Hypotheses:

1. Experts will make a higher percentage of meaningful transitions than novices.
2. Expert transitions will deviate further from what is expected by chance than novice transitions.
3. Expert transitions will deviate further from what is expected by random chance than novices.

METHODS

Participants and Design:

Chemistry professors (n=7) and students who had taken no chemistry or only general chemistry one (n=13), participated in this study. All participants possessed good vision and passed the eye-tracking calibrations. Participants were shown a series of fourteen slides each containing nine boxed images. The first two slides were insignificant and used to familiarize the students with the process. Each subsequent slide presented a different chemistry concept, increasing in difficulty as the slides progressed. The position of the images was randomly generated for each participant to reduce view bias. Each slide included images of macroscopic, submicroscopic, and symbolic concepts in chemistry, utilizing the levels of the Triplet Relationship. The slides contained conceptual groupings, defined as “chemical concepts not explicitly available in the representations,” and surface groupings, defined as “based only on the explicit features of the representations in the group” 1 Slide eleven provided no complete conceptual concept. Slides seven, nine, and ten provided one complete conceptual concept and three surface level concepts. The remaining slides provided three complete conceptual concepts and three surface level concepts.

Response Analysis:

Participants were instructed to select three images for each slide that represented a single grouping. They were then asked to provide a verbal response explaining their selection. The responses were then evaluated by the principle investigator and two external faculty members in coordination with the process described by Kozma and Russel’s study 1. Each grouping was examined along with the verbal response and deemed to be a conceptual, surface, or invalid selection. For example, if the participant selected a conceptual grouping, but gave an incorrect explanation of grouping, the selection would be deemed invalid.

Data Collection:

The GazePoint *whatever the model is* and the accompanying API was used to gather eye movement data for the study at *whatever the number is* Hz. For all slides, each of the images were mapped to an area of interest labeled one through nine. The data was then filtered to eliminate all saccades and fixations that occurred outside of an area of interest. If consecutive fixations occurred within the same area of interest, the fixations were combined with the summed duration. Each fixation was then given a surface, meaningful, or invalid transition tag depending on if the sequential fixations were within the predefined groupings. All transitions that occurred within the first two seconds were eliminated from the data to account for initial image recognition transition patterns (*source needed*)

Data Analysis:

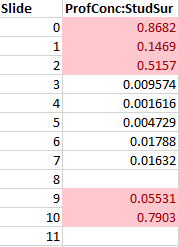
Several statistical analyses were performed on the data. Each analysis looked only at professors who selected a conceptual grouping and students who selected a surface grouping for each slide. Note that this eliminated slides that have no conceptual grouping as well as those where no professor selected a conceptual grouping. First, a t-test was performed comparing the percentage of meaningful transitions made by professors and students (for Hypothesis 1). A two-way ANOVA test with interactions was performed on each slide with a specific focus on the interaction effect of the percentage of meaningful and surface transitions for both the professors and the students. This test utilized the Tukey p-value adjustment method to reduce type one error (*source*). A lack of significant interaction indicates that the participant’s meaningful transitions do not affect the amount of surface transitions, but rather an increase of meaningful transitions will be linked to less invalid transitions (for Hypothesis 2). A Chi-squared analysis was performed on each slide to test the professor’s and student’s deviation from the transitions of what would be expected if the searching occurred by random chance. For example, if the slide contained three conceptual and three surface groupings it would be expected that 12.5% of the transitions would be meaningful, 12.5% would be surface level, and 75% would be invalid. It is expected that a participant who has a conceptual understanding of the images would deviate further from random than a participant that lacked the same understanding (for Hypothesis 3)

RESULTS

T-Test

To examine hypothesis one, a t-test was performed on the percentage of meaningful transitions made by professors who made conceptual selections and students who made surface selections. The results for these tests are shown in Table 1. A significant p-values indicates that that there is a difference in the percentage of meaningful transitions made by the professors and the students. Of the ten slides, five returned statistically significant p-values. (*probably want to add a look at effect size for this section)*

Table 1. Results of the t-test of percentage of meaningful transitions

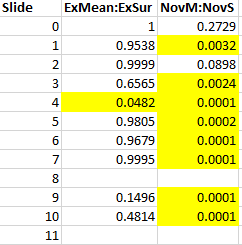


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Expert | Novice |  |  |  |
| Slide | M | M | df | t | p |
| 0 | 0.342 | 0.333 | 7.982 | 0.171 | 0.868 |
| 1 | 0.255 | 0.156 | 3.8614 | 1.811 | 0.147 |
| 2 | 0.260 | 0.224 | 10.674 | 0.672 | 0.516 |
| 3 | 0.375 | 0.173 | 3.727 | 5.072 | 0.0096 |
| 4 | 0.382 | 0.147 | 5.253 | 5.9505 | 0.0016 |
| 5 | 0.328 | 0.199 | 8.7848 | 3.7536 | 0.0047 |
| 6 | 0.284 | 0.169 | 14.891 | 2.661 | 0.018 |
| 7 | 0.168 | 0.0631 | 8.394 | 2.994 | 0.016 |
| 9 | 0.143 | 0.0754 | 8.226 | 2.231 | 0.055 |
| 10 | 0.0714 | 0.0469 | 1.149 | 0.332 | 0.790 |

ANOVA

Table 3 shows the p-values for each comparison of the interactions between novices and experts for the percentage of surface and the percentage of meaningful transitions. A significant p-value demonstrates that the factor is dependent on the level of the other factor, thus showing that the percentage of meaningful transitions affects the percentage of surface transitions. Only one slide resulted in a significant value for the experts, indicating that there is no correlation between the percentage of meaningful and surface level transitions that experts make. The novices displayed a strong dependency between surface and meaningful transitions with eight of the ten slides providing a significant result.

Table 2. Surface to Conceptual interactions for students and professors for two-way ANOVA test.

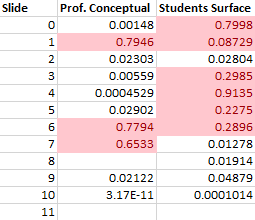


|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ExpertMeaningful:ExpertSurface | | | NoviceMeaningful:NoviceSurface | | |
| Slide | SE | t-ratio | p-value | SE | t-ratio | p-value |
| 0 | 0.0595 | 0.144 | 1.0 | 0.0595 | 1.833 | 0.469 |
| 1 | 0.0649 | -0.517 | 1.0 | 0.0530 | -4.233 | 0.0038 |
| 2 | 0.117 | 0.073 | 1.0 | 0.0781 | -2.488 | 0.125 |
| 3 | 0.111 | 1.163 | 1.0 | 0.0586 | -4.260 | 0.0028 |
| 4 | 0.0635 | 2.927 | 0.066 | 0.0449 | -6.413 | 0.0001 |
| 5 | 0.0948 | 0.383 | 1.0 | 0.0571 | -4.959 | 0.0002 |
| 6 | 0.0644 | 1.799 | 1.0 | 0.0494 | -5.733 | <0.0001 |
| 7 | 0.0438 | -0.110 | 1.0 | 0.0331 | -7.211 | <0.0001 |
| 9 | 0.0687 | -2.201 | 0.221 | 0.0595 | -2.876 | 0.0476 |
| 10 | 0.168 | -1.456 | 0.966 | 0.0751 | -5.864 | 0.0001 |

Chi-Squared

A chi-squared analysis was performed to test novice’s and expert’s transitions in comparison to what would be expected by random chance. Figure 1 displays a plot of the combined transitions for all slides, filtered by professors who made conceptual selections and students who made surface selections. From this plot alone, it is demonstrated that students did not make meaningful transitions at a rate that that statistically varied from random and made more surface level transitions than the experts. Table 3 shows the results of the chi-squared analysis on each slide. A significant p-value indicates that the group is not behaving as expected by random chance. The professors provided significant values on seven of the ten applicable slides while the students provided significant values on five of the eleven applicable slides. Additionally, the experts generated more significant results on seven of the ten applicable slides.

Table 3. Results of the chi-squared test of deviation from random by chance transitions.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Expert | | | Novice | | |
| Slide | X | df | p | X | df | P |
| 0 | 25.125 | 8 | 0.00148 | 4.595 | 8 | 0.800 |
| 1 | 1.679 | 4 | 0.795 | 16.457 | 10 | 0.0873 |
| 2 | 14.664 | 6 | 0.023 | 28.434 | 16 | 0.028 |
| 3 | 10.374 | 2 | 0.0056 | 16.247 | 14 | 0.299 |
| 4 | 20.215 | 4 | 0.00045 | 4.646 | 10 | 0.914 |
| 5 | 10.790 | 4 | 0.029 | 24.349 | 20 | 0.228 |
| 6 | 3.230 | 6 | 0.779 | 25.157 | 22 | 0.290 |
| 7 | 4.173 | 6 | 0.653 | 25.461 | 12 | 0.0123 |
| 8 |  |  |  | 32.505 | 18 | 0.0191 |
| 9 | 14.879 | 6 | 0.0212 | 28.967 | 18 | 0.0488 |
| 10 | 48.347 | 2 | <0.0001 | 49.149 | 18 | 0.0001 |

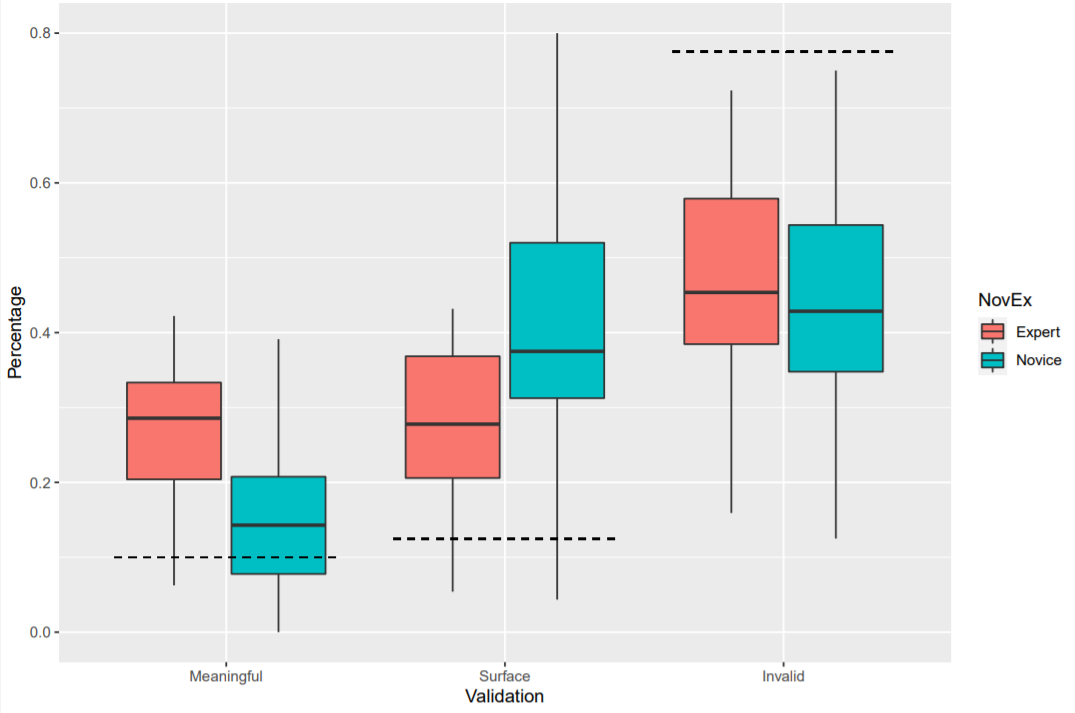


Figure 1. Transition type across all slides. The dotted line represents what would be expected if transitions were made by chance.

Findings:

For this analysis experts were designated as professors who made conceptual selections and novices were designated as students who made surface level selections for each slide (per … validation). All participants who made an invalid selection for a slide was discarded. Any transitions that occurred within the first two seconds of the presentation of a slide were discarded as analysis indicated that both novices and experts made more transitions during this time in an assumed attempt to gain an initial comprehension of the images (in accordance with …)

A two-way ANOVA test with interactions was performed for each slide and results are reported in Table 1. Note that for slide eight no professors made a conceptual selection and slide eleven contained no conceptual connections, so those slides possess no data for this analysis. The comparison between surface and meaningful was significant in eight out of ten slides, however all this shows is that the percentage of surface and meaningful transitions were not equivalent for each slide without accounting for whether the induvial was an expert or a novice. The applicable results come from inspecting the interactions within all four categories, which showed significant results in seven of the ten slides.

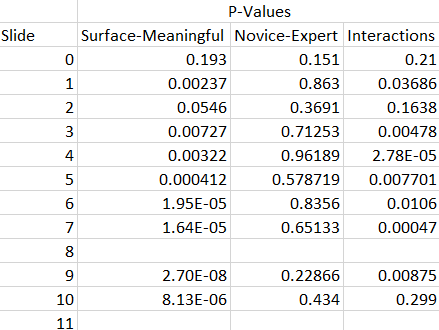
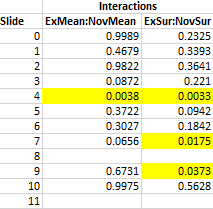


Table 1: P-values from two-way ANOVA test with interactions

*Table 2 shows the p-values for each the comparison for the interactions between novices and experts for the percentage of surface and the percentage of meaningful transitions. A significant p-value indicates that there is a statistical difference between the two percentages. The Tukey adjustment method was used to reduce type one error for this analysis. For the comparison of the percentage of meaningful interactions, only one slide provided a significant result, thus hypothesis one cannot be accepted. The comparison of surface transitions provided more significant results, yet no conclusions can be drawn as only three of the slides provide sufficient p-values.*

**

*Table 2: Interactions from two-way ANOVA test*

Table 3 shows the p-values for each the comparison for the interactions between novices and experts for the percentage of surface and the percentage of meaningful transitions. A significant p-value demonstrates that the factor is dependent on the level of the other factor. The Tukey adjustment method was used to reduce type one error for this analysis. Only one slide resulted in a significant value for the experts, indicating that there is no correlation between the percentage of meaningful and surface level transitions that experts make. The novices displayed a strong dependency between surface and meaningful transitions with eight of the slides providing a significant result.

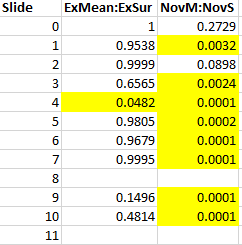


Table 3: Interactions from two-way ANOVA test

A chi-squared analysis was performed to test novice’s and expert’s transitions in comparison to what would be expected by random chance. For example, if a slide contained one complete concept and three surface level concepts, it would be expected that a meaningful transitions would be made 4.166% of the time and a surface transition would be made 12.5% of the time. Table 4 displays a plot of the combined transitions for all slides, filtered by professors who made conceptual selections and students who made surface selections. From this plot alone, it is demonstrated that students did not make meaningful transitions at a rate that that statistically varied from random and made more surface level transitions than the experts.

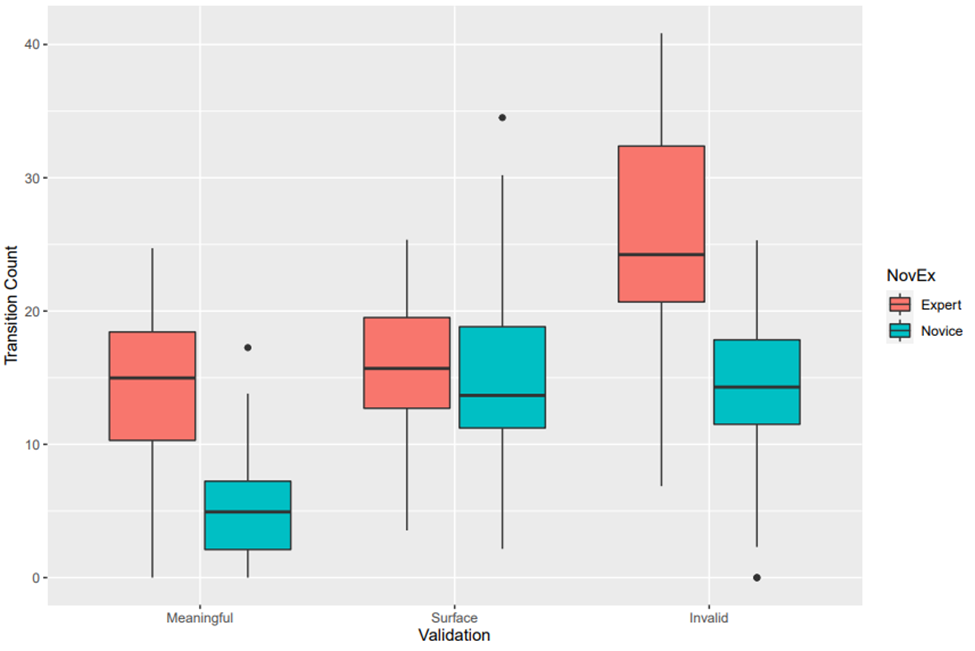


Figure 2. Transitions type average across all slides

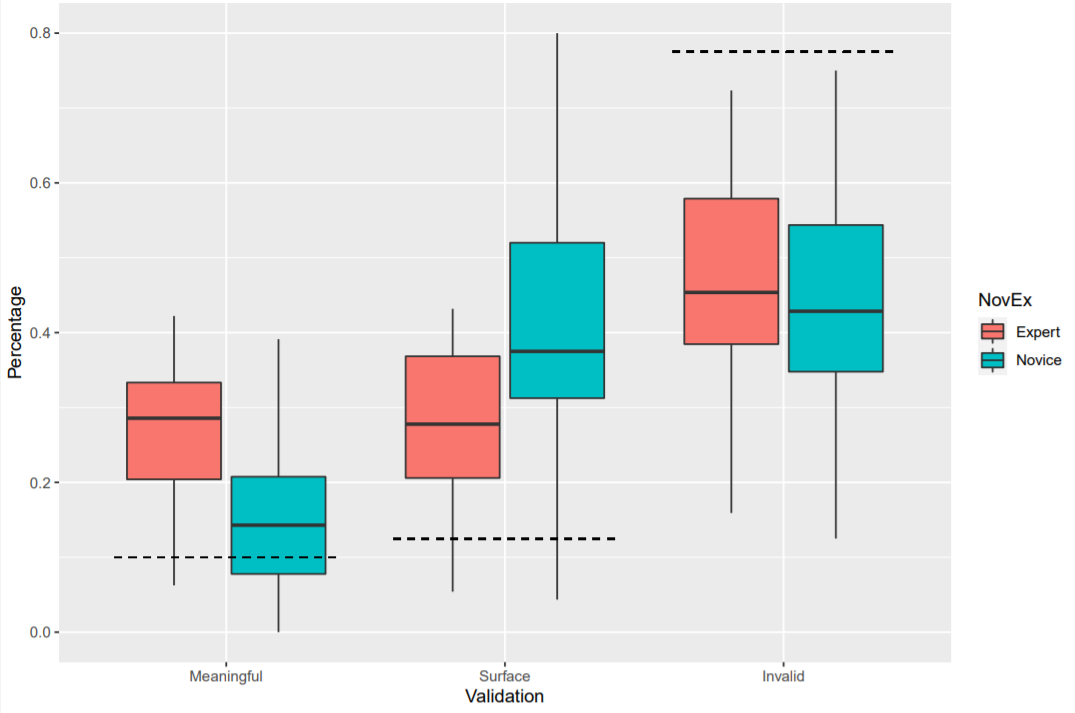


Table 4: Transition type across all slides. The dotted line represents what would be expected if transitions were made by chance.

Figure 2: Total transitions performed by students and professors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Page of Study | Concepts Represented | Number of Conceptual Groupings | Expert Meaningful Selection | Novice Surface Selection |
| 0 | Three phases of matter | 3 | 8 | 5 |
| 1 | Three different solids (various colors) | 3 | 7 | 6 |
| 2 | Equilibrium, solubility, and acids | 3 | 4 | 9 |
| 3 | Three different solids (all white) | 3 | 5 | 8 |
| 4 | Polymers, long alkanes, hydrogen bonding | 3 | 6 | 6 |
| 5 | Solid, liquid, and gas combustion | 3 | 6 | 11 |
| 6 | Solutions of sulfuric acid, ammonia, sodium hydroxide | 3 | 8 | 12 |
| 7 | Titration reaction | 1 | 7 | 7 |
| 8 | Carboxylic acids, ketones, alcohols | 3 | 0 | 10 |
| 9 | Molecular geometry | 1 | 7 | 10 |
| 10 | Thermodynamics | 1 | 2 | 10 |
| 11 | No concept | 0 | 0 | 0 |

Hypotheses:

1. Experts will make a higher percentage of meaningful transitions than novices.
2. Expert transitions will deviate further from what is expected by chance than novice transitions.
3. Expert transitions will deviate further from what is expected by random chance than novices.

Discussion:

The demonstration that the experts make more total transitions aligns with previous studies (Gray, Hope 2015, and Sharma 2012). As expected, experts who selected a complete MER grouping had more transitions between meaningful groupings (Pande 2015). Although experts made a lower percentage of surface transitions, they made about the same number of them. This appears to indicate that experts are making the same connections between surface level groupings as the students, but the increased amount of meaningful and invalid transitions may show that experts are attempting to further search for conceptual groupings. This is also supported by the expert’s lack of interaction between the percentage of surface and meaningful transitions in the ANOVA analysis. One would expect that an increase in meaningful transitions would directly lead to a decrease in both invalid and surface transitions, but the lack of interaction shows that the change only affects the percentage of invalid transitions. This alone would not be particularly interesting if not for the fact that novices showed very strong interactions between the same categories, demonstrating a key difference in behaviors.

The expert’s larger variation from what would have been expected by random chance is also indicative of different behavior between novices and experts. In the chi-squared test showed a significant variation from random in seven out of ten slides while students were only significant in five out of eleven slides. It should also be noted that two statistically insignificant slides for the experts were also insignificant for the novices, allowing for the possibility that slide design affected those two results.

Introduction:

Novices and experts in any field have a working memory that is limited to about seven items of information at once (Miller 1956, Baddeley 1992). This is the information that can be consciously processed; other information is stored in the long-term memory accessing it is a result of the working memory filtering through the stored knowledge. Thus, the difference between novices and experts is not their ability to store more information, but through complex mental schema (Kirschner 2002). A mental schema is formed by combining pieces of information that would normally take up multiple sections of the working memory into one concept that acts as a single piece of information. For example, rain drops and a lake are different representations but based on experience they can be combined into the single schema of water.

This understanding of memory can be applied to chemistry through the triplet relationship. Most chemistry concepts can be portrayed by the triplet relationship which utilizes macroscopic, submicroscopic, and symbolic representations (Johnstone 1982, Gilber and Treagust 2009). Due to their more complex mental schema, an expert in chemistry would more easily draw connections between a macroscopic, submicroscopic, and symbolic representation (Talanquer 2011, Taber 2013, Johnstone 2000). Kozma and Russell’s study presented professors and students with several different representations of chemical equilibrium concept and then asked them to group them into meaningful groups, demonstrating that the professors were more likely to make meaningful groupings (1997). Our study replicated their work, but with a more specific focus on the triplet relationship. The alone would only validate the results of past work. Thus, the primary focus became to demonstrate how experts and novices viewed the representations differently, not simply demonstrate that they were viewed differently.

To understand how novices and experts were approaching the representations, eye-tracking was used. Other eye-tracking studies have demonstrated that novices focus primarily on graphical and model representations while neglecting equations (Stieff 2011) and that experts tend to have larger spatial jumps for their eye fixations (Pande 2015); both indicating a difference in search patterns between experts and novices. If an understanding of the difference between experts and novices can be understood, then it may be possible to develop an eye-tracking assessment that maps a student’s understanding of a concept beyond what would be possible with a traditional assessment.

Research Questions:

What are the differences in how novices and experts transition between groups of images with different representations of a chemistry concept?

Are experts making the same initial connections that the students are and then deeper connections or are they making a completely different set of connections?