Other	
Session 1						
Before Instruction (Pre-test)						
Throughout	1	11	0	0	16	**28**
Before-after	3	8	0	1	12	**24**
During Instruction						
Throughout	2	7	2	4	1	**16**
After Instruction (Post-test)						
Throughout	1	5	1	3	0	**10**
Before-after	3	2	0	2	1	**8**
Session 2 *(Delayed-test)*						
Throughout	3	12	0	2	9	**26**
Before-after	2	10	0	2	10	**24**
No-prompt	0	12	0	2	10	**24**

drawing quality. Hence, we qualitatively analyze the types and accuracy of drawings.

With respect to the *types of drawings*, Table 5 shows that students chose to generate a variety of drawing types. At each time of drawing, students across conditions chose to draw similar types of drawings. However, when comparing drawings over time, the throughout and before-after condition seem to generate fewer "Other" representations from session 1 before instruction to session 2.

With respect to *the accuracy of drawings*, Table 6 shows that mean accuracy scores differed by drawing type, time of drawing, and condition. Recall that types differ with respect to how many conceptual aspects they can depict. Hence, students' choice for type of drawing affects the accuracy of their drawings. Qualitative inspection of total means shows that students generated less accurate drawings when they were first prompted to draw: The drawings of students in the no-prompt condition in session 2 achieved accuracy scores ($M = 1.83$) as low as the scores for the drawings of students in the throughout ($M = 1.83$) and before-after conditions ($M = 1.86$) in session 1 before instruction. For the two prompted conditions, drawing accuracy improved equally over time in session 1: at the post-test after instruction, accuracy scores were similar ($M = 3.10$ for the throughout condition and $M = 3.25$ for the before-after condition). In session 2, drawing accuracy was highest for the throughout condition ($M = 2.42$), lower for the before-after condition ($M = 2.08$), and lowest for the no-prompt condition ($M = 1.83$).

Table 6
Mean drawing accuracy scores for each type of representation by time of prompt and condition. Drawing scores range from 0 to 4. Underlined text highlights the mean score for first drawing prompt for each condition, Test times (pre-test, post-test, delayed-test) are shown to facilitate analysis of change over time.

	Type of Representation					**Total**
	Lewis	Bohr	Energy	Orbital	Other	
Session 1						
Before Instruction (Pre-test)						
Throughout	4	2.82	–	–	1.06	**1.86**
Before-after	3.33	2.36	–	1	1.17	**1.83**
During Instruction						
Throughout	3.5	2.43	1.5	2.75	–	**2.38**
After Instruction (Post-test)						
Throughout	4	3.4	1	3	–	**3.10**
Before-after	4	2.5	–	4	1	**3.25**
Session 2 *(Delayed-test)*						
Throughout	4	3.17	–	2.5	0.89	**2.42**
Before-after	4	2.3	–	2.5	1.4	**2.08**
No-prompt	–	2.58	–	3	0.7	**1.83**

5. Discussion

This experiment examined whether adding drawing prompts to an educational technology that supports verbal sense making of traditional representations enhances students' learning of domain knowledge. We first considered instructional effectiveness, which controlled for time-on-task at the level of condition by reducing instructional activities for conditions with more drawing prompts. With respect to Research Question 1, we compared two experimental conditions in which drawing prompts were provided (1) throughout instruction and (2) before and after instruction to (3) a control condition without drawing prompts. Our findings suggest that providing drawing prompts is not more effective than typical instructional activities (H1a not supported). With respect to Research Question 2, we investigated whether the frequency of drawing prompts affects instructional effectiveness by comparing the two prompted conditions. Our findings show that, on a delayed posttest, providing drawing prompts *throughout* instruction is more effective than providing prompts only before and after instruction (H2a partially supported).

Second, we considered instructional efficiency that accounted for time-on-task at the level of the individual student. With respect to Research Question 1, our findings show that drawing prompts enhance instructional efficiency (H1b supported). However, providing drawing prompts throughout instruction does not further increase instructional efficiency compared to providing prompts before and after instruction (H2b not supported).

Finally, we analyzed how the quality of students' drawings changed over time to gain insights into how drawing prompts affect students' visual sense-making processes. In support of the quantitative results, our findings suggest that more frequent drawing prompts enhance the quality of student drawings over time and that this effect is prominent after a delay. When receiving more drawing prompts, students seem to generate drawings that increasingly align with traditional representations, which may reflect students' development of increasingly more accurate and sophisticated mental models of the domain knowledge.

We now discuss each of our main findings in turn. One of our main findings is that the frequency of prompts affects instructional effectiveness: providing drawing prompts *throughout instruction*—particularly prompts to *revise* drawings—was more effective at a delayed posttest than providing prompts only before and after instruction. In fact, after one week without instruction, students who were prompted to draw throughout instruction achieved higher quality drawings and *learning gains* at the delayed posttest. Our results align with prior research suggesting that prompting students to review their own drawings is effective because it engages students' mental models with domain-relevant concepts (Gadgil et al., 2012; Vosniadou, 1994). Recall that students monitor, compare, and revise their representations in the third phase of the CMDC (Van Meter & Firetto, 2013). Providing revise prompts throughout instruction may help students identify and remedy inaccuracies in their own mental models.

However, these effects were limited to the *delayed posttest* one week after instruction. We did not find differences at the immediate posttest. The qualitative analysis of students' drawings corroborates this result. At the immediate posttest, drawing accuracy was similar for the two prompted conditions. However, at the delayed posttest, students in the throughout condition generated more accurate drawings than students in the before-after condition. We interpret these findings in light of prior research, which shows that drawing activities—by virtue of engaging students more deeply with the content—can increase cognitive load and interfere with students' immediate performance (Schwamborn, Thillmann, Opfermann, & Leutner, 2011; Van Meter et al., 2006).

After a delay, the effects of cognitive load fade, and students' performance reveals long-term benefits of deep engagement with the content (Schweppe & Rummer, 2016). A follow-up experiment will investigate the possible role of cognitive load, which was not measured in this experiment. However, our findings illustrate that delayed posttests are useful to reveal the effects of drawing prompts.

A second main finding was that drawing prompts enhance *instructional efficiency*. That is, students who received drawing prompts showed higher learning gains in less instructional time than students who did not receive these prompts. Specifically, we found that while drawing prompts throughout instruction demonstrated instructional efficiency at the delayed posttest, prompts provided before and after instruction demonstrated instructional efficiency immediately after instruction. In accordance with the second phase of the CMDC (Van Meter & Firetto, 2013), we propose that drawing prompts increase instructional efficiency because they direct students' attention to concepts they may not yet understand and help them organize relevant visual and verbal information. This interpretation aligns with prior research, which shows that drawing prompts guide students' interactions with traditional representations in a way that helps them integrate domain knowledge in their existing mental models (Valanides et al., 2013; Zhang & Linn, 2011). Our finding extends prior research by showing that such support for visual sense-making processes can increase learning from instructional activities that support verbal sense-making processes—in less instructional time.

These two findings make several contributions to the literature on learning and instruction. First, our quantitative and qualitative analyses suggest that students do not revise their mental models sufficiently unless frequently prompted to do so throughout instruction. We propose that revise prompts facilitate mental model revision, in which students visually make sense of traditional representations, compare their own mental models to traditional representations, and revise their mental models accordingly. Moreover, we find that the effects of this process may only be apparent after a delay. Second, our findings show that drawing prompts are an efficient means to enhance learning from instructional activities that support verbal sense-making of traditional representations. We propose that visual sense-making processes complement verbal sense-making processes, potentially serving as an intermediate step that helps students translate concepts from traditional visual representations into verbal explanations. Third, our experiment expands prior research by stringently controlling for time-on-task in two ways. We replaced instructional activities with drawing prompts to address the methodological issue of adding instructional time for drawing activities. We also controlled for instructional time by individual students and found that drawing prompts help students achieve higher learning gains in less instructional time. Taken together, our findings suggest that drawing prompts can engage students in visual sense-making processes that help them learn from instructional activities that support verbal sense making of traditional visual representations.

5.1. Implications for instruction

Our findings have several implications for instruction. First, our findings suggest that instructors should add drawing prompts to instructional activities with traditional visual representations. Our experiment shows that asking students to draw with generic prompts to generate and revise drawings with paper and pen can enhance an educational technology. Such drawing prompts are easy to add to existing instructional activities.

Second, our findings suggest that instructors should prompt students to draw not only before and after instruction, but throughout instruction. Given that students typically do not spontaneously draw (Leutner & Schmeck, 2014; Van Meter et al., 2006), our findings suggest that students benefit from frequent prompts, especially prompts to revise drawings throughout instruction.

Finally, our findings yield instructional design recommendations for educational technologies. Replacing instructional time in educational technologies with drawing prompts throughout instruction may enhance students' learning from educational technologies. Further, our experiment implemented simple drawing prompts without guidance on how to draw or feedback on the drawings. Hence, simply providing generic drawing prompts may enhance the effectiveness of educational technologies, when instructional time is limited.

5.2. Limitations and future directions

Our results should be interpreted against the following limitations. First, we tested the effects of drawing prompts in the context of a particular educational technology. Albeit a realistic and effective educational technology for learning with traditional visual representations, future research should investigate whether our results generalize to other instructional activities that are not technology based. Second, we conducted the current experiment in a laboratory setting to increase internal validity, assess drawing accuracy, and control for instructional time. For external validity, future research will investigate whether our findings generalize to realistic educational settings. Third, our experiment focused on a specific chemistry topic: atomic structure. While atomic structure is similar to many other STEM topics because it uses multiple representations that depict visual-spatial concepts, future research should determine whether our results generalize to other STEM topics. Fourth, participants were non-science-major undergraduate students. Their study motivation and interactions with drawing prompts may differ from students majoring in chemistry or other STEM domains. Future research will investigate whether the effects of drawing prompts generalize to other student populations.

Fifth, we conducted a qualitative analysis of drawing quality between conditions because only a subset of students received drawing prompts during instruction, and thus our sample of drawings is too small. Future research will collect a larger sample of drawings across conditions to assess quantitative changes in drawing quality. Sixth, we prompted all students to draw before the delayed posttest to compare drawings across all students. Because this drawing prompt may affect outcomes for students who were previously not prompted to draw, future studies will prompt students to draw after the delayed posttest. Seventh, increased retrieval of information, known as the mnemonic effect (Roediger & Karpicke, 2006), may explain differences in drawing quality between students who received drawing prompts throughout instruction and students who received prompts before and after. Future studies comparing different timing and types of drawing prompts will test this possibility.

Eighth, efficiency of drawing prompts that replaced instructional activities in the educational technology could also be interpreted as inefficiency of the replaced instructional activities. We do not think this is likely because prior research has demonstrated the effectiveness of these instructional activities (Rau et al., 2015c). However, future research should test whether decreasing the number of activities without adding drawing prompts would be equally efficient.

Finally, we found that providing drawing prompts throughout instruction was more effective at the delayed posttest compared to prompts before and after instruction, but not compared to the control condition. Our results align with a recent study that compared drawing prompts to self-explanation prompts and found no significant differences in learning outcomes because both instructional activities support sense-making processes (Scheiter et al., 2017). We also found that students who were prompted to draw during instruction produced higher quality drawings, compared to students not prompted to draw. We speculate that differences in students' encoding of traditional visual representations between verbal sense-making processes (e.g., descriptions) and visual sense-making processes (e.g., imagery) may explain the drawing outcomes (Schnotz, 2014). The students not prompted to draw during instruction may have to translate verbal descriptions of traditional representations into visual imagery, leading to lower drawing quality. It may be that verbal sense making does not adequately prepare students for visual sense making of representations; thus, visual representations must be learned via visual sense-making processes. A future think-aloud study will examine these potential processes underlying visual and verbal sense making with traditional visual representations.

6. Conclusion

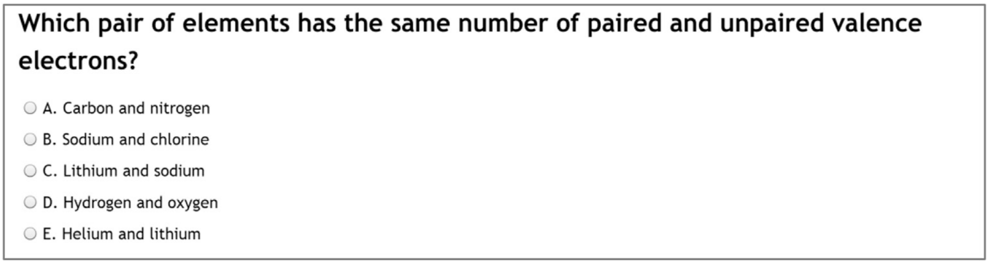
Our results show that drawing prompts can enhance learning with traditional visual representations in an educational technology. Our findings suggest that drawing prompts facilitate visual sense-making processes, which complement verbal sense-making processes that are typically supported by instructional interventions. We extend prior research by demonstrating the efficiency of drawing prompts when controlling for time-on-task by replacing instructional activities with drawing prompts. Further, our experiment suggests that prompts to generate and revise drawings should be provided throughout instruction rather than only before and after. Because drawing prompts can be added to any type of instructional intervention without requiring additional time, our findings have broad practical implications for classroom-based instruction.

Funding

This research was supported by the Wisconsin Center for Education Research, the National Science Foundation through Award #1611782, and the Institute of Education Sciences, U.S. Department of Education through Award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or the National Science Foundation.

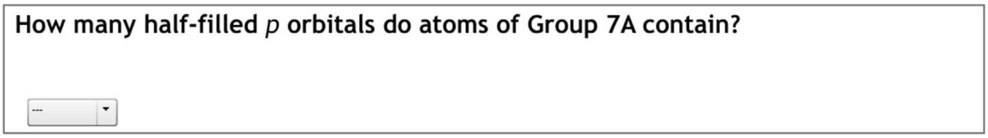
Appendix A

Example chemistry test questions.

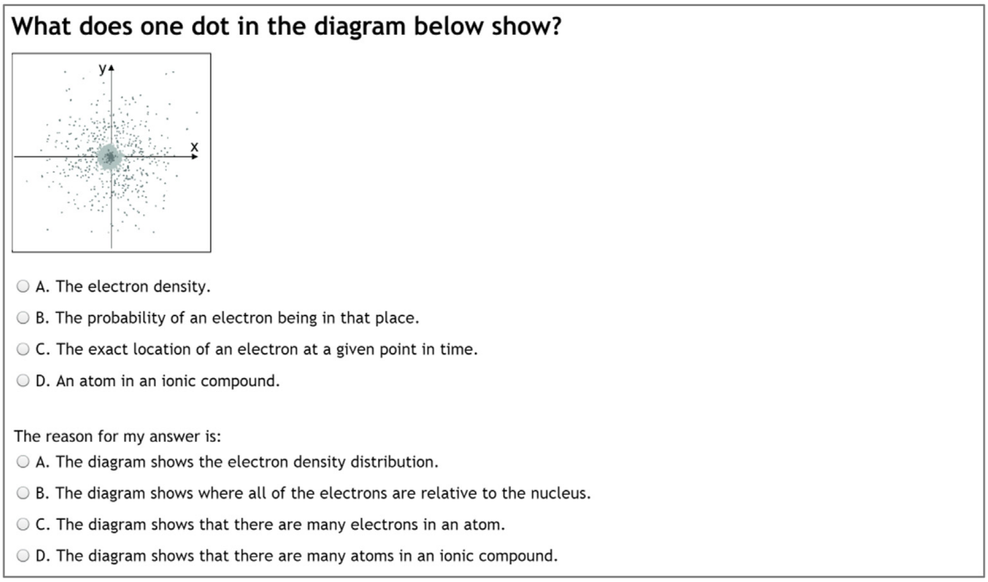


Which pair of elements has the same number of paired and unpaired valence electrons?

○ A. Carbon and nitrogen
○ B. Sodium and chlorine
○ C. Lithium and sodium
○ D. Hydrogen and oxygen
○ E. Helium and lithium



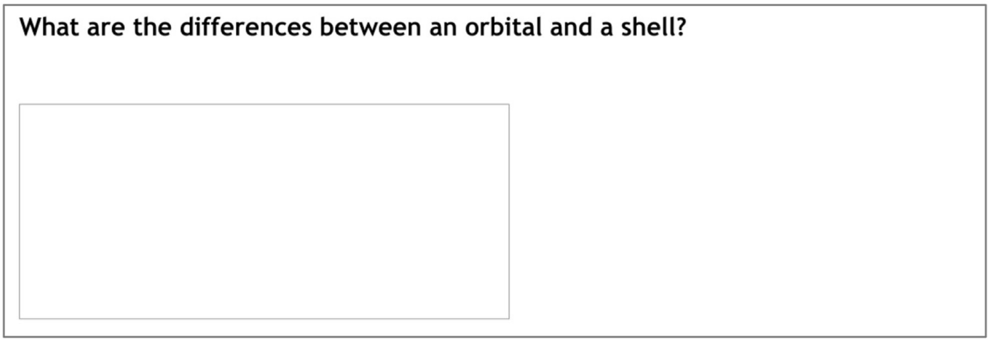
How many half-filled *p* orbitals do atoms of Group 7A contain?



What does one dot in the diagram below show?

○ A. The electron density.
○ B. The probability of an electron being in that place.
○ C. The exact location of an electron at a given point in time.
○ D. An atom in an ionic compound.

The reason for my answer is:
○ A. The diagram shows the electron density distribution.
○ B. The diagram shows where all of the electrons are relative to the nucleus.
○ C. The diagram shows that there are many electrons in an atom.
○ D. The diagram shows that there are many atoms in an ionic compound.

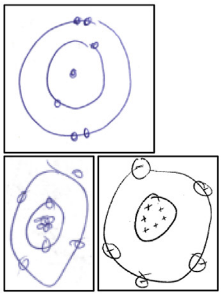
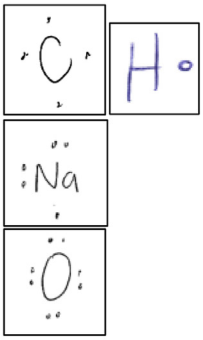
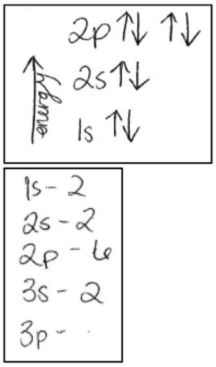
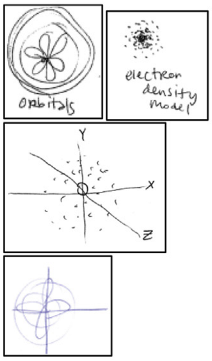
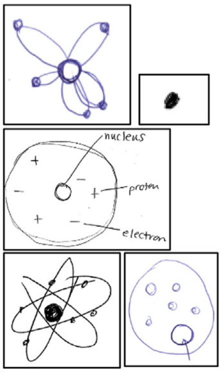


What are the differences between an orbital and a shell?

Appendix B

Coding scheme for drawings.

Step 1
Code by type of representation that best match the drawing.

	Bohr model	**Lewis structure**	**Energy diagram**	**Orbital diagram**	**Other**
Nucleus	One dot/circle or cluster of dots/circle in the center	Atomic symbol in the center	Not shown	Not shown; density at the center; origin of x-y plot	No nucleus; dot or cluster of dots/circles
Electrons	Dots/circles on rings around the nucleus	Dots around atomic symbol	Arrows of opposite directions; numbers in rows by energy level	No distinct electrons shown; orbitals filled in with dots or shaded in	No electrons
Energy levels	Circular paths around the nucleus	No energy levels shown	Rows of specific energy levels	Spherical/loop shapes around center	No energy levels; overlapping loops
Optional features	Protons as +; electrons as – or e⁻		Arrow pointing up to indicate energy	x-y axes	
Examples					

Step 2
Code drawing by the accuracy of features shown within each type of representation.

Representation	Coding	Description	Example	Notes
Bohr	−1 point	Incorrect number of shells	2 shells for Hydrogen	
	−1 point	Incorrect number of electrons	13 electrons for Sodium	
	−1 point	Incorrect number of electrons in any shell	4 electrons in the first shell	−1 also for missing electrons
	−1 point	Incorrect pairing of electrons	Electrons not paired; electrons paired when they should be spread out	−1 also for missing electrons
Lewis	−1 point	Incorrect atomic symbol	"S" for sodium	
	−1 point	Incorrect number of electrons		
	−1 point	Incorrect pairing of electrons	Electrons not paired; electrons paired when they should be spread out	
Energy	−1 point	Incorrect number of orbitals/misplaced orbitals	2p before 2s	
	−1 point	Incorrect number of electrons	6 electrons for oxygen	
	−1 point	Incorrect number of electrons in any orbital	6 electrons in 2s	
	−1 point	Incorrect pairing of electrons/spin	Electrons not paired; electrons paired when they should be spread out	

(continued on next page)

Step 2 (*continued*)

Representation	Coding	Description	Example	Notes
Orbital	−1 point	Incorrect number of orbitals	2p orbitals without a 2s orbital	
	−1 point	Misplaced orbitals	Orbitals not centered around nucleus	
	−1 point	Incorrect orbital shapes; Inaccurate electron density	Overlapping ovals for p-orbitals; lack of random electrons	
	−1 point	Missing 1s orbital	Only p-orbitals	There is always an 1s orbital
Other	-? points	Use all coding for each of the above types of representation that apply	Wrong shells/electrons OR orbital shapes/placement for “Jimmy Neutron” atom	Calculate possible points for each type of representation and use the best score

References

Ainsworth, S. E. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*(1), 70–87. https://doi.org/10.1037/a0013247.

Bobek, E., & Tversky, B. (2014). Creating visual explanations improves learning. In *Proceedings of the 36th annual conference of the cognitive science society* (pp. 206–211). Austin, TX: Cognitive Science Society.

Brooks, M. (2009). Drawing, visualisation and young children's exploration of “big ideas.” *International Journal of Science Education, 31*(3), 319–341. https://doi.org/10.1080/09500690802595771.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182. https://doi.org/10.1016/0364-0213(89)90002-5.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Erlbaum.

Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education, 92*(5), 848–867. https://doi.org/10.1002/sce.20262.

Davis, E. A. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education, 22*(February 2015), 819–837. https://doi.org/10.1080/095006900412293.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction, 22*(1), 47–61. https://doi.org/10.1016/j.learninstruc.2011.06.002.

Gilbert, J. K. (2005). *Visualization: A metacognitive skill in science and science education. In visualization in science education.* Netherlands: Springer.

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations-a meta-analytic review. *Educational Psychology Review, 22*, 245–269. https://doi.org/10.1007/s10648-010-9126-7.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x.

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representational competence. *Visualization in Science Education*, 121–145.

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction, 22*(1), 16–26. https://doi.org/10.1016/j.learninstruc.2011.05.005.

Leutner, D., & Schmeck, A. (2014). The generative drawing principle in multimedia learning. In R. Mayer (Ed.), *The cambridge handbook of multimedia learning* (pp. 433–448). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.022.

Mason, L., Lowe, R., & Tornatora, M. C. (2013). Self-generated drawings for supporting comprehension of a complex animation. *Contemporary Educational Psychology, 38*(3), 211–224. https://doi.org/10.1016/j.cedpsych.2013.04.001.

Mathewson, J. H. (1999). Visual-spatial thinking: An aspect of science overlooked by educators. *Science Education, 83*(1), 33–54.

van der Meij, J., & de Jong, T. (2011). The effects of directive self-explanation prompts to support active processing of multiple representations in a simulation-based learning environment. *Journal of Computer Assisted Learning, 27*(5), 411–423. https://doi.org/10.1111/j.1365-2729.2011.00411.x.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032.

Prain, V., & Tytler, R. (2012). Learning through constructing representations in science: A framework of representational construction affordances. *International Journal of Science Education, 34*(17), 2751–2773. https://doi.org/10.1080/09500693.2011.626462.

Rau, M. A. (2016). Conditions for the effectiveness of multiple visual representations in enhancing STEM Learning. *Educational Psychology Review*, 1–45. https://doi.org/10.1007/s10648-016-9365-3.

Rau, M. A., Aleven, V., Rummel, N., & Pardos, Z. (2014). How should intelligent tutoring systems sequence multiple graphical representations of fractions? A multi-methods study. *International Journal of Artificial Intelligence in Education, 24*. https://doi.org/10.1007/s40593-013-0011-7.

Rau, M. A., & Wu, S. P. W. (2015a). ITS support for conceptual and perceptual processes in learning with multiple graphical representations. In C. Conati, N. Heffernan, A. Mitrovic, & F. Verdejo (Eds.), *Proceedings of the 17th international conference of artificial intelligence in education* (pp. 398–407). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-642-39112-5.

Rau, M. A., Aleven, V., & Rummel, N. (2015b). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology, 107*(1), 30–46. https://doi.org/10.1037/a0037211.

Rau, M. A., Michaelis, J. E., & Fay, N. (2015c). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers and Education, 82*, 460–485. https://doi.org/10.1016/j.compedu.2014.12.009.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education, 0*(123456789), 1–13. https://doi.org/10.1007/s11858-017-0834-z.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1467-8721.2008.00612.x.

Roelle, J., Lehmkuhl, N., Beyer, M., & Berthold, K. (2015). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instructions. *Journal of Educational Psychology, 107*(3), 705–723. https://doi.org/10.1037/edu0000010.

Schank, P., & Kozma, R. (2002). Learning chemistry through the use of a representation-based knowledge building environment. *Journal of Computers in Mathematics and Science Teaching, 21*(3), 253–279.

Scheiter, K., Schleinschok, K., & Ainsworth, S. E. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in Cognitive Science*, 1–17. https://doi.org/10.1111/tops.12261.

Schmeck, A., Mayer, R. E., 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. https://doi.org/10.1016/j.cedpsych.2014.07.003.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (pp. 72–103).

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior, 27*(1), 89–93. https://doi.org/10.1016/j.chb.2010.05.028.

Schweppe, J., & Rummer, R. (2016). Integrating written text and graphics as a desirable difficulty in long-term multimedia learning. *Computers in Human Behavior, 60*(August), 131–137. https://doi.org/10.1016/j.chb.2016.02.035.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012.

Talanquer, V. (2013). Chemistry education: Ten facets to shape us. *Journal of Chemical Education, 90*(7), 832–838. https://doi.org/10.1021/ed300881v.

Valanides, N., Efthymiou, I., & Angeli, C. (2013). Interplay of internal and external representations: Students' drawings and textual explanations about shadow phenomena. *Journal of Visual Literacy, 32*(2), 67–84.

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. https://doi.org/10.1016/j.cedpsych.2005.04.001.

Van Meter, P., & Firetto, C. M. (2013). Cognitive model of drawing construction. *In Learning Through Visual Displays*, 247–280.

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. https://doi.org/10.1007/s10648-005-8136-3.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69. https://doi.org/10.1016/0959-4752(94)90018-3.

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (pp. 413–432). Cambridge University Press.

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching, 48*(10), 1177–1198. https://doi.org/10.1002/tea.20443.