How does Sustaining and Interleaving Visual Scaffolding Help Learners? A Classroom Study with an Intelligent Tutoring System

Tomohiro Nagashima (tnagashi@cs.cmu.edu)
Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Elizabeth Ling (elizabethling@college.harvard.edu)
Harvard University, Massachusetts Hall, Cambridge, MA 02138, USA

Bin Zheng (binzheng@andrew.cmu.edu)
Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Anna N. Bartel (anbartel@wisc.edu)
University of Wisconsin, Madison, 1025 W. Johnson Street, Madison, WI 53706, USA

Elena M. Silla (esilla@udel.edu)
University of Delaware, 16 West Main Street Newark, DE 19716, USA

Nicholas A. Vest (navest@wisc.edu)
University of Wisconsin, Madison, 1025 W. Johnson Street Madison, WI 53706, USA

Martha W. Alibali (martha.alibali@wisc.edu)
University of Wisconsin, Madison, 1025 W. Johnson Street, Madison, WI 53706, USA

Vincent Aleven (aleven@cs.cmu.edu)
Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Abstract

Integrating visual representations in an interactive learning activity effectively scaffolds performance and learning. However, it is unclear whether and how *sustaining* or *interleaving* visual scaffolding helps learners solve problems efficiently and learn from problem solving. We conducted a classroom study with 63 middle-school students in which we tested whether sustaining or interleaving a particular form of visual scaffolding, called anticipatory diagrammatic self-explanation in an Intelligent Tutoring System, helps students' learning and performance in the domain of early algebra. Sustaining visual scaffolding during problem solving helped students solve problems efficiently with no negative effects on learning. However, in-depth log data analyses suggest that interleaving visual scaffolding allowed students to practice important skills that may help them in later phases of algebra learning. This paper extends scientific understanding that sustaining visual scaffold does not over-scaffold student learning in the early phase of skill acquisition in algebra.

Keywords: Visual Representations; Intelligent Tutoring System; Problem Solving; Algebra

Introduction

Studies show that visual representations can benefit learners in interactive learning environments, such as in Intelligent Tutoring Systems (ITSs) (Rau et al., 2015; Yung & Paas, 2015). Visual representations make complex concepts, which are difficult to understand with only verbal information, more accessible and comprehensible (Larkin & Simon, 1987). Therefore, visual representations are often considered a type of instructional scaffolding (Nagashima et al., 2021; Rittle-Johnson & Koedinger, 2005). When designed well, visual representations support learners in performing and learning from tasks that they would otherwise be unable to solve (Pea, 2004).

Despite the reported effectiveness of visual representations, a persistent instructional challenge involves how to fade visual scaffolding in interactive learning environments (i.e., how and when to reduce the amount and level of scaffolding given to learners, Koedinger & Aleven, 2007; Sharma & Hannafin, 2007). Researchers have argued that scaffolds should serve as temporary support that is provided with an intention to help learners independently solve problems and eventually succeed without the support (Puntambekar & Hubscher, 2005). However, while some studies have investigated questions around the fading of visual scaffolding (Rau et al., 2013, 2010), we do not yet fully understand the effects of sustaining vs. fading visual scaffolding in interactive learning environments. From a cognitive science perspective, an investigation on students' cognitive processes involving learning with and without visual scaffolding would help elucidate how visual scaffolding benefits learners and informs how to design instruction with visual representations.

One domain in which visual representations are commonly used as instructional scaffolding is early algebra (Ayabe et al., 2021; Murata, 2008). A specific type of visual representation called "tape diagrams" has been used widely in practice and empirically evaluated (Booth & Koedinger, 2012; Chu et al., 2017; Nagashima et al., 2020). Tape diagrams use bar-like representations to visualize quantitative relationships (Figure 1). Prior research has shown that integrating tape diagrams in an algebra problem-solving activity enhances students' problem-solving performance, both with interactive technologies (Nagashima et al., 2021) and without (Booth & Koedinger, 2012; Chu et al., 2017).

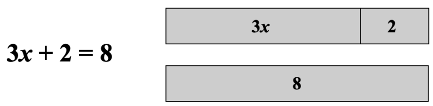

Figure 1: Example tape diagram (for $3x + 2 = 8$).

In the context of early algebra, sustaining or fading visual scaffolding presents an important question both from a scientific and a practical perspective. To date, research investigating learning with visual representations for algebra has either provided visual support all the time or has not provided it at all (Booth & Koedinger, 2012; Nagashima et al., 2020). Practically, this *all-or-nothing* contrast does not reflect how classroom teaching is conducted; many mathematics textbooks mix problems with visuals and without visuals (Fukuda et al., 2021). Research on sustaining or fading visual scaffolding in algebra will extend current scientific knowledge with new insight into whether providing visual support all the time might or might not *over-scaffold* learning. For instance, always providing diagrams may effectively foster conceptual understanding of problem-solving procedures because students can connect the visual information depicted in a diagram with symbolic representations (Nagashima et al., 2020). On the other hand, always using such scaffolding in solving equation problems runs the risk that students become overly reliant on tape diagrams when solving symbolic equations. That is, students' learning might get focused on rather superficial *diagram-to-symbols translation knowledge* that might not help to acquire deeper knowledge related to the use of visual representations. Students might not learn how to strategically solve symbolic equations without the aid of visual representations, which may be detrimental when equations become more complex (for which tape diagrams are no longer useful).

In the current study, we test whether sustaining visual scaffolding to support problem solving (i.e., providing visual scaffolding on all problem-solving opportunities), previously shown beneficial, might over-scaffold student learning in algebra. We compare sustained scaffolding against partial visual scaffolding, which we implemented by *interleaving* scaffolded problems and non-scaffolded problems (i.e., providing visual scaffolding on every other problem).

In what follows, we report findings of the study, which took place in a middle-school in the U.S. To investigate students' learning processes, we conducted analyses of log data from the tutoring system. Specifically, we explored how students' performance (e.g., problem-solving accuracy, time spent) differed when they solved problems *with* and *without* visual support. Further, we conducted Knowledge Component modeling (Nguyen et al., 2019) to better understand how the presence (and absence) of visual scaffolding affected the types of fine-grained problem-solving skills that students practice. The current research extends theoretical understanding of whether and how visual representations scaffold learning and performance in an interactive learning environment.

Anticipatory Diagrammatic Self-Explanation

In the current study, we test the effects of a visual scaffolding strategy called *anticipatory diagrammatic self-explanation*, embedded in an ITS (Nagashima et al., 2021). Anticipatory diagrammatic self-explanation is a form of self-explanation (Chi et al., 1989) with visual representations that aims to support students' learning of problem-solving procedures (Figure 2). In anticipatory diagrammatic self-explanation, students *explain* what to do next in the form of (auxiliary) diagrams (i.e., students would think, "what would be a correct and strategic step to take next?" during problem solving) to support problem solving with the target representation (i.e., symbolic representation). There is evidence that anticipatory diagrammatic self-explanation can lead to more efficient learning (i.e., better performance in the ITS) with comparable posttest performance (Nagashima et al., 2021). However, prior studies provided anticipatory diagrammatic self-explanation for *all practice problems*, leaving open the important question of whether and how partially providing the visual scaffolding may affect students' performance during problem-solving practice and learning.

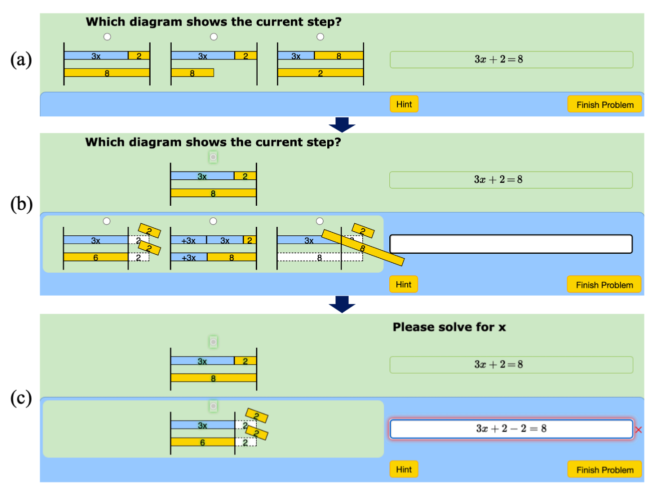

Figure 2: In anticipatory diagrammatic self-explanation, students first (a) select a correct diagram representation for the given equation. Then, students (b) *explain* (by selecting a diagram) a correct next problem-solving step. After selecting the correct representation, students (c) solve the step using symbols (on the right-hand side of the screen). The ITS gives feedback on students' input. Hints are also available for both diagrammatic and symbolic steps.

In the current study, we investigate the following research questions:

RQ1: Does interleaving problems with and without visual scaffolding during algebra problem solving (interleaved condition) lead to better learning (by reducing over-scaffolding), compared to sustaining the visual scaffolding (sustained condition)? Literature on interleaved practice argues that interleaving different problem types may support learning because it requires extra cognitive effort conducive to learning (Rohrer et al., 2015). In the context of anticipatory diagrammatic explanation, interleaving visual scaffolding might help foster deeper thinking about problem-solving procedures because, on problems without visual scaffolding, students would have to plan problem steps on their own without visual support. Such deeper thinking and engagement might result in enhanced conceptual and procedural knowledge (Crooks & Alibali, 2014; Rittle-Johnson & Siegler, 1998). As well, when students receive only problems in which visual scaffolding is provided, they might engage in shallow processing of the content (i.e., *translating* what is shown in the tape diagrams to algebra notation without deeply engaging with their conceptual and procedural meanings). Therefore, we hypothesize: *H1. Students who receive interleaved visual scaffolding will make greater gains in conceptual and procedural knowledge, compared to students who receive the visual scaffolding all the time.*

RQ2: Does providing visual scaffolding at every problem-solving opportunity during algebra problem solving support efficient problem-solving performance in the ITS? RQ2 pertains to students' performance during practice with the ITS. For the current study, we expect that receiving visual scaffolding will improve problem-solving performance in the ITS and that interleaving the visual scaffolding would negatively impact problem-solving performance due to the lack of scaffolding on the half of the problems. Therefore, we hypothesize: *H2. Students who receive sustained visual scaffolding will solve problems in the ITS more efficiently compared to students who receive interleaved visual scaffolding.*

RQ3: How does the visual support influence students' performance in the ITS? We examine RQ3 to uncover how students interact with the visual scaffolding during problem solving. We specifically examine students' performance across the two conditions 1) on problems on which all students, regardless of the visual support frequency, received visual support and 2) on problems in which students with interleaved practice received no visual support whereas students who were given the visual scaffolding all the time did. We investigate where any observed differences between the conditions (if any) come from. We hypothesize: *H3.1. Students in the interleaved condition will not perform differently from those in the sustained condition on problems with visual support* and *H3.2. Students in the interleaved condition will perform worse on problem-solving items in the ITS on problems on which only students in the sustained condition receive the scaffolding.*

Lastly, to gain further insights into what types of skills students practice with and without visual scaffolding, we conducted "Knowledge Component modeling" (a standard technique used in the field of educational data mining, Long et al., 2018; Nguyen et al., 2019). Specifically, we investigate to what extent students might be using overlapping vs. separate knowledge on symbolic steps with diagrams and without diagrams, respectively. By labeling Knowledge Components, or fine-grained problem-solving skills in an intelligent tutor (Koedinger et al., 2012) differently for solving problems with and without the visual scaffolding, we can examine if students' actual performance can be modeled better with such a separation of knowledge. Such an understanding will help uncover possible mechanisms that may influence any learning and performance differences. We hypothesize: *H3.3. A Knowledge Component model that considers problem solving with and without visual scaffolding separately will show a better fit.*

Method

Participants

We conducted an "in-vivo" classroom experiment (Koedinger et al., 2009) at a public middle school in the Eastern United States. Participants were 77 7th-grade students who were taught in five class sections by one teacher. The school is the only middle school in the school district where over 65% of students came from low-income families, and 44.7% of students were considered "below basic" in terms of their academic performance in 2019. We conducted the experiment in May 2021 when the school was operating under a *hybrid* teaching mode due to the COVID-19 pandemic. Thirty students participated remotely from their own home environment and the remaining 47 joined from their classroom with their teacher. The teacher noted that students' prior exposure to tape diagrams was minimal. Of the 77 students, 16 students had Individualized Education Plans (IEP). Students in each class were randomly assigned to either the sustained condition or the interleaved condition. Students with IEPs were separately and randomly assigned to the conditions. Based on teacher-reported information regarding students' regular class participation mode (i.e., remote or in-person), we randomly assigned students to conditions separately for those joining remotely and those joining from the classroom so that the conditions were balanced with respect to these variables.

Materials

Pretest and Posttest A web-based pretest and posttest were developed for the study. The tests were designed to measure students' conceptual understanding and procedural skills (Crooks & Alibali, 2014; Rittle-Johnson & Siegler, 1998). Test items included six conceptual knowledge items (CK) and seven procedural knowledge items, developed partly based on items used in the literature (e.g., Rittle-Johnson et al., 2011). Conceptual knowledge items assessed multiple concepts of conceptual understanding of algebra, such as math equivalence. The procedural knowledge items consisted of four items with no tape diagrams (PK-NoDiagram) and three problems that show a corresponding tape diagram (PK-Diagram) (Figure 3). Two isomorphic versions were created and assigned to students in a counter-balanced way.

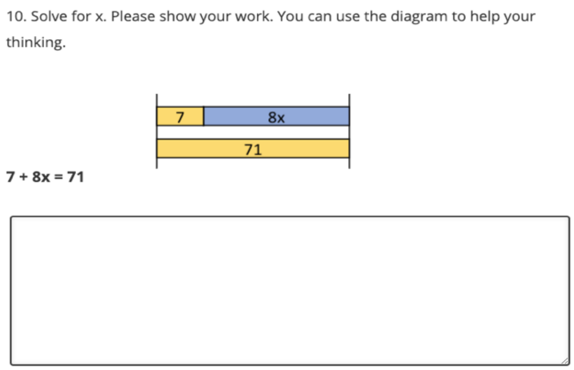

Figure 3: An example procedural item with tape diagrams.

Intelligent Tutoring System with Anticipatory Diagrammatic Self-Explanation An ITS with anticipatory diagrammatic self-explanation (Figure 2) was used in the study. Students in the sustained condition used a version of the ITS that provided anticipatory diagrammatic self-explanation support for all problems whereas those in the interleaved condition used a version that provided such support only for odd-numbered problems (Figure 4). For even-numbered problems, students in the interleaved condition received problems with no diagrammatic steps available. These two ITS versions differed only in whether the ITS provided diagrams or not on even-numbered problems. Regardless of their assigned condition, students received the same set of algebra problems in a fixed order. The following problem types were included in both versions: $x + a = b$, $ax + b = c$, $ax = bx + c$, and $ax + b = cx + d$.

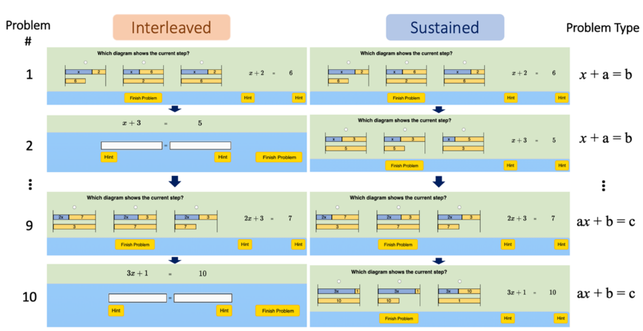

Figure 4: Sustaining and interleaving visual scaffolding in the ITS. In the interleaved condition, students received the visual scaffolding only on odd-numbered problems.

Procedure

The study took place during five regular class periods, in which approximately half of the students and the teacher were present live in the actual classroom, and remote learners and the experimenters joined through a video conferencing system. In the first class session, students worked on the pretest for 20 minutes. Then the experimenter showed students in both conditions a five-minute video describing how to use the ITS and what tape diagrams represent. Starting in the second class period, students spent 15-20 minutes using the assigned ITS version to practice algebra problem solving in each class period (the total ITS learning time in both conditions was approximately 60 minutes). On the final day, students took the web-based posttest for 20 minutes. Students were given access to both ITS versions about a week after the study.

Results

Of the 77 participants who completed the pretest (40 in the interleaved, 37 in the sustained condition), we excluded 14 who did not complete tutor learning, the posttest, or both. The high attrition may be due to the hybrid mode of instruction (e.g., the teacher had to pay attention to both in-person and remote students). The following analyses focus on the remaining 63 students (32 in the interleaved, 31 in the sustained condition). The attrition rate did not significantly differ between the conditions, X^2 (1, N = 77) = .18, p = .67.

Learning

Table 1 presents students' pretest and posttest scores. To test H1, we conducted three separate linear regressions, with posttest scores on conceptual knowledge (CK), procedural knowledge items with tape diagrams (PK-Diagram), and procedural knowledge items without tape diagrams (PK-NoDiagram) as dependent variables, respectively. In all of the models, condition (as a binary variable, coded as sustained = 0, interleaved = 1) and prior knowledge (i.e., pretest scores) served as predictors. There was no significant main effect of condition for CK (β = -0.42, t(62) = -1.22, p = .23), PK-Diagram (β = -0.17, t(62) = -0.77, p = .44), or PK-NoDiagram (β = -0.29, t(62) = -1.12, p = .27). Therefore, H1 was not supported. The evidence did not support the notion that interleaving visual scaffolding leads to greater knowledge gains, nor did it support the notion that sustained visual scaffolding over-scaffolds learning.

Table 1: Students' test scores (standard deviations in parentheses) by condition. PK-D and PK-NoD denote PK-Diagram and PK-NoDiagram, respectively.

Condition	Pretest			Posttest		
	CK	PK-D	PK-NoD	CK	PK-D	PK-NoD
Sustained	2.82 (1.53)	1.27 (1.15)	2.18 (1.42)	3.30 (1.55)	1.45 (1.25)	2.21 (1.45)
Interleaved	2.78 (1.39)	1.03 (1.06)	1.84 (1.22)	2.75 (1.50)	1.12 (1.13)	1.72 (1.45)

Performance in the ITS

To address H2, we ran three separate linear regressions with three problem-solving performance measures that are typically investigated in the ITS literature (average number of hints requested per symbolic step, average number of incorrect/error attempts per symbolic step, average time spent per symbolic step) as dependent variables (Long & Aleven, 2013). To compare the performance measures across the conditions, we only compared students' performance on symbolic steps, and we excluded interactions with the diagram steps. In all models, condition (as a binary variable, coded as sustained = 0, interleaved = 1) and pretest score were included as independent variables. Table 2 shows descriptive data on the performance measures.

There was a significant main effect of condition on the average number of hint requests per step (β = 0.28, $t(60)$ = 2.83, $p < .01$) and average time spent per step (β = 3.32, $t(60)$ = 2.21, p = .03), but not on the number of incorrect attempts made per step (β = 0.31, $t(60)$ = 1.67, p = .10). Overall, students in the sustained condition solved problems with fewer hint requests and less time on symbolic steps, suggesting that students in the sustained condition solved symbolic steps more efficiently than those in the interleaved condition, partially supporting H2.

Table 2: Average performance measures per symbolic step (standard deviations) by condition.

Condition	Ave. number of hints used	Ave. number of incorrect attempts	Ave. time spent
Sustained	0.16 (0.23)	0.45 (0.63)	6.85 (2.70)
Interleaved	0.53 (0.64)	0.86 (0.92)	12.0 (6.92)

Visual Scaffolding Effects

Performance with and without Visual Scaffold For H3.1 and H3.2, we conducted further analyses to unpack how the visual scaffolding might have helped students perform in the learning environment. Specifically, to understand where the performance difference between the conditions came from, we looked at students' performance on *odd-numbered* problems, where students in both conditions received the visual scaffolding, and *even-numbered* problems where only students in the sustained condition received scaffolding (Table 3). Investigating students' performance on odd-numbered and even-numbered problems will allow us to find out if the diagrams' scaffolding effect is only observed for the problems in which the scaffolding is present and how students in the interleaved condition performed differently on problems with no visual scaffolding.

For H3.1, we compared students' performance on odd-numbered problems (i.e., problems with visual scaffolding in both conditions; see Figure 4). We ran three separate linear regressions with the number of hints used per symbolic step, the number of incorrect attempts per symbolic step, and time spent per symbolic step as dependent variables, and condition and pretest scores as predictors. Students in the sustained condition used significantly fewer hints (β = 0.29, $t(60)$ = 2.48, p = .02) and trended towards spending less time (β = 5.87, $t(60)$ =1.89, p = .06). No significant difference was found for the number of incorrect attempts per symbolic step, β = 0.19, $t(60)$ = 0.78, p = .44. Therefore, H3.1 was not supported; students in the sustained condition performed better on problems in which students in both conditions received the same scaffolding.

Table 3: Performance measures per symbolic step (standard deviations) by condition for problems with and without visual scaffolding.

Condition	Odd-numbered problems			Even-numbered problems		
	Hints used	Errors	Time spent	Hints used	Errors	Time spent
Sustained	0.23 (0.31)	0.70 (1.02)	15.1 (9.15)	0.08 (0.19)	0.17 (0.29)	7.51 (5.14)
Interleaved	0.62 (0.75)	0.99 (1.01)	21.4 (16.0)	0.37 (0.55)	0.49 (0.50)	14.6 (10.4)

Then, for H3.2, we compared students' performance on even-numbered problems to test whether students in the interleaved condition performed less well on problems with no visual scaffolding (see Figure 4). Students in the sustained condition requested significantly fewer hints (β = 0.24, $t(60)$ = 2.67, p = .01), made significantly fewer incorrect attempts (β = 0.30, $t(60)$ = 2.85, p = .01), and spent significantly less time (β = 6.64, $t(60)$ = 3.31, p = .01) on symbolic steps. Thus, students in the sustained condition did better on problems in which only those students in the sustained condition received the scaffolding, supporting H3.2.

Knowledge Component Modeling Finally, for H3.3, we conducted Knowledge Component modeling to investigate potential mechanisms that may have influenced the observed differences between the conditions. A Knowledge Component (KC) is defined as "an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks" (Koedinger et al., 2012). Studies on ITSs have used Knowledge Component modeling (i.e., modeling student's knowledge state and growth based on student's performance on a set of KCs) to design and improve instruction in the software (Huang et al., 2021). KC models use a specialized form of logistic regression known as Additive Factors Models (Rivers et al., 2016). Improving KC models is critical for better understanding student learning and performance, and for better designing instructional support in intelligent software. Model fit can be evaluated by three metrics, namely, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and 3-fold cross validation metrics reported as root mean squared error (RMSE), where lower values suggest better model fit (Nguyen et al., 2019). We applied the KC modeling approach to our dataset to investigate whether the scaffolding effect of having diagrams can be manifested in the model fit. To conduct the KC modeling analysis, we used LearnSphere's DataShop (https://pslcdatashop.web.cmu.edu/).

In our ITS log data, the original KC model had nine general algebra problem-solving KCs (e.g., the skill of *subtracting a* *constant term*). To see if the visual scaffolding effect would be manifested in KCs, we created an additional set of KCs that treats the various skills involved in "solving equations with diagrams" as separate skills, depending on whether the problems had visual scaffolding or not. For example, for the skill of *subtracting a constant term,* we included the skills of *subtracting a constant term when diagrams are absent* and *subtracting a constant term when diagrams are present.* This process doubled the total number of KCs, resulting in 18 KCs. We compared the original KC model with an updated model that considers "solving equations in symbols without diagrams" and "solving equations with diagrams," only for the interleaved condition (because the sustained condition had diagrams for all problem-solving opportunities). We found that the updated model improved the model fit on AIC and all the RMSE values (but not for BIC, see Table 4), which suggests that treating problem-solving skills with and without diagrams as distinct better represents the actual student behavior (Stamper et al., 2013). This suggests that students were solving equations using different skills between problems with anticipatory diagrammatic self-explanation and problems without visual support (H3.3 supported).

Table 4: Model metrics values for the original and updated KC models. Three types of RMSE values differ slightly in how the grouping was done in the data.

KC model	AIC	BIC	RMSE (student blocked)	RMSE (item blocked)	RMSE (un-blocked)
Original	4,431.81	4,894.95	0.3682	0.3088	0.3075
Updated	4,347.60	4,933.32	0.3675	0.3048	0.3047

Discussion

Providing the right amount of timely visual scaffolding is a challenging, important instructional design problem. While scaffolding can support performance when the scaffold is present, giving too much scaffolding could result in a detrimental, *over-scaffolding* effect. We investigated the effect of visual scaffolding on students' learning and performance using an ITS for early algebra. Our work focused on anticipatory diagrammatic self-explanation, a form of interactive visual scaffolding that has been shown effective in supporting student performance in an ITS. In the following, we discuss the results from this experiment.

First, providing visual scaffolding for every problem in the ITS did not *over-scaffold* learning. We did not find any difference in posttest performance between students who received *sustained* scaffolding and students who received *interleaved* visual scaffolding. Yet, sustained visual scaffolding led to markedly better problem-solving performance in the ITS. In fact, the students in the interleaved condition did not have a very smooth learning experience in the ITS; for them, problem solving was harder and slower, and did not lead to enhanced learning. The difference in performance between the conditions existed not only on even-numbered problems, in which only students in the sustained condition received the scaffolding, but also on odd-numbered problems, in which students in both conditions received visual scaffolding. These results indicate that the overall performance differences in the tutor did not come only from problems with no visual scaffolding, but rather came from the entire learning experience, including students' interaction with the scaffolded problems.

Why did sustaining visual scaffolding benefit students? The Knowledge Component modeling analysis provides evidence that students in the interleaved condition exercised different types of skills (i.e., Knowledge Components) for problems with visual scaffolding and those without visual scaffolding. Students in the sustained condition, on the other hand, were consistently practicing the skills of "solving problems with diagrams." It may be that students who received the scaffolding for every problem-solving opportunity benefited because their learning experience was focused and consistent.

However, the findings from the Knowledge Component modeling also indicate that students in the interleaved condition were engaged in learning that students in the sustained condition did not practice (i.e., solving equations without visual scaffolding). Given that students eventually need to be able to solve equation problems without visual scaffolding (e.g., more advanced equation problems), it could be that students' practice with interleaved visual scaffolding may lead to better learning outcomes in later phases of equation solving that involve more complicated problem types. The current study did not capture this potential benefit because these later stages were not reached. Future research could explore this possibility.

We acknowledge several limitations of the study. Most important, we are uncertain whether and how far the results will generalize. The current study used a specific form of visual scaffolding in a specific domain in an ITS. Other types of interactive visual scaffolding in other domains might yield different results as they may involve different kinds of cognitive processes in using visual scaffolding. Also, because the study was conducted with a small sample of students at one school which was operating under a hybrid instruction mode, future studies are needed to understand how sustained vs. interleaved visual scaffolding influences learning and performance with more students and with schools using different teaching modes (e.g., in-person teaching).

The current paper extends scientific understanding of how visual scaffolding during problem-solving activities influences student performance and learning. The study yielded evidence that over-scaffolding due to visual scaffolding may not occur very early in skill acquisition, and that fading (specifically, in the form of interleaving) may need to be introduced later in skill acquisition. Practically, the study highlights the benefits of sustaining visual scaffolding to help students have efficient problem-solving experiences.

Acknowledgements

This research was supported by NSF Award #1760922 and by the Institute of Education Sciences, U.S. Department of Education, through Award #R305B150003 to the University of Wisconsin–Madison. The opinions expressed do not represent views of NSF or the U.S. Department of Education. We thank Max Benson, Octav Popescu, Jonathan Sewall, Stephanie Tseng, and all the participating students and teacher.

References

Ayabe, H., Manalo, E., Fukuda, M., & Sadato, N. (2021). What Diagrams Are Considered Useful for Solving Mathematical Word Problems in Japan? *International Conference on Theory and Application of Diagrams* (pp. 79–83). Springer, Cham.

Booth, J. L., & Koedinger, K. R. (2012). Are diagrams always helpful tools? Developmental and individual differences in the effect of presentation format on student problem solving: Development of diagram use. *British Journal of Educational Psychology*, *82*(3), 492–511.

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

Chu, J., Rittle-Johnson, B., & Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving. *British Journal of Educational Psychology*, *87*(2), 273–287.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*(4), 344–377.

Fukuda, M., Manalo, E., & Ayabe, H. (2021). The presence of diagrams and problems requiring diagram construction: Comparing mathematical word problems in Japanese and Canadian textbooks. *International Conference on Theory and Application of Diagrams* (pp. 353–357). Springer, Cham.

Huang, Y., Lobczowski, N. G., Richey, J. E., McLaughlin, E. A., Asher, M. W., Harackiewicz, J. M., Aleven, V., & Koedinger, K. R. (2021). A general multi-method approach to data-driven redesign of tutoring systems. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference* (pp. 161–172).

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. *Handbook of Metacognition in Education*, 897–964.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757–798.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100.

Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 219–228). Springer, Berlin, Heidelberg.

Long, Y., Holstein, K., & Aleven, V. (2018). What exactly do students learn when they practice equation solving?: refining knowledge components with the additive factors model. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 399–408).

Murata, A. (2008). Mathematics teaching and learning as a mediating process: The case of tape diagrams. *Mathematical Thinking and Learning*, *10*(4), 374–406.

Nagashima, T., Bartel, A. N., & Yadav, G., Tseng, S., Vest, N. A., Silla, E. M., Alibali, M. W., & Aleven, V. (2021). Using anticipatory diagrammatic self-explanation to support learning and performance in early algebra. In E. de Vries, J. Ahn, & Y. Hod (Eds.), *15th International Conference of the Learning Sciences* (pp. 474–481). International Society of the Learning Sciences.

Nagashima, T., Bartel, A., Silla, E., & Vest, N., Alibali, M. W., & Aleven, V. (2020). Enhancing conceptual knowledge in early algebra through scaffolding diagrammatic self-explanation. In M. Gresalfi & I. S. Horn (Eds.), *14th International Conference of the Learning Sciences* (pp. 35-43). International Society of the Learning Sciences

Nguyen, H., Wang, Y., Stamper, J., & McLaren, B. M. (2019). Using knowledge component modeling to increase domain understanding in a digital learning game. In *Proceedings of the 12th Internatioanl Conference on Educational Data Mining.* International Educational Data Mining Society.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*(3), 423–451.

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*(1), 1–12.

Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction*, *23*, 98–114.

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

Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus interleaved practice with multiple representations in an intelligent tutoring system for fractions. In *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 413–422).

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing knowledge scaffolds to support mathematical problem solving. *Cognition and Instruction*, *23*(3), 313–349.

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, *103*(1), 85–104.

Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. *The Development of Mathematical Skills*, 75–110.

Rivers, K., Harpstead, E., & Koedinger, K. R. (2016). Learning curve analysis for programming: Which concepts do students struggle with? *ICER'16*, (pp. 143–151).

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*(3), 900–908.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, *15*(1), 27–46.

Stamper, J., Koedinger, K., & Mclaughlin, E. (2013). A comparison of model selection metrics in datashop. In *Proceedings of the 6th Internatioanl Conference on Educational Data Mining.* International Educational Data Mining Society.

Yung, H. I., & Paas, F. (2015). Effects of cueing by a pedagogical agent in an instructional animation: A cognitive load approach. *Journal of Educational Technology & Society*, *18*(3), 153-160.