Unlocking potential: Individual differences in the use of concurrent scientific visualizations

Scott R. Hinze (s-hinze@northwestern.edu)
David N. Rapp (rapp@northwestern.edu)
School of Education and Social Policy
Department of Psychology, Northwestern University
Evanston, IL 60208, USA

Vickie M. Williamson (Williamson@tamu.edu)
Department of Chemistry, Texas A&M University
College Station, TX 77843, USA

Mary Shultz (mary.shultz@tufts.edu
Department of Chemistry, Tufts University
Medford, MA, 02155, USA

Ken C. Williamson (kcwilli@tamu.edu)
Department of Construction Science, Texas A&M University
College Station, TX 77843, USA

Ghislain Deslongchamps (ghislain@unb.ca)
Department of Chemistry, University of New Brunswick
Fredericton, New Brunswick, E3B 5A3, Canada

Abstract

The goal of this study was to identify whether individual differences among students might influence the effectiveness of visualizations; in particular, concurrently presented alternative visualizations of chemical molecules. Thirty beginning organic chemistry students of varying prior knowledge completed: (1) a battery of tests measuring reasoning ability, spatial ability, and need for cognition and (2) an eye-tracking session, in which they viewed both ball-and-stick and potential plot representations, and answered interpretation and application questions. Eye movement patterns indicated that students tended to initially rely on the already familiar ball-and-stick representations, especially for more difficult application questions. As the task unfolded, though, students' choice of which representation to refer to was moderated by prior expertise and general reasoning ability.

Keywords: Education, Multimedia, Eye Tracking, Prior Knowledge, Expert vs. Novice Comprehension

Introduction

The goal of this study was to identify whether individual differences among students might influence their reliance on scientific visualizations. Previous work has focused on the notion that visualizations are a useful and generally effective means of enhancing comprehension of scientific concepts, principles, and hypotheses (van Gog & Scheiter, 2010). However, the utility of such visualizations might be mediated by individual differences in cognitive abilities, functions, and propensities. To test this possibility, we specifically focused on two types of chemistry visualizations that are popularly used in both science textbooks and classrooms; ball-and-stick representations of molecules, and potential plot representations of molecules. (We looked at these representations for alcohols, carboxylic acids, and hydroxycarboxylic acids.) Our examination of students' reliance on and comprehension of chemistry concepts with these representations employed eye tracking, response accuracy and explanation data. This multi-method approach afforded the means for identifying which representations the students used, or didn't use, to reason about molecules.

Electrostatic potential (ESP) plots are color-coded surface maps that visually encode charge distribution in molecules. ESP maps can be informative since many chemical interactions are governed by the principle that positive and negative charges attract each other. These visualizations are widely used for computer-aided drug design and materials development, so student familiarity with ESP plots carries benefits beyond immediate learning and testing issues. The more familiar ball-and-stick models have been used in chemistry classrooms for decades, and visual depictions of those models are similarly popular. These models depict the atoms and bonds of a molecule, but do not provide a visualization of the distribution of electrons, like in the ESP plots. Neither ESP maps nor Ball-and-Stick graphics are necessarily a superior visualization. Rather, their utility depends on task demands. Ball-and-Stick plots may be most useful for reasoning about molecular structure, while ESP maps are most useful for reasoning about relative atomic charges. Here, we were interested in whether students would be willing or able to use unfamiliar ESP maps for questions specifically designed to be answered based on relative charges.

Although we were interested in determining whether students differentially rely on these different representations, ESP plots on their own present interesting fodder for investigation. Chemical educators have advocated the use of ESP maps in both general and organic chemistry (Shusterman & Shusterman, 1997), and to our knowledge, only a few studies have shown potential benefits of such plots for STEM understanding (e.g., Sander & Badger, 2001), and those studies have often confounded plot presentation with additional classroom activities. It remains unclear whether and how these novel plots are utilized by students, or whether different students differentially rely on and benefit from these visualizations.

The effective use of visualizations is not guaranteed nor is it a trivial activity (e.g., Ainsworth, 2006), so it is important to know whether students have the representational competence necessary to utilize both readily familiar and new visualizations. Thus, the purpose of this experiment was to examine whether students actually use different representations to answer exam-type chemistry questions. In addition, because visualization use and effectiveness is potentially moderated by individual differences (e.g. Canham & Hegarty, 2010; Kalyuga, Ayres, Chandler & Sweller, 2003) we were also interested in whether prior knowledge, reasoning or spatial abilities was associated with effective or ineffective use and learning from the visualizations.

Method

Participants

Participants were recruited from a large organic chemistry course (N = 225). While these students had access to ESP maps in textbooks, no formal instruction on the use of these maps had occurred in class prior to the experiment. High and low prior knowledge participants were recruited from this sample based on their scores on a pre-test. A total of 30 students, (18 high prior knowledge; 12 low prior knowledge) volunteered and completed all aspects of the experiment and were compensated $20 for their time. The 30 participants consisted of 19 female and 11 male students; the age ranged between 19-21 years (M = 19.47, SD = .68). Most (N = 27) identified themselves as neither a chemistry major or minor.

Materials and Apparatus

Pre-test. A 19 question general chemistry knowledge quiz, created for this project, asked students to answer questions regarding relevant chemistry concepts, including the definition of electronegativity, the identification of atoms in a ball-and-stick representation, and the interpretation of ESP plots. This pre-test served as the basis for recruiting high knowledge (top 1/3 of scores) and low knowledge (bottom 1/3 of scores) participants.

Eye Tracker. For some parts of the experiment, a Tobii T60 eye tracker was used to track eye movements while participants completed the problem set (see below).

Instructional Materials. Participants were introduced to ESP maps with a 391 word multimedia text, presented using Tobii Studio software. The text explained key concepts such as the attraction of opposite charges and the role of electronegativity differences in facilitating this process. This concept was made more concrete by the use of two examples: the ionic bond between Na^+ and Cl^- based on their widely different electronegativity, and the less extreme case of the bonds in H_20. Two ESP maps were presented to show positive (blue), negative (red) and neutral (green) charges.

Problem Set. The main task of this experiment asked participants to use the ball-and-stick and/or ESP map representations of molecules to answer questions about molecules. Examples of the questions and representations are provided in Figure 1; all problems consisted of a question displayed at the top of the screen along with simultaneous presentation of both an ESP map and Ball-and-Stick representation of a molecule. There were a total of six different molecules depicting alcohols, carboxylic acids, and hydroxycarboxylic acids, half of which presented the ESP map on the left side of the screen and half which presented the ESP map on the right side of the screen. Each participant answered the four questions shown for each of the six molecules, for a total of 24 items, with the order of presentation of the items randomized. Three answer choices (one correct) numbered 1-3, were presented on both representations. All questions were designed to be best answered using ESP maps, since the maps provide a visual cue as to the location of relative charges in a molecule, information that was relevant to all questions. Two of the questions for each molecule, as shown in the top two panels of Figure 1, required participants to identify an atom which met certain basic criteria (i.e. greatest positive charge, highest electron density); we refer to these as *Identification* questions. The two other questions for each molecule, as exemplified in the bottom two panels of Figure 1, required participants to determine where a positive or negative charge would be attracted in the molecule. This necessitated not only identifying the basic characteristics of the atoms, but also the application of this information in the context of a chemical interaction; we refer to these as *Application* questions.

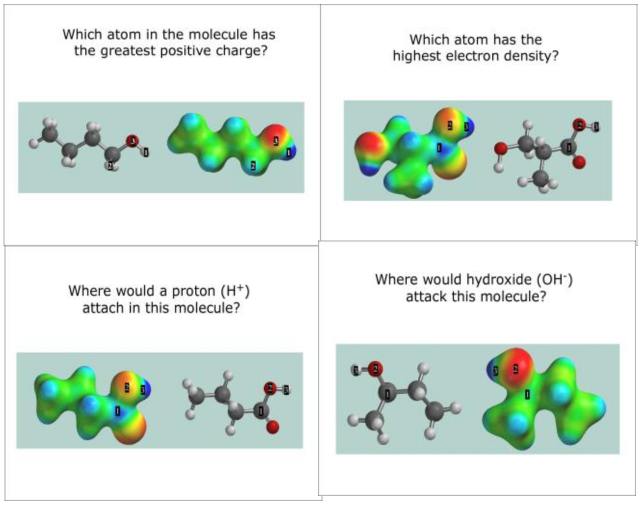

Figure 1: Example problems. The top two panels represent Interpretation questions, while the bottom two panels represent Application questions.

Baseline Presentations of Molecules. Prior to completion of the problem sets, participants freely viewed the two representations of all six molecules for five seconds each. This baseline presentation served to make the participants generally aware of the molecules they would be working with and the format of the representations. It also served as a check for whether participants' attention would be generally drawn toward one representation over the other (e.g. participants could simply be more intrigued by the colorful ESP plots). Such a baseline recording was necessary to attribute eye-movement patterns to the attempts at answering the questions, not simply to interest in one representation over the other.

Individual Differences Measures. Participants completed a battery of tasks presented via computer to assess their reasoning abilities, spatial abilities, and need for cognition. The reasoning abilities task was an electronic version of the Test of Logical Thinking (TOLT; Tobin & Capie, 1981), a 10-question test containing a variety of reasoning problems, which is generally positively correlated with successful chemistry performance (Bunce & Hutchinson, 1993).

Spatial abilities were assessed using measures intended to tap a variety of spatial components. All tests were conducted using computers, some of which were adapted from paper versions. Two tests were used to measure the speed of mental rotation including the Card Rotation Test (CRT; Ekstrom, French, Harman, & Derman, 1976) and the Mental Rotation Test (MRT; Vandenberg & Kuse, 1978). Complex visualization abilities were measured with the Purdue Visualization of Rotations Test (ROT; Bodner & Guay, 1997) and Guay's Visualization of Viewpoints (GVVT; Guay & McDaniels, 1976). Gestalt identification ability was assessed using the Hidden Patterns Test (HPT; Ekstrom et al., 1976). Participants also completed the Need for Cognition scale (NFC; Cacioppo, Petty, & Kao, 1984).

Eye Tracking Metrics

To examine eye movement patterns of participants during their completion of the problem sets, we concentrated mainly on the pattern of fixations, which occur when the eye stops for a period of time in a relatively stable spatial location. We were mainly interested in fixations to the areas containing either the ESP map or the Ball-and-Stick representation for each item. These Areas of Interest (AOIs) were similar in size (~20% of the area of the screen was taken up by each AOI) and the size of the AOIs was held constant across all trials. We calculated measures related to the number and length of fixations within an AOI, the number and length of visits to each AOI, and the total amount of time spent within an AOI for a given trial.

Procedure

The assessment of prior knowledge and the TOLT measure were both completed prior to participation in the experiment as part of the organic chemistry course. All other activities took place in two counterbalanced laboratory sessions with each participant's second session occurring a week after the first. Informed consent was provided at the beginning of the first session, and payment was provided upon completion of the second session.

Eye tracking session. Each participant's eye tracking session presented the instructional materials on ESP maps and the problem set. A monitor containing the eye tracking mechanism presented all tasks. After calibration, participants read through the instructional materials at their own pace. Next, participants examined the six molecules in the baseline presentation. Each molecule was presented for 5 seconds and the six molecules were presented in the same order for all participants. Next, participants answered the 24 items in the problem set, in random order. After providing an answer to each question both by keypress and verbally (recorded by an experimenter and by a microphone) participants rated their confidence and provided an oral explanation of why they chose their answer.

Individual differences session. This session was completed in small groups in a private computer lab. Participants completed the NFC followed by all spatial abilities tasks in random order.

Results

Accuracy

Response accuracy was aggregated across molecules (alcohol, carboxylic acid, etc.), to obtain a mean for each type of question (Interpretation, Application) across twelve items. This within-subject variable of question type was crossed with the between-subject variable of prior knowledge. We predicted: 1) Application questions would be more difficult than Interpretation questions; 2) Participants with low prior knowledge would make more errors than participants with high prior knowledge; 3) The effect of prior knowledge would be strongest for the more difficult Application questions.

Figure 2 displays the means which were subject to a 2 X 2 mixed factors ANOVA. As expected, Application questions were more difficult than Interpretation questions (F(1, 28) = 10.71, p = .003, η_p^2= .28), and overall participants with low prior knowledge performed worse than participants with high prior knowledge, F(1, 28) = 12.11, p = .002, η_p^2= .30. These main effects were moderated by the predicted interaction (F(1, 28) = 4.06, p = .05, η_p^2= .13); the difference between prior knowledge groups was larger for Application (.13, t(28) = 2.09, p = .05, d = 1.11) than for Interpretation questions (.09, t(28) = 3.25, p = .003, d = .69).

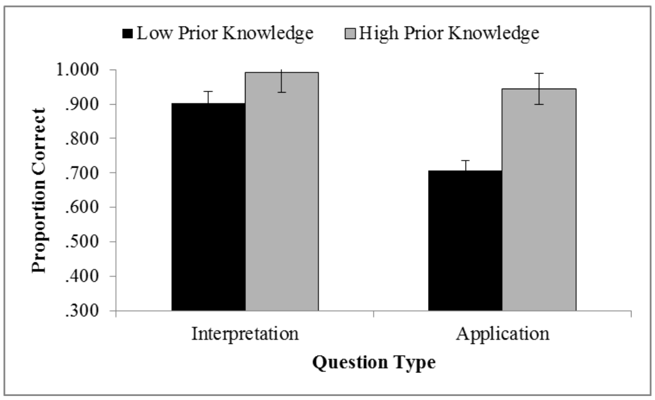

Figure 2: Mean response accuracy for prior knowledge groups based on type of question. Error bars represent 1 SEM.

Participants categorized as high prior knowledge demonstrated better performance on the problem set in general, and the Application questions were more challenging, especially for participants with low prior knowledge.

Cognitive Abilities and Accuracy

We correlated NFC and the cognitive abilities measures with response accuracy for Interpretation and Application questions. Possibly because of near-ceiling accuracy, no individual difference measure correlated significantly with performance on Interpretation questions (N = 30, all r's < .14). For Application questions, participants with higher scores on the gestalt spatial Hidden Patterns Test, were generally more accurate (r = .37, p = .05), and a trend suggested participants with higher reasoning abilities, as measured by the TOLT, were more accurate as well (r = .31, p = .10). Interestingly, the HPT and TOLT were highly interrelated (r = .62, < .001) indicating that these measures may have been tapping similar constructs. There were no significant relationships between Application accuracy and any other spatial measures (r's < .18) or NFC (r = -.04).

Eye Movements and Use of Representations

In the following analyses, we present data based on the total fixation time in the AOIs associated with Ball-and-Stick and ESP plots. While we also calculated a variety of other metrics, the pattern of results was similar for all metrics, so the total fixation duration metric will be used for simplicity.

Baseline visualizations. These screens without question prompts serve to determine whether high or low knowledge participants were drawn to a certain type of representation naturally. Participants spent roughly equivalent amounts of time focusing on Ball-and-Stick (M = 2.37 s, SD = .35) and ESP representations, M = 2.30 s, SD = .39, F(1, 28) = .60. This was not moderated by prior knowledge (F = .23), indicating that neither high nor low knowledge participants' visual attention was drawn to one visualization type over the other without question prompts.

Problem set. Since questions regarding electrostatic potential or the attraction of positive or negative ions could be easily answered using the color-coded ESP maps, we expected that more attention to these plots would be related to success on the tasks. Similarly, since the Ball-and-Stick representations are likely to be more familiar, we expected that participants who struggle to understand the relatively new ESP plots would rely more heavily on the Ball-and-Stick representations. Thus, we predicted that participants with low prior knowledge, and participants who did not accurately answer the questions, would fixate longer on the Ball-and-Stick representation than the ESP representation. This pattern of results may be especially prominent for the more difficult Application questions as compared to the Interpretation questions, as the Interpretation questions could be answered without a sophisticated understanding of the ESP maps.

These predictions were tested with a 2 (Visualization; Ball-and-Stick, ESP) X 2 (Question Type; Interpretation, Application) X 2 (Prior Knowledge; Low, High) mixed factors ANOVA with prior knowledge a between-subjects variable and Visualization and Question Type within-subjects variables. Figure 3 displays the mean results. Overall, participants with high prior knowledge spent less time inspecting the problems than participants with low prior knowledge as suggested by a marginal main effect of Prior Knowledge, F(1, 28) = 3.09, p = .09, η_p^2= .10. The main effects of Question Type (F(1, 28) = 9.43, p = .01, η_p^2= .25) and Visualization (F(1, 28) = 5.23, p = .03, η_p^2= .01) were significant. However, each of these main effects was qualified by significant 2-way interactions (all p's < .01). Most importantly, the 3-way interaction between Prior Knowledge, Visualization, and Question Type was significant, F(1, 28) = 5.99, p = .02, η_p^2= .18. As seen in the left panel of Figure 3, participants with low prior knowledge spent much more time on Ball-and-Stick Representations than ESP plots for the difficult Application questions (t(11) = 3.53, p = .01, d = 1.02), but did not exhibit the same over-reliance on the Ball-and-Stick representations for

Interpretation questions, $t < 1$. In contrast, as seen in the right panel of Figure 3, participants with high prior knowledge did not differentially rely on either representation for the difficult Application questions, $t(17) < 1$. Interestingly, high knowledge participants actually spent significantly more time inspecting the ESP plots than Ball-and-Stick representations for the Interpretation questions, $t(17) = 2.58$, $p = .02$, $d = .42$. In sum, more challenging application problems led students with low prior knowledge to rely on the more familiar Ball-and-Stick representations. In contrast, less difficult problems that could easily be answered by looking at the surface features of the ESP plots led high-knowledge participants to rely more on those plots than the Ball-and-Stick representations.

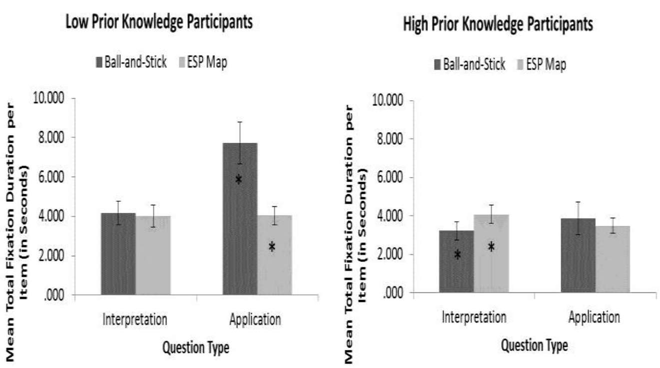

Figure 3: Mean total fixation duration on Ball-and-Stick vs. ESP maps based on Prior Knowledge and Question Type. Error bars represent 1 SEM.

Use of representations and cognitive abilities. To explore whether spatial abilities, reasoning abilities, or NFC influenced fixation time on Ball-and-Stick or ESP representations, we created a single continuous measure reflecting relative use of the two representations. We divided the total fixation duration on the Ball-and-Stick AOI by the sum of the total fixation duration on Ball-and-Stick and ESP AOIs (effectively controlling for time-on-task). High scores on this measure indicate more time looking at Ball-and-Stick than ESP representations.

Reasoning ability (as measured by the TOLT) correlated negatively with relative Ball-and-Stick use on Application questions ($r(30)= -.48$, $p = .01$), and showed a similar though not significant pattern for Interpretation questions ($r = .30$, $p = .11$.). That is, participants with higher TOLT scores tended to spend less time viewing the familiar Ball-and-Stick plots, and more time viewing the ESP plots, while answering the questions. All measures of spatial ability demonstrated small negative correlations with Ball-and-Stick use, but none reached significance (all p's > .24), though the effect of HPT was marginal ($r(30) = -.35$, $p = .06$). NFC also had no significant relationship with relative Ball-and-Stick use for application ($r = -.01$) or interpretation ($r = .02$) questions.

Since TOLT and prior knowledge were related to reliance on the representations, we checked whether high and low prior knowledge groups differed on TOLT scores. The high knowledge group had significantly higher scores on the TOLT $t(28) = 2.12$, $p = .04$, $d = .75$). Thus, part of the difference between high and low prior knowledge groups may be related to the cognitive abilities which underlie TOLT scores.

Use of representations over practice. Aggregating across question trials obscures the role that practice with the plots might have on use of the representations over the course of the experiment. Since the ESP plots were novel for these participants, students might only come to rely on those plots, rather than the more familiar Ball-and-Stick, after experience with them. We examined whether any developing familiarity might emerge differentially for participants with high and low prior knowledge.

We again used the relative use of representations measure described above as the DV in an ANOVA with Question Type (Interpretation, Application), Practice Trial (1-6), and Prior Knowledge (Low, High) as independent variables. As can be seen in the two panels of Figure 4, Practice appeared to be moderated by Question Type, $F(5, 140) = 2.37$, $p = .04$. Because of this, we analyzed the data for Interpretation and Application questions separately.

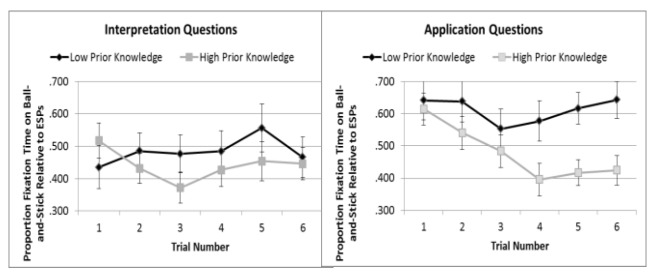

Figure 4: Mean relative use of Ball-and-Stick representation over practice trials based on question type and prior knowledge.

For Interpretation questions, there was no main effect of Practice ($F(5, 140) = .74$) and no main effect of Prior Knowledge ($F(1, 28) = .52$). With respect to the interaction between Practice and Prior Knowledge, there was a significant quadratic trend ($F(1, 28) = 4.47$, $p = .04$). Participants with low prior knowledge did not change their relative use of Ball-and-Stick representations with practice ($F(1, 11) = .91$), while participants with high prior knowledge showed a significant quadratic trend (i.e. a reduction and then leveling off of Ball-and-Stick use; $F(1, 18) = 4.71$, $p = .05$).

For Application questions, there was a significant effect of Practice ($F(5, 140) = 4.39$, $p = .001$) and a significant effect of Prior Knowledge, $F(1, 28) = 4.69$, $p = .04$. But, as can be seen in the right panel of Figure 4, the effect of Practice differed for high and low prior knowledge participants as confirmed by a linear interaction, $F(1, 28) = 9.12$, $p = .01$. Participants with low prior knowledge demonstrated no linear effect of Practice on Ball-and-Stick use (i.e. they continued to view Ball-and-Stick representations more than ESPs throughout the experiment, $F(1, 11) = .01$); in contrast, participants with high prior knowledge demonstrated a linear decline in relative Ball-and-Stick use with Practice, $F(1, 17) = 23.35$, $p < .001$. In sum, participants with high prior knowledge initially relied on the familiar Ball-and-Stick representation, just as participants with low prior knowledge did. However participants with high prior knowledge learned to utilize the ESP over the course of the experiment. Participants with low prior knowledge, however, continued to rely on the Ball-and-Stick representations more than ESP plots for these Application questions.

Discussion

Chemistry educators advocating the use of ESP maps as instructional tools have suggested that the models are easy to understand, allowing students to efficiently appraise relevant information about electron density (Shusterman & Shusterman, 1997). In some ways, our data support this contention. After only a short introduction, nearly all students could correctly identify positive charges and relative degrees of electron density. However, participants did not immediately utilize the ESP plots when asked about chemical interactions, relying instead on the more familiar ball-and-stick representations. Interestingly, while participants with higher prior knowledge (and reasoning abilities) began to effectively apply the ESP maps on application problems with practice, low prior knowledge participants (who also had lower reasoning abilities) maintained their use of ball-and-stick representations, and continued to answer incorrectly. A preliminary analysis of verbal explanations suggests that incorrect answers were often associated with unnecessarily complicated (and often faulty) reasoning about the atoms in the molecules, when the correct answer could be more simply inferred based on the information in the ESP maps. This suggests the importance of developing competence with these visualizations in reducing cognitive burdens and errors.

As with many experiences, individuals in the current experiment relied on what they are already familiar with. With practice, individuals with knowledge of chemistry and reasoning skills moved beyond those familiarities to employ less familiar visualizations. Clearly the effectiveness of visualizations, both in drawing attention and influencing learning, depends upon a host of individual differences; visualizations are not necessarily the panacea they have been made out to be, but they can help under particular circumstances. These results demonstrate that the processes and products of multimedia experiences emerge from the interactions of prior knowledge, cognitive abilities, and task demands, and that understandings of these interactions is imperative for the design of effective learning interventions.

Acknowledgments

This project was funding by REESE grant #0908130 from the National Science Foundation, David N. Rapp, Principal Investigator.

References

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

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator, 2,* 1-18.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155-166.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). Manual for kit of factor-referenced cognitive tests. *Educational Testing Service.* Princeton: New Jersey.

Guay, R., & McDaniels, R. E. (1976).The visualization of viewpoints. The Purdue Research Foundation. West Lafayette, IN.

Kalyuga, S., Ayres, P, Chandler, P, & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Sanger, M. J., & Badger, S. M. II. (2001). Using computer-based visualization strategies to improve students' understanding of molecular polarity and miscibility. *Journal of Chemical Education*, *78*, 1412-1416.

Shusterman, G. P., & Shusterman, A. J. (1997). Teaching chemistry with electron density models. *Journal of Chemical Education*, *74*, 771-776.

Vandenberg, S. G., & Kuse, A. R..(1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-601.

van Gog, T., & Scheiter, K. (Eds.). (2010). Eye tracking as a tool to study and enhance multimedia learning [Special issue]. *Learning and Instruction,20.*