

Visualizing the Invisible: Generating Explanations of Scientific Phenomena

Eliza Bobek (ebobek@packer.edu)

The Packer Collegiate Institute
Brooklyn, NY 11238 USA

Abstract

This study investigated the ability of learner-generated visualizations to improve learning in science. The hypothesis was tested in two domains, a mechanical system and a chemical system, and the results were analyzed separately to compare low and high spatial ability learners. The production of visual explanations of a mechanical system, a bicycle tire pump, increased understanding of the pump particularly for participants with low spatial ability. In the domain of chemical bonding, visual explanations were more effective than verbal explanations for all participants. Visual explanations often included crucial yet invisible features; their accurate construction requires and provides a check for completeness of explanations.

Keywords: learning; drawing; external representation; structure; function; spatial ability; self-generated explanation

Introduction

Many topics in science are notoriously difficult for students to learn. Mechanisms and processes that exist on a scale outside student experience, such as gravitational pull, chemical bonding, and cellular processes, present particular challenges. When students attempt to learn these phenomena, they often experience difficulty because they must understand not only the individual components of the process (structure) but also the interactions and mechanisms (function). While instruction often involves visualizations, students typically explain in words, spoken or written. Visualizations have many advantages over verbal explanations, especially for science, so asking student to produce visual rather than verbal explanations should improve their learning.

Learner-generated Explanations

When learners make connections between information, knowledge, and experience, by generating headings, summaries, pictures, and analogies, deeper understanding develops (Wittrock, 1990). Mayer and colleagues have conducted several experiments that have shown a learning benefit to generative activities in domains involving invisible components, including electric circuits (Johnson & Mayer, 2010), lightning formation (Johnson & Mayer, 2009), and the chemistry of detergents (Schwamborn et al., 2010). Hausmann & Vanlehn (2007) addressed the possibility that generating explanations is beneficial because learners merely spend more time with the content material. In their study in the domain of physics, provided

explanations were not as effective as generated explanations.

Learner-generated Explanations in Visual and Verbal Formats

The cognitive processes underlying the development of understanding may differ for visual and verbal explanations. Language has words for some parts, configurations, actions, and causes, but complex and complete descriptions of spatial and dynamic systems can be difficult to produce. Visualizations can readily depict the parts, shape, and configuration of a system, but it may be more difficult to depict the operation of a system, its functionality, and its causal mechanisms. Of course, the configuration provides clues for the system's operation and causality, and visual information can be supplemented with non-depictive diagrammatic devices, notably arrows (Heiser & Tversky, 2006; Tversky et al., 2000, Tversky, 2002, 2011). Importantly, visual explanations demand completeness. Like other types of models, all of the essential parts of a system need to be represented in the proper configuration for it to work. In this way, drawings provide a visual check for completeness that verbal descriptions do not require. Inferences can then be made from diagrams that preserve and map the parts and configuration of the represented system or process. In an experiment that asked students to take notes while reading a text that they could later use to answer questions about the text, many students used only language, but those who made diagrams performed better (Schneider et al., 2010). Furthermore, requiring diagrams benefited all students.

Some researchers have demonstrated visual explanations' superiority over written explanations. Gobert & Clement (1999) investigated the effectiveness of student-generated diagrams versus student-generated summaries on understanding plate tectonics after reading an expository text. Students who generated diagrams scored significantly higher on a post-test measuring spatial and causal/dynamic content, even though the diagrams contained less domain-related information. Hall, Bailey, & Tillman (1997) showed that learners who generated their own illustrations from text performed equally as well as learners provided with text and illustrations. Both groups outperformed learners only provided with text. In a study concerning the law of conservation of energy, participants who generated drawings scored higher on a post-test than participants who wrote their own narrative of the process (Edens & Potter, 2003). In addition, the quality and number of concept units present in the drawing/science log correlated with

performance on the post-test. Van Meter (2001) found that drawing while reading a text about Newton's Laws was more effective than answering prompts in writing. Finally, Witherspoon et al. (2007) showed that generating external representations while studying the circulatory system increased scores compared to re-reading the provided text.

The Role of Spatial Ability in Learner-generated Explanations

Developing an ability to visually manipulate a model of scientific processes is complicated. In constructing a visual representation of a scientific process, people may need to first imagine actions. Kozhevnikov, Hegarty, & Mayer (2002) found that low spatial ability participants interpreted graphs as pictures, whereas high spatial ability were able to construct more schematic images and manipulate them spatially. Hegarty & Just (1993) found that the ability to mentally animate mechanical systems correlated with spatial ability, but not verbal ability. In their study, low spatial ability participants made more errors in movement verification tasks. However, Leutner, Leopold, & Sumfleth (2009) found no effect of spatial ability on the effectiveness of drawing compared to mentally imagining text content.

Experiment 1: Explaining the Function of a Bicycle Tire Pump

Method

Participants Participants were 127 7th and 8th grade students, ages 12-14, enrolled in an independent school in New York City. Of the 127 students, 59 were females, and 68 were males.

Materials Each participant was given a 12-inch Spalding bicycle tire pump, a blank 8.5 x 11 sheet of paper, a 16 question post-test, and the Vandenberg-Kuse Mental Rotation Test (MRT). Half of the participants received instructions to create a verbal explanation in writing; the other half received instructions to create a visual explanation in a drawing.

Procedure On the first of two non-consecutive school days, participants completed the MRT as a whole-class activity. Participants were read aloud the instructions, and were given untimed practice on several items. They were then given three minutes to complete items 1-10, and an additional three minutes to complete items 11-20. On the second day, participants were given the pump and instructions to try to understand how it worked. This segment was untimed. The next set of instructions asked students to verbally explain how the pump worked (in words) or to visually explain how the pump worked (in a drawing). Upon completion of the explanation, participants were given the 16 question post-test.

Coding

Coding for Structure and Function. A maximum score of twelve points was awarded for the inclusion and labeling of six structural components: chamber, piston, inlet valve, outlet valve, handle, and hose. Information was coded as functional if it depicted or described the function/movement of an individual part, or the way multiple parts interact. There was no maximum imposed on the number of functional units.

Coding of Essential Features. Explanations were also coded for the inclusion of information essential to its function according to a four-point scale (adapted from Hall, Bailey, & Tillman, 1997). One point was given if both the inlet and the outlet valve were clearly present in the drawing or described in writing, one point was given if the piston inserted into the chamber was shown or described to be airtight, and one point was given for each valve if they were shown or described to be opening/closing in the correct direction. The maximum score for essential features was four points.

Coding of Invisible Features. The presence of three invisible features (the inlet valve, the outlet valve, and the movement of air) were coded separately, with one point given for the presence of each valve, and three points given for movement of air (entering, moving through, and exiting the pump). The maximum score for invisible features was thus five points.

Coding Visual Elements: Arrows and Multiple Steps. Arrows were coded for three purposes: label for a part or action, to show motion, or to indicate sequence. Each use of arrows was coded for one of these purposes and a score tallied for each use. The use of multiple steps/frames was used to show starting and ending positions, and change in location of parts of the pump and air.

Results

Spatial ability. Participants scores' on the MRT were used to divide participants into low and high spatial ability groups based on a median split in the data. Scores on the MRT range from 0-20; the mean score for participants was 10.56, and the median was 11. Scores were significantly higher for males ($M = 13.5$, $SD = 4.4$) than for females ($M = 8.8$, $SD = 4.5$), $F(1, 126) = 19.07$, $p < .01$.

Structure and Function. Both visual and verbal explanations contained from two to ten structural components. Visual explanations contained a significantly greater number of structural components ($M = 6.05$, $SD = 2.76$) than verbal components ($M = 4.27$, $SD = 1.54$), $F(1, 126) = 20.53$, $p < .05$, while there was no difference in the number of expressed functional components between visual and verbal explanations.

Essential Features. Scores for the inclusion of essential information were significantly higher for visual explanations ($M = 1.78$, $SD = 1.0$) than for verbal explanations ($M = 1.20$, $SD = 1.21$), $F(1, 126) = 7.63$, $p < .05$. No significant differences were found between low ($M = 1.34$, $SD = 1.04$) and high spatial participants ($M =$

1.45, $SD = 1.2$). Essential features were also found to positively correlate with delayed post-test scores, $r = .197$, $p < .05$.

Invisible Features. Scores for the inlet valve were higher for visual explanations ($M = .67$, $SD = .45$) than verbal explanations ($M = .51$, $SD = .5$), however the effect was only marginally significant, $F(1, 126) = 3.13$, $p = .07$. Scores for air movement also showed a marginally significant difference, $F(1, 126) = 2.93$, $p = .09$, with visual explanations ($M = 2.35$, $SD = 1.28$) containing a greater number than verbal explanations ($M = 1.88$, $SD = 1.45$). No significant differences between visual ($M = .92$, $SD = .43$) and verbal explanations ($M = .79$, $SD = .65$) were found for the outlet valve. Analysis of the invisible parts between low and high spatial participants also failed to show any significant differences in the inclusion of the inlet valve, the outlet valve, or air movement. Finally a total score for the inclusion of invisible parts was calculated for each participant by totaling the scores for the inlet valve, the outlet valve, and for air movement. The mean score was 3.26, $SD = 1.25$. The data was analyzed using linear regression, and revealed that the total score for invisible parts significantly predicted scores on the post-test, $F(1, 118) = 3.80$, $p = .05$.

Multiple Steps. The number of steps used by participants ranged from one to six. Participants whose explanations contained more than a single step scored significantly higher ($M = .76$, $SD = .18$) on the post-test than participants whose explanations consisted of a single step ($M = .67$, $SD = .19$), $F(1, 126) = 5.02$, $p < .05$.

Learning Outcomes. Scores on the post-test by group and spatial ability are shown in Figure 1. A test of the overall interaction between group and spatial ability was significant, $F(1, 124) = 4.094$, $p < .01$. In particular, low spatial participants who generated verbal explanations had significantly lower scores ($M = .609$, $SD = .145$) than low spatial participants who drew explanations ($M = .716$, $SD = .121$). Analyzing structure and function questions separately on the post-test found no differences in performance between low and high spatial participants on structural questions. However, analyzing performance on functional questions found a significant effect: low spatial participants who generated verbal explanations ($M = .502$, $SD = .194$) scored significantly lower than low spatial participants that drew ($M = .678$, $M = .122$), see Figure 2.

Discussion

The results of Experiment 1 show that low spatial ability participants were able to learn as successfully as high spatial ability participants when they first generated an explanation in a visual format. Importantly, this result was particularly strong for functional understanding. Visual explanations were more likely to contain certain invisible features of the pump, such as the valves. Including the inlet valve and attempting to explain its function is crucial because then it is performing its function it is inside the chamber and air entering or exiting cannot be felt by the user.

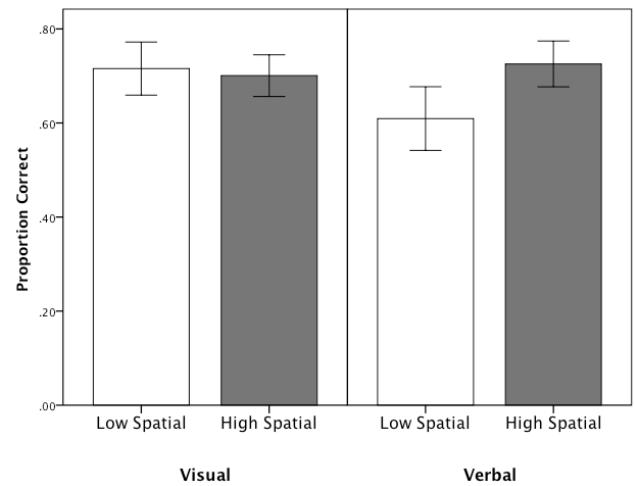


Figure 1: Scores on the post-test, by group and spatial ability.

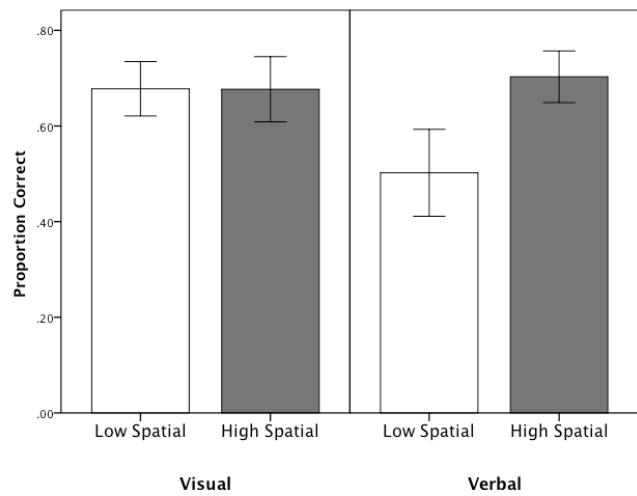


Figure 2: Scores on functional questions on the post-test, by group and spatial ability.

As mentioned previously, drawing encourages completeness. They force learners to decide on the size, shape, and location of parts/objects, and how the parts are related. Understanding the “hidden” function of the invisible parts is key to understanding the function of the entire system and requires an understanding of how both the visible and invisible parts interact. The visual format may have been able to elicit components and concepts that are invisible and difficult to integrate into the formation of a mental model.

Finally, an analysis of the visual explanations revealed that 67% also added written components to accompany their explanation. Arguably, some types of information may be difficult to depict visually, and our verbal language has many possibilities that allow for specificity. Indeed, several studies by Mayer and colleagues have found that

understanding a system is enhanced when text and pictures are presented simultaneously to learners (e.g. Mayer & Gallini, 1990).

The utility of visual explanations may differ for scientific phenomena that are more abstract, or contain elements that are invisible due to their scale. ‘

Experiment 2: Explaining the Process of Chemical Bonding

Method

Participants Participants were 126 8th grade students, ages 13-14, enrolled in an independent school in New York City. Of the 126 students, 58 were females, and 68 were males.

Materials Each participant was given an immediate post-test, a delayed post-test, a blank 8.5 x 11 sheet of paper, and the Vandenberg-Kuse Mental Rotation Test (MRT). Half of the participants received instructions to create a verbal explanation in writing; the other half received instructions to create a visual explanation in a drawing. In addition, the experimenter showed all participants a pre-recorded video lesson on bonding (13 minutes, 22 seconds long). The video began with a brief review of atoms and their structure, and introduced the idea that atoms combine to form molecules. Next, the lesson discussed how location in the periodic table affects behavior and reactivity of atoms, and makes atoms more or less likely to gain, lose, or share electrons. Examples of atoms, their valence shell structure, stability, charges, transfer and sharing of electrons, and the formation of ionic, covalent, and polar covalent bonds were discussed. The immediate post-test and delayed post-test each consisted of seven multiple-choice items and three free-response items.

Procedure On the first of three non-consecutive school days, participants completed the MRT as a whole-class activity, following the same procedures as Experiment 1. On the second day, participants viewed the recorded lesson on chemical bonding. They were instructed to pay close attention to the material but were not allowed to take notes on material presented in the video. Immediately following the video, participants were administered the immediate post-test of chemical bonding knowledge. Participants were given twenty minutes to complete the test; all participants finished within this time frame. On the third day, the participants were randomly assigned to either the visual or verbal condition. The next set of instructions asked students to either visually or verbally explain how atoms bond and how ionic and covalent bonds differ). Upon completion of the explanation, participants were given the delayed post-test.

Coding

Coding for Structure and Function. Visual and verbal explanations were coded for structural and functional

components. Table 1 and Table 2 show the components that were coded for structure and function, respectively.

Table 1: Coding Guide for Structure

Structural Components (1 pt. each)
Atoms with the correct number of electrons/valence electrons
Atoms with the correct charges (magnitude, positive/negative)
Bond between appropriate elements (i.e. between non-metals for covalent molecules and between a metal and a non-metal for ionic molecules)
Ionic bonds depicted/described as crystalline structure
Covalent bonds depicted/described as individual molecules

Table 2: Coding Guide for Function

Functional Components (1 pt. each)
Transfer of electrons in ionic bonds
Sharing between atoms in covalent bonds
Attraction between ions of opposite charges
Outcome of bonding shows atoms with stable valence electron shell configurations.
Outcome of bonding shows molecules with overall neutral charge

Coding system for arrows. Arrows were present in 92% of visual explanations. Their use was categorized into the use of arrows as labels and to show movement/action. Each use was tallied for each explanation.

Coding system for the use of specific examples. Explanations were coded for the use of specific atoms, such as NaCl to illustrate ionic bonding.

Coding for the use of multiple representations. Explanations were coded as symbolic (e.g. NaCl), atomic (showing structure of atom(s), and macroscopic (visible).

Results

Spatial ability. As in Experiment 1, participants' scores on the MRT were used to divide participants into low and high spatial ability groups based on a median split in the data. Scores were significantly higher for males ($M = 12.5$, $SD = 4.8$) than for females ($M = 8.0$, $SD = 4.0$), $F(1, 125) = 24.49$, $p < .01$.

Structure and Function. The maximum score for structural and functional information was five points. Visual explanations contained a significantly greater number of structural components ($M = 2.81$, $SD = 1.56$) than verbal

components ($M = 1.30$, $SD = 1.54$), $F(1, 125) = 13.69$, $p < .05$, while there was no difference in the number of expressed functional components between visual and verbal explanations. Structural information was more likely to be depicted in pictures ($M = 3.38$, $SD = 1.49$) than described in words ($M = .429$, $SD = 1.03$), $F(1, 62) = 21.49$, $p < .05$, but functional information was equally likely to be expressed in pictures ($M = 1.86$, $SD = 1.10$) and words ($M = 1.71$, $SD = 1.87$). Functional information in words added to visual explanations significantly predicted scores on the post-test, $F(1, 62) = 21.603$, $p < .01$. As in Experiment 1, there were no significant differences in the amount of structural information contained in explanations created by low and high spatial ability participants. However, explanations created by high spatial participants contained significantly more functional components, $F(1, 125) = 7.13$, $p < .05$.

Arrows. 83% of visual explanations contained arrows. The use of arrows was positively correlated with scores on the post-test, $r = .293$, $p < .05$.

Specific examples. High spatial participants ($M = 1.6$, $SD = .69$) used specific examples in their explanations more often than low spatial participants ($M = 1.07$, $SD = .79$). The difference was marginally significant $F(1, 125) = 3.65$, $p = .06$. There were no significant differences in the use of specific examples between visual and verbal groups. The inclusion of a specific example was positively correlated with scores on the delayed post-test, $r = .555$, $p < .05$.

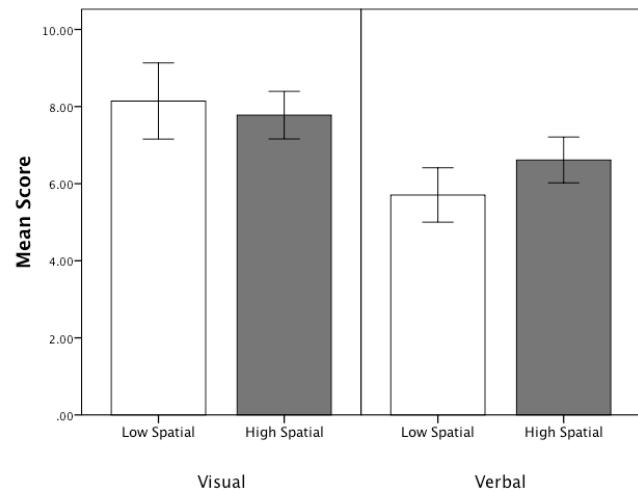
Multiple representations. Multiple representations were included in 65% of the explanations. Participants generated significantly more when creating visual explanations ($M = 1.79$, $SD = 1.20$) compared to verbal explanations ($M = 1.33$, $SD = .48$), $F(1, 125) = 6.03$, $p < .05$. However, the use of multiple representations did not significantly correlate with delayed post-test scores.

Learning outcomes. The immediate post-test was scored so that the maximum score was ten points. Each of the seven multiple choice questions and three free-response questions was given one point for the correct answer. The mean score (defined by proportion correct) on the immediate post-test was .463, $SD = .469$. Scores did not differ significantly between participants in the visual group ($M = .486$, $SD = .308$) and the verbal group ($M = .443$, $SD = .260$), $F(1, 125) = .740$, $p > .05$. Scores between high spatial ($M = .532$, $SD = .421$) and low spatial participants ($M = .402$, $SD = .390$) also did not differ significantly, $F(1, 125) = 2.72$, $p > .05$.

The mean score on the delayed post-test (after participants generated explanations) was .704, $SD = .299$. Participants in the visual group improved significantly from the immediate post-test ($M = .822$, $SD = .208$), $F(1, 125) = 51.24$, $p < .01$, Cohen's $d = 1.27$. Participants in the verbal group also showed significant increases from the immediate post-test ($M = .631$, $SD = .273$), $F(1, 125) = 15.796$, $p < .05$, Cohen's $d = .71$.

A comparison of the delayed post-test scores between groups found significant differences. Figure 3 shows scores on the post-test by group and spatial ability. Participants

generating visual explanations ($M = .822$, $SD = .208$) scored higher on the post-test than participants generating verbal explanations ($M = .631$, $SD = .273$), $F(1, 125) = 19.707$, $p < .01$, Cohen's $d = .88$. In addition, high spatial participants ($M = .824$, $SD = .273$) scored significantly higher than low spatial participants ($M = .636$, $SD = .207$), $F(1, 125) = 19.94$, $p < .01$, Cohen's $d = .87$ (Figure 4-12). The results of the test of the interaction between group and spatial ability was not significant. A separate analysis comparing performance on multiple choice questions and free response questions did not show any differences between visual and verbal groups or between low and high spatial ability groups.



Discussion

The results of Experiment 2 supported those of Experiment 1: learner-generated visual explanations provided an advantage over learner-generated verbal explanations. Visual explanations resulted in higher scores on the post-test for both low spatial and high spatial participants.

No difference was found between low and high spatial participants in the amount of structural information contained in the explanations, but high spatial participants included more function, were more likely to use specific examples, and scored higher on the delayed post-test.

An interesting finding of Experiment 2 was that the use of arrows significantly correlated with scores on the delayed post-test. How does their use lead to greater understanding? Arrows were most often used to label structure, or to label an action. They were also used to differentiate an initial versus and ending state, to show change. Previous research has shown arrows to serve a number of purposes. Notably, studies have shown the addition of arrows able to convey functional information in a structural diagram (Heiser & Tversky, 2006). While the purpose of this study was to examine student-generated explanations, these results support those of previous work that shows when arrows are used in diagrams in a way that encourages the development of mental models, they become more effective.

Conclusion

Experiment 1 showed that learning about a physical bike pump and generating visual explanations was primarily a benefit to low-spatial ability participants. The measures of learning (from a true/false post-test) were of course limited, and it is possible that higher-order learning by high spatial ability participants was not revealed. Experiment 2 showed that viewing a class lesson on chemical bonding and generating visual explanations benefited both low and high spatial ability students, although to different degrees (high spatial ability participants scored significantly higher on the post-test). In the generation of visual explanations, learners use the information they gather from new material to create internal representations that become richer with the integration of verbal and non-verbal representations, forming a mental model that then informs and direct the creation of visual explanations. Learners with high spatial ability are more adept at forming and manipulating mental images; this may make the generation of visual explanations easier for them. Learners with low spatial ability may find the task difficult, but may be able to be more successful with generating visual explanations if support is provided.

Together, the results from the two experiments support the use of learner-generated visual explanations as a learning strategy in science. Future studies should explore how this strategy mediates the comprehension of concepts presented in physical models, experiments, and textbooks, and posttest performance. Students live in a macroscopic world, where objects have mass and occupy space. Understanding “invisible” processes in science, then, presents a challenge. Generating visual explanations through drawing is likely an underused method of monitoring and supporting students’ understanding of scientific concepts.

References

Gobert, J.D., & Clement, J.J. (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.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology*, 89(4), 677-681.

Hausmann, R. G. M., & Vanlehn, K. (2007). Explaining self-explaining: A contrast between content and generation. In Luckin, R., Koedinger, K. R., & Greer, J. (Eds). *Artificial intelligence in education: Building technology rich learning contexts that work*, 158 (pp. 417-424). Amsterdam: IOS Press.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717-742.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, 30, 581-592.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101, 621-629.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior*, 26, 1246-1252.

Kozhevnikov, M., Hegarty, M., & Mayer (2002). Spatial abilities in kinematics problem solving. In M. Anderson, B. Meyer & P. Olivier (Eds). *Diagrammatic Representation and Reasoning* (pp. 155-171). Berlin: Springer-Verlag.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, 25, 284-289.

Mayer, R.E. & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82, 715-726.

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, 872-879.

Tversky, B., Zacks, J., Lee, P. U., & Heiser, J. (2000). Lines, blobs, crosses, and arrows: Diagrammatic communication with schematic figures. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Theory and application of diagrams*, (pp. 221-230). Berlin, Germany: Springer.

Tversky, B. (2002). What do Sketches Say about Thinking? *2002 AAAI Spring Symposium, Sketch Understanding Workshop*, Stanford University, AAAI Technical Report SS 02-08, March, 2002, pp. 148-151.

Tversky, B. (2011). Visualizing Thought. *Topics in Cognitive Science*, 3, 499-535.