Educational Technology Support for Collaborative Learning With Multiple Visual Representations in Chemistry

Martina A. Rau, Educational Psychology, University of Wisconsin – Madison, marau@wisc.edu
Sally P. W. Wu, Educational Psychology, University of Wisconsin – Madison, pwwu@wisc.edu

Abstract: Educational technologies have two features that can enhance collaborative learning. First, they can provide collaboration scripts that adaptively react to student actions and prompt them to engage in effective collaborative behaviors. Second, collaboration often involves multiple visual representations. But many students have difficulties in making sense of representations. Educational technologies can support students in doing so by adapting to how they construct, interpret, and connect representations. We conducted a quasi-experiment with 61 undergraduate chemistry students to test the effectiveness of an adaptive collaboration script that prompts students to discuss visual representations. A control condition collaboratively solved worksheet problems with multiple visual representations without a collaboration script. An experimental condition solved the same problems using an educational technology with the script. The experimental condition showed significantly higher learning gains on a transfer posttest and on complex questions on a midterm exam three weeks later.

Introduction

Educational technologies play an increasingly important role in undergraduate instruction in science, technology, engineering, and math (STEM) (Freeman et al., 2014). One reason for this trend is that practice guides recommend engaging students in authentic problem-solving activities to help them reason about concepts in the same way as experts do (NRC, 2006). Educational technologies offer two key features that may make them particularly effective platforms for such problem-solving activities. First, because experts often solve problems collaboratively (Kozma, Chin, Russell, & Marx, 2000), STEM instruction often involves collaborative activities (Freeman et al., 2014). Educational technologies can provide adaptive support for collaboration, for example by providing collaboration scripts that adapt to student needs (Walker, Rummel, & Koedinger, 2009). Second, experts often use multiple visual representations to solve problems (Kozma et al., 2000). Therefore, STEM instruction often asks students to do the same. For example, chemistry students may collaboratively construct, interpret, and connect ball-and-stick models (Figure 1A) and wedge-dash structures (Figure 1B) when they learn about isomers (i.e., chemical compounds made of the same atoms that differ only in the spatial arrangement of their atoms, which can have dramatic effects on the properties of chemical compounds). Educational technologies can provide adaptive support for learning with visual representations, for example by grading student-generated representations automatically, by providing real-time feedback on students' interpretations of the representations, and by prompting them to connect multiple representations (Rau, 2016a; Seufert, 2003).

Consequently, combining adaptive support for collaboration with adaptive support for using visual representations may significantly enhance students' learning of content knowledge. The following brief review of prior research shows that this question remains open because (1) research on adaptive collaboration scripts has not focused on supporting students in making sense of visual representations, while (2) research on learning with visual representations has mostly focused on individual learning.

To address this limitation, we conducted a quasi-experiment within a 3-hour lab session in an undergraduate chemistry course. A control condition worked on a traditional version of an activity about isomers. Students collaboratively constructed ball-and-stick models (see Figure 1A) and drew wedge-dash structures (see Figure 1B) on a worksheet. Students in the experimental condition worked on the same activity, except that they drew wedge-dash structures using an educational technology that incorporated an adaptive collaboration script. The script prompted students to collaboratively discuss mistakes they made in their drawings. We tested effects on learning gains assessed with an immediate posttest and a midterm three weeks later.

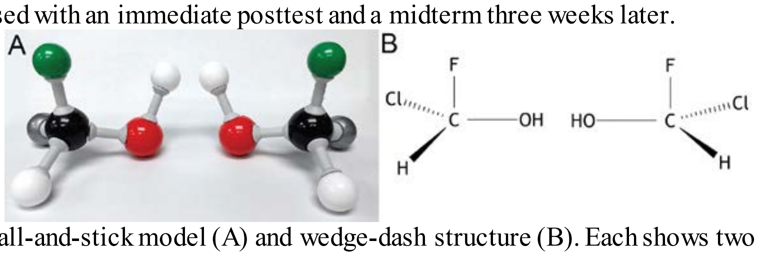

Figure 1. Physical ball-and-stick model (A) and wedge-dash structure (B). Each shows two chlorofluoromethanol isomers that have the same molecular formula but different 3d arrangement of the atoms.

Adaptive collaboration scripts

Collaboration can significantly enhance students' learning, but it is not always effective (Lou, Abrami, & d'Apollonia, 2001). The effectiveness of collaborative activities depends on the quality of interactions among students. They need to actively co-construct meaning, for instance by discussing divergent views and sharing information rather than splitting the work (Miyake & Kirschner, 2014). Students often fail to spontaneously engage in effective collaborative behaviors (Lou et al., 2001).

Collaboration scripts provide an effective means to support collaboration by suggesting sequences of interactions (e.g., analyze the problem, critique partner's analysis, respond to critiques), posing questions for students to discuss (e.g., do you understand the problem?), or prompting them to engage in particular behaviors (e.g., ask your partner to explain the rationale for the solution). Such collaboration scripts can significantly improve the quality of students' collaboration (Fischer, Kollar, Stegmann, & Wecker, 2013). However, results on students' learning of content knowledge are mixed. Several studies found null effects on content knowledge—even if collaboration quality was improved (e.g., Stegmann, Weinberger, & Fischer, 2007; Walker et al., 2009).

The lack of evidence for the effectiveness of collaboration scripts for learning of content knowledge has been attributed to the fact that they do not adapt to students' needs for support. That is, scripts may provide too much or too little support, or support at the wrong time (Rummel, Walker, & Aleven, 2016). Inadequate support can have negative effects on students' affect because they may perceive it as annoying or distracting (Rummel et al., 2016). In contrast, human instructors adapt the amount, timing, and type of support to students' state (e.g., current knowledge level) (Gweon, Rose, Carey, & Zaiss, 2006).

Educational technologies can make adaptive collaboration support scalable by tailoring collaboration scripts to the students' needs (Walker et al., 2009). At a technical level, adaptation is achieved by computational model that detects the students' needs in real time and formalizes the procedure for tailoring support to these needs. For example, the model may infer the students' current knowledge level from their action (e.g., an answer to a problem). Based on the inferred knowledge level, the model can dynamically adjust the amount, timing, and type of support the collaboration script provides (Magnisalis, Demetriadis, & Karakostas, 2011).

Thus far, evidence for the effectiveness of adaptive collaboration scripts for students' learning of content knowledge is mixed (Magnisalis et al., 2011). While some studies show that adaptive collaboration scripts enhance students' learning of content knowledge (Karakostas & Demetriadis, 2011), several studies have failed to show that activities with adaptive collaboration scripts are more effective compared to activities with non-adaptive collaboration scripts and compared to individual learning (e.g., Walker et al., 2009). We are not aware of studies that compared adaptive collaboration scripts to collaborative activities without scripts.

Support for learning with visual representations

Many collaborative activities involve visual representations. Indeed, visual representations and collaborative activities may mutually enhance one another. On the one hand, visual representations can enhance the quality of collaboration. Visual representations allow students to externalize their reasoning, which can reduce cognitive load in the group (Kirschner, Paas, & Kirschner, 2010). Further, externalizing reasoning through visual representations can help the group reach a consensus about how to explain a complex concept or how to solve a task (Suthers & Hundhausen, 2003). On the other hand, collaboration can enhance students' ability to make sense of visual representations. When working individually, students often fail to spontaneously reflect on their understanding of visual representations (Ainsworth, Bibby, & Wood, 2002). When students collaborate with visual representations, they may realize that they hold divergent views on how to interpret, construct, or connect visual representations. This, in turn, may prompt students to engage more deeply in making sense of the representations (Gnesdilow, Bopardikar, Sullivan, & Puntambekar, 2010).

Helping students make sense of visual representations is a key goal of STEM instruction (Ainsworth, 2008; NRC, 2006). Because any individual visual representation shows only a particular aspect of the concepts, instruction typically uses *multiple* visual representations that depict complementary information (Ainsworth, 2008). Besides understanding how each representation depicts information, students need to make connections among the different representations to integrate this information into a coherent mental model (Rau, 2016a). Connection making is a major stumbling block that interferes with students' learning of content knowledge in many STEM domains (Ainsworth, 2008). For example, in chemistry, failure to make connections among representations can yield misconceptions that interfere with learning of crucial concepts (De Jong & Taber, 2014). In the example in Figure 1, if students fail to understand that the wedge-dash structure on the left is *not identical* to the ball-and-stick model on the right, they may incorrectly infer that the melting point of a sample that contains both isomers is equal to the melting point of a sample that contains only one of the isomers.

Much research shows that educational technologies can enhance students' learning of content knowledge by helping them make sense of visual representations (e.g., Ainsworth, 2008). Effective technology-based support typically provides real-time feedback on student-generated visual representations (Rau, 2016b), asks them to map representations to concepts (Seufert, 2003), and prompts them to explain connections between representations (Rau, 2016b). Experiments show that such technology-based support can enhance students' learning of content knowledge compared to educational technologies without such support (Seufert, 2003).

Two limitations of research on learning with visual representations need to be addressed. First, the effectiveness of technology-based support over traditional activities with visual representations remains to be shown. We are not aware of a study that study has systematically compared technology support for sense making of representations to traditional activities without an educational technology. Second, prior research has mostly focused on individual students in using visual representations. This stands in contrast to the fact that visual representations are often used collaboratively for problem solving in STEM instruction, as discussed above.

Research question

In sum, educational technologies can enhance learning by prompting *individual sense making* of visual representations and by *scripting collaboration*. Prior research has not investigated whether an educational technology can enhance learning by prompting *collaborative sense making* of visual representations. Further, research has not compared educational technology support for visual representations or for collaboration to traditional activities without technology support. Therefore, we investigate the following question: Does a technology-based adaptive collaboration script that prompts students to collaboratively make sense of visual representations enhance learning of content knowledge?

Methods

Participants and setting

To address this question, we conducted a quasi-experiment with 69 students in an undergraduate chemistry course at a university in the U.S. Midwest. The course involved two weekly 50-minute lectures, two weekly 50-minute discussion sessions, and one weekly 3-hour lab session. The lecture was attended by all students. Lab and discussion sessions were held in smaller sections; namely four sections of about 18 students each. The lab and discussion sessions were led by two teaching assistants (TAs) who went through the same training program at the beginning of the semester. During the semester, students worked in small groups of 2-3 students during discussion and lab sessions. Our quasi-experiment took place in the lab session in week 5 of the semester.

Experimental design

We assigned two of the four lab sections of the course to the control condition ($n = 37$ students) and two to the experimental condition ($n = 32$ students). Students selected lab sections at the beginning of the semester so that they fit well into their class schedule. We do not have any reason to believe that systematic differences exist between sections. In addition, we took the following steps to ensure equivalency of the conditions. To counterbalance potential effects of class period, each control session was held concurrently with an experimental session. To counterbalance TA effects, each TA led one control and one experimental session. We also counterbalanced the sequence in which the TAs led control and experimental sessions. Both conditions worked on problems collaboratively in the same small groups as in discussion and lab sessions throughout the semester.

Control condition

The control condition received the traditional version of the problem-solving activities: a worksheet that consisted of ten multi-step problems about isomers. In each problem, students had to construct physical ball-and-stick models that represent specific molecules. Students worked on this step collaboratively, using a shared modeling kit to construct these models. After constructing each model, they had to draw a wedge-dash structure of the same molecule. Students drew the structures individually on their own worksheet, but they were encouraged to consult with their partner. Each activity also required students to answer conceptual questions about the molecule. Students wrote down their answers individually, again while being encouraged to consult with their partner. At the end of the 3-hour lab session, students handed their worksheets to the TAs who provided written feedback on the problem solutions and on the wedge-dash drawings in the following week's lab session.

Experimental condition

The experimental condition received the technology-enhanced version of the same problems. To ensure equivalency to the worksheet version, the technology-enhanced problems contained the same steps, the same conceptual questions, and the same molecules. Problems were presented in the same order and required students to build the same physical ball-and-stick models. TAs led the sessions in the same way as for the control sessions (e.g., they were available answer questions about the problems). The difference to the control condition was that problems were presented and answered within an educational technology, shown in Figure 2. Students used the educational technology to draw wedge-dash structures and to answer conceptual questions via mouse and keyboard. The technology incorporated an adaptive collaboration script that prompted students to discuss specific concepts when they made a mistake in their wedge-dash drawing. At a technical level, the script used a computational model that detects conceptual errors students often make when drawing a wedge-dash structure or answering conceptual questions. When the computational model identified an error and a misconception that may have led to this error, the educational technology highlighted the feature of the wedge-dash structure that students had drawn incorrectly and prompted students to discuss the concept with their partners while using the ball-and-stick model.

In sum, the only difference between experimental and control conditions was that students in the experimental condition drew wedge-dash structures using an educational technology with an adaptive collaboration script. The script changed the nature of the collaboration in several ways. First, the timing of feedback differed: while the control condition received written feedback on their worksheets in the following week, the experimental condition received immediate feedback from the technology. Second, the form of feedback differed: while the control condition received only correctness feedback, the experimental condition received feedback in the form of collaboration prompts to discuss concepts that students may have misunderstood. Third, the consequentiality of feedback differed: while the control condition did not have to revise their answers, the experimental condition had to submit a correct answer before students could continue.

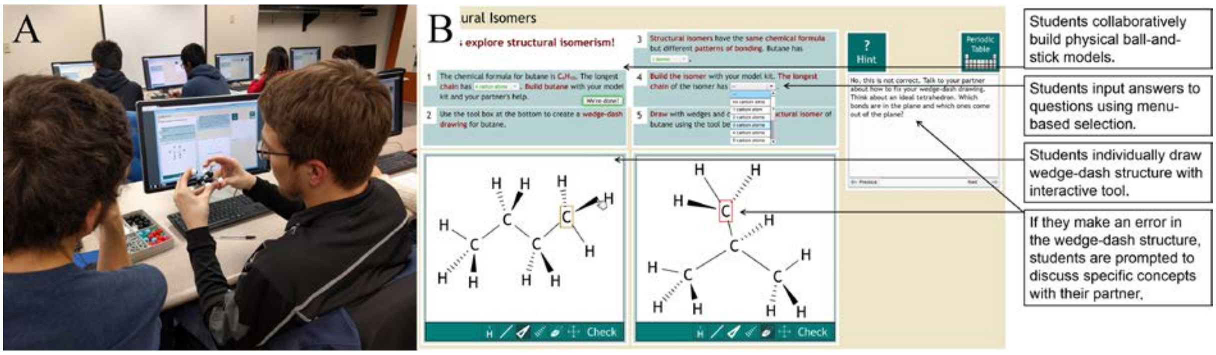

Figure 2. Students in the experimental condition built physical ball-and-stick models (A) and drew wedge-dash structures in an educational technology (B).

Assessments

To assess students' learning of content knowledge, we created a *pretest* and *posttest* on isomerism concepts. The test had two scales. The *reproduction* scale had six multiple-choice items that assessed students' ability to recall and understand the concepts (i.e., levels 1 and 2 of Bloom's taxonomy, as defined by Anderson & Krathwohl, 2001). The *transfer* scale had four multiple-choice items that assessed students' ability to apply and analyze the concepts (i.e., levels 3 and 4 of Bloom's taxonomy). Hence, the reproduction scale assessed simple concepts; the transfer scale assessed complex concepts. Two versions of the test were counterbalanced across pretest and posttest. The tests were optional, but students received course credit for completing them.

To assess students' long-term retention of content knowledge, we used data from two exams that were provided as part of the course. A *pre exam* in the second week of the semester assessed students' prior understanding of chemistry concepts that they may be expected to have covered in high school courses. A *midterm exam* in the eighth week of the semester (i.e., three weeks after the experiment) assessed students' understanding of the chemistry concepts covered in the course thus far. We focused on one question on the midterm exam that assessed the isomerism concepts covered in the lab session in which we conducted our quasi-experiment. This *isomerism question* was one of five advanced questions on the midterm exam, and students had to choose three of these five advanced questions. This question asked students to draw wedge-dash structures and to transfer their knowledge about isomers to novel tasks. We coded students' responses to this question by giving points for each of 20 aspects that were correctly drawn. In addition, we coded for errors that indicated students' difficulties in remembering the target chemistry concepts (level 1 in Bloom's taxonomy), to understand and apply the concepts (level 2 and 3 in Bloom's taxonomy), to analyze and evaluate the concepts (levels 4 and 5 in Bloom's taxonomy), and to make novel inferences (level 6 in Bloom's taxonomy).

Procedure

Figure 3 shows how the experiment aligned with course activities in the entire semester (i.e., two weekly 50-minute lectures, two weekly 50-minute discussion sessions, and one weekly 3-hour lab session). In the second week of the semester, students took a pre exam. A lecture in the fourth week of the semester covered stereoisomerism and related concepts. Our experiment took place in the fifth week. The pretest was made available online three days prior to the lab. Up to this point, all course activities were identical for students in the control and experimental conditions. Then, students attended the version of the 3-hour lab session that corresponded to their condition. All following activities were again identical for both conditions. On the following day, the posttest was made available online for three days. The following discussion and lecture sessions did not focus on isomers. The midterm exam was given in the eighth week of the semester.

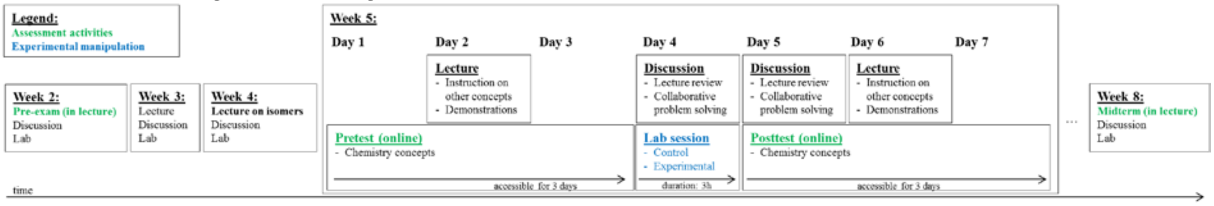

Figure 3. Timeline of assessment (green) and experimental manipulation (blue) in the chemistry course.

Results

Prior checks

As mentioned, students were free to choose whether or not to complete the pretest and posttest for extra course credit, and whether to choose the isomerism question on the midterm exam. Therefore, we first tested for differences between students who chose to complete the tests to those who did not. Eight students did not complete the pretest and posttest, yielding $N = 61$ for these analyses ($n = 30$ in the control condition, $n = 31$ in the experimental condition). Students who did not choose to complete the pretest and posttest did not differ from included students on their pre-exam scores ($F < 1$). Forty students chose to complete the isomerism question ($n = 20$ in the control condition, $n = 20$ in the experimental condition). Students who did not choose this question did not differ from students who chose it on reproduction pretest, $F(1, 55) = 2.647$, $p = .109$, or transfer pretest ($F < 1$), but had significantly lower pre-exam scores, $F(1, 55) = 4.383$, $p = .031$, p. $\eta^2 = .074$. [1]

Because we developed the pretest and posttest specifically for this experiment, they had not been evaluated. Therefore, we conducted a factor analysis to evaluate the separation of the reproduction and transfer scales. A factor analysis showed that a two-factor model that separates the reproduction and transfer scales had a better model fit than a one-factor model. A reliability analysis showed that the reproduction scale had poor reliability (Cronbach's $\alpha = .525$), whereas the transfer scale had good reliability (Cronbach's $\alpha = .851$).

Next, we tested for differences between conditions prior to the experiment. There were no significant differences on pre exam ($F < 1$), reproduction pretest, $F(1, 59) = 1.190$, $p = .280$, or transfer pretest ($F < 1$).

Finally, we tested whether students' understanding of isomerism improved as a result of the interventions. To this end, we used a repeated-measures ANOVA with test time (i.e., pretest and posttest) as the repeated within-subjects factor. Pre-exam scores were not a significant predictor and were hence not used in this analysis. There was no significant effect of test time on the reproduction test ($F < 1$). There was a significant effect of test time on the transfer test, $F(1, 59) = 8.776$, $p = .004$, p. $\eta^2 = .128$, showing that students' ability to transfer knowledge about isomers to novel tasks improved significantly from pretest to posttest.

Differences between conditions on learning outcomes

To test whether the adaptive collaboration script enhanced learning of content knowledge, we used an ANCOVA with condition as independent factor, scores on the reproduction posttest and transfer posttest as dependent measures, and scores on the respective pretests as covariate. The pre exam was not included because it was not a significant predictor. Figure 4 shows the estimated marginal means on the posttests that control for pretest. There was no significant effect of condition on the reproduction posttest ($F < 1$), suggesting that the adaptive collaboration script did not enhance knowledge reproduction. There was a significant effect of condition on the transfer posttest, $F(1, 59) = 4.256$, $p = .044$, p. $\eta^2 = .068$, such that the experimental condition outperformed the control condition. This suggests that the adaptive collaboration script enhanced knowledge transfer.

Next, we tested the effect of condition on overall midterm exam scores using an ANCOVA with condition as the independent factor, scores on the midterm exam as dependent measure, and scores on the pre exam as the covariate. We included scores on the pre exam as a covariate in this model because they were a significant

[1] We report effect sizes using p. η^2: p. η^2 of .01 corresponds to small, .06 to medium, and .14 to large effects.

predictor of students' midterm exam scores. The reproduction pretest and transfer pretest were not included because they were not significant predictors. Results revealed no significant differences on the overall midterm exam scores ($F < 1$). Using the same ANCOVA model to test for differences on the isomerism question for the 40 students who chose this question, we found no differences between conditions on this question ($F < 1$).

A more fine-grained assessment was provided by the errors on the isomerism question, which indicated difficulties in using the isomerism concepts with respect to Bloom's taxonomy levels 1 (remember), 2-3 (understand/apply), 4-5 (analyze/evaluate), and 6 (novel inferences). The same ANCOVA model showed no effects of condition on level 1-5 errors (Fs < 1), suggesting that the adaptive collaboration script did not enhance students' learning of concepts of simple to medium complexity. There was a significant effect on level-6 errors, $F(1, 33) = 4.272$, $p = .047$, p. $\eta^2 = .115$, such that the control condition made more level-6 errors (i.e., difficulties in making inferences about complex concepts). This result suggests that the adaptive collaboration script enhanced students' learning of complex concepts and that this effect persisted three weeks after our quasi-experiment.

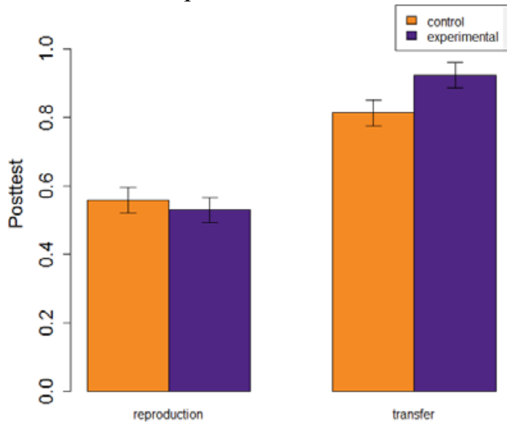

Figure 4. Estimated marginal means for control condition (orange) and experimental condition (purple) on reproduction and transfer posttest, controlling for pretest. Error bars show standard errors of the mean.

Discussion

We conducted a quasi-experiment to test whether an adaptive collaboration script can enhance students' learning of content knowledge from problems that involve connection making among visual representations. Results on students' learning outcomes show a medium-size advantage of the adaptive collaboration script on the transfer posttest over the traditional worksheet version of the same activity. There were no effects on students' scores on the reproduction posttest. There were no effects on overall midterm exam scores or on the isomerism question on the midterm exam three weeks after our experiment. Yet, a fine-grained analysis of the isomerism question showed a medium-sized reduction of errors for the experimental condition on questions that required students to make novel inferences based on complex concepts. This suggests that an adaptive collaboration script can enhance learning with visual representations, but that this effect is confined to complex concepts.

These findings extend prior research on *individual sense making* of visual representations. There is abundant evidence that connection making among visual representations is a difficult but crucial mechanism through which students acquire content knowledge. There is also abundant evidence that educational technologies can enhance students' learning of content knowledge by helping them make sense of the connections. Even though much prior research suggests that collaboration can enhance students' connection making, a limitation of this research is that it has focused mostly on individual rather than collaborative learning. Our findings provide a first affirmation that prompting students to collaboratively make sense of connections when they encounter difficulties in making connections can enhance their learning of content knowledge.

Our findings also extend research on *collaborative learning*. Even though effects of collaboration scripts on the quality of students' collaboration are well established, few studies have found effects on learning of content knowledge. We show that an adaptive collaboration script can significantly enhance learning of content knowledge, compared to a traditional version of the same problems without a collaboration script. Specifically, adaptive collaboration scripts that focus students' collaboration on connection making among visual representations when they struggle with the connections may be effective.

We found effects on complex concepts (i.e., the transfer scale of the posttest and on level-6 concepts on the isomerism question on the midterm exam) but not on simpler concepts (i.e., the reproduction scale of the posttest and lower-level concepts on the isomerism question). The fact that we did not find effects on overall midterm exam scores is not surprising because the midterm exam contained questions about all content covered up to the midterm, and not just on the content covered in the lab session in which we situated our quasi-experiment. The null effects on the reproduction scale of the test may result from the fact that we did not see significant learning gains on this test, which may in turn result from poor reliability of this scale. In future research, we plan to revise the reproduction scale of the test.

The fact that we found effects on the scales that assessed complex concepts can be interpreted in light of research on sense making of visual representations. Integrating information from multiple visual representations is more important for learning of complex concepts than for simple concepts. For this reason, it seems plausible that students are more likely to make mistakes when connection making involves complex concepts. Further, they may be more likely to hold divergent views on complex concepts. Hence, collaboration that yields deeper engagement in connection-making processes may pay off more for complex than for simple concepts.

The finding that effects of the adaptive collaboration script are confined to complex concepts can also be interpreted in light of research on collaborative learning. Discussing complex concepts is cognitively demanding. External representations can be used to off-load these cognitive demands (Kirschner et al., 2010). Hence, prompting students to focus collaborative interactions on the visual representations may benefit their learning of complex concepts more so than their learning of simple concepts. Thus, if complex concepts require connection making more so than simple concepts and if collaboration can help students make these connections, we expect adaptive collaboration scripts to be more effective for complex than for simple concepts.

Limitations

Our findings should be interpreted in the context of the following limitations. First, quasi-experimental designs provide less stringent causal evidence than randomized control trials. Even though we found no differences between conditions prior to the experiment and took steps to ensure equivalency of conditions, unmeasured differences may have affected the results. Hence, a randomized control trial should replicate the results.

Second, while most students completed the pretest and posttest, eight students did not. Even though we did not find differences between these students, it is possible that they differed in unmeasured aspects. Further, students who chose not to complete the isomerism question had lower pre-exam scores, so we do not know whether findings on this question generalize to students with low prior knowledge. We suggest that future research should replicate our findings in a setting that allows for compulsory testing.

Third, our quasi-experiment investigated whether a carefully designed educational technology that contains an adaptive collaboration script is more effective than a traditional version of the same activity. We did not attempt to compare collaborative to individual learning and hence cannot conclude that adaptive collaboration scripts are more effective than individual learning with or without the technology. Likewise, we did not aim at comparing an educational technology with an adaptive collaboration script to an educational technology without a script. Similarly, we did not compare non-adaptive scripts to adaptive scripts. Therefore, we cannot conclude that an adaptive collaboration script enhances the effectiveness of educational technologies. Finally, we did not compare different versions of adaptive collaboration scripts. Hence, we cannot conclude that scripts that adapt to connection making are more effective than scripts that adapt to other aspects of collaboration.

Finally, although we consider the realistic context a particular strength of our study, it limits the conclusions we can draw. Of particular importance may be that students had worked in the same groups since the beginning of the semester and may have had an established collaboration routine. It is possible that the adaptive collaboration script was not maximally effective in altering this routine. Because we did not assess collaboration quality, future research should examine the effects of adaptive collaboration scripts on collaboration quality and examine if a script introduced before students establish a collaboration routine may be more effective.

Conclusion

A quasi-experiment in an undergraduate chemistry course shows that an adaptive collaboration script that supports students in making connections among visual representations enhanced their learning of content knowledge more so than a traditional version of the same collaborative activity without a script. Effects were of medium size and were found immediately after and three weeks after the experiment. We extend research on learning with visual representations by showing that an adaptive collaboration script can support sense making of visual representations. We extend research on collaborative learning by showing that an adaptive collaboration script focused on visual representations can enhance learning of content knowledge.

References

Ainsworth, S. (2008). The Educational Value of Multiple-Representations When Learning Complex Scientific Concepts. In J. K. Gilbert, M. Reiner & A. Nakama (Eds.), *Visualization: Theory and Practice in Science Education* (pp. 191-208). Netherlands: Springer.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the Effects of Different Multiple Representational Systems in Learning Primary Mathematics. *Journal of the Learning Sciences, 11*(1), 25-61.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman

De Jong, O., & Taber, K. S. (2014). The Many Faces of High School Chemistry. In N. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education* (pp. 457-480). New York, NY: Routledge.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist, 48*(1), 56-66.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Gnesdilow, D., Bopardikar, A., Sullivan, S. A., & Puntambekar, S. (2010). *Exploring Convergence of Science Ideas through Collaborative Concept Mapping*. Paper presented at the 9th International Conference of the Learning Sciences.

Gweon, G., Rose, C., Carey, R., & Zaiss, Z. (2006). Providing Support for Adaptive Scripting in an on-Line Collaborative Learning Environment *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 251-260). New York, NY: ACM Press.

Karakostas, A., & Demetriadis, S. (2011). Enhancing Collaborative Learning through Dynamic Forms of Support: The Impact of an Adaptive Domain-Specific Support Strategy. *Journal of Computer Assisted Learning, 27*(3), 243-258.

Kirschner, F., Paas, F., & Kirschner, P. A. (2010). Task Complexity as a Driver for Collaborative Learning Efficiency: The Collective Working-Memory Effect. *Applied Cognitive Psychology, 25*(4), 615-624.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*, 105-143.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small Group and Individual Learning with Technology: A Meta-Analysis. *Review of Educational Research, 71*(3), 449-521.

Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field. *IEEE Transactions on Learning Technologies, 4*(1), 5-20.

Miyake, N., & Kirschner, P. A. (2014). The Social and Interactive Dimensions of Collaborative Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2 ed., pp. 418-438). New York, NY: Cambridge University Press.

NRC. (2006). *Learning to Think Spatially*. Washington, D.C.: National Academies Press.

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

Rau, M. A. (2016). A framework for discipline-specific grounding of educational technologies with multiple visual representations. *IEEE Transactions on Learning Technologies*. doi: 10.1109/TLT.2016.2623303

Rummel, N., Walker, E., & Aleven, V. (2016). Different Futures of Adaptive Collaborative Learning Support. *International Journal of Artificial Intelligence in Education*.

Seufert, T. (2003). Supporting Coherence Formation in Learning from Multiple Representations. *Learning and Instruction, 13*(2), 227-237.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating Argumentative Knowledge Construction. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421-447.

Suthers, D. D., & Hundhausen, C. D. (2003). An Experimental Study of the Effects of Representational Guidance on Collaborative Learning Processes. *Journal of the Learning Sciences, 12*(2), 183-218.

Walker, E., Rummel, N., & Koedinger, K. R. (2009). Ctrl: A Research Framework for Providing Adaptive Collaborative Learning Support. *User Modeling and User-Adapted Interaction, 19*(5), 387-431.

Acknowledgments

This work was supported by the Wisconsin Alumni Research Foundation, the Wisconsin Center for Education Research, and by the University of Wisconsin – Madison Educational Innovation fund. We thank John Moore, Hannah Bowman, Judy Hines, Abe Wu, and Rachel Bain for their help.