Collaborative Learning on Multi-Touch Interfaces: Scaffolding Elementary School Students

Lara Johanna Schmitt and Armin Weinberger
l.schmitt@edutech.uni-saarland.de, a.weinberger@edutech.uni-saarland.de
EduTech, Saarland University

Abstract: Multi-touch interfaces allow for direct and simultaneous input by several co-present learners. Additional scaffolding may or may not be needed to ease or problematize tasks that involve intuitive bodily experiences. In this study, a tablet app ("Proportion") is supposed to enable two novices (about 10 years old) to collaboratively construct an understanding of proportional relations. In a 2×2 factorial design ($n = 162$), effects of facilitating strategy prompts (with / without) and problematizing verbalization prompts (with / without) regarding the variables task focus, emotions, quality of dialogue and learning gains have been investigated. While the strategy prompts did not have any significant influence, the verbalization prompts had versatile effects: On one hand, quality of talk was improved, on the other hand, task focus and emotions were negatively affected. Learning gains were limited to near transfer task types and comparable over conditions.

Keywords: collaborative learning, embodiment, proportional reasoning, scaffolding, tablets

Multi-touch interfaces for collaborative embodied learning

Multi-touch interfaces allow for co-present collaborative learning and, specifically, for equal, simultaneous, and direct manipulation of a learning environment (Roschelle et al., 2010). Using a multi-touch interface together can support beneficial forms of interactions, like whole-group discussions, fluid interactions (Alvarez, Brown, & Nussbaum, 2011), equal participation, and joint time on task, while process losses are reduced (Mercier, Higgins, & da Costa, 2014). The direct manipulation is supposed to reduce the "cognitive distance between intent and execution" (Rick, 2012; p. 316) and also enables forms of embodied learning experiences (Schneps et al., 2014), using the body to interact with a tool in order to construct knowledge. Despite the possible benefits of embodied approaches in learning mathematics (Abrahamson & Lindgren, 2014), embodied hands-on learning activities may provoke off-task behavior and distract from the actual learning goals (e.g., Danish, Enyedy, Saleh, Lee, & Andrade, 2015), thus come at the cost of verbalization, abstraction, and reflection of knowledge. Hence, additional scaffolding for reflection may constructively complement, but may also disrupt the embodied learning experiences. To address this issue, the present study examines to what extent scaffolding can support learning in an embodied learning environment or not.

Scaffolding CSCL with collaboration scripts

Supporting learners' (inter-)actions for knowledge construction can take various forms. One widely applied form is scripting of collaborative processes, which focuses on fostering reflection (Kobbe et al., 2007). Building and sharing arguments together may productively coalesce with embodied learning experiences, helping to translate between embodied experience and abstract conceptualization. A large body of research shows that argumentative practices can be facilitated in CSCL scenarios through scripting (e.g., Gijlers, Weinberger, van Dijk, Bollen, & van Joolingen, 2013).

One of the core design questions of scripting to scaffold collaborative learning processes is whether scripts should make the task easier or harder (Reiser, 2004): On one hand, learners often need structural or strategic support to proceed in task solution processes; providing this support, by e.g. reducing the complexity of a task or learning environment, is making the task easier. For example, scaffolding can guide learners to better understand what they have to do, what is a sensible sequence of activities, or where should they focus their attention (Jackson, Krajcik, & Soloway, 1998). Those strategic types of scaffolds are prospective, i.e. directed to future behavior. One goal of scripting is internalization of effective scripts by the learner (Fischer, Kollar, Stegmann, & Wecker, 2013). On the other hand, learners' understanding can be enhanced by problematizing aspects of the tasks; to this end, prompting reflection and verbalization is feasible; contrary to the first way of scaffolding, it is making the task harder (Reiser, 2004). Stopping learners from superficial and fast problem solving and prompting them to engage in verbalizations, aims at fostering deep elaboration of the learning material (King, 1990). The quality of learners' interactive talk has been found to positively relate to learning (Paus, Werner, & Jucks, 2012; Teasley, 1995; Webb, 1989). Scaffolding learners to reflect typically is a retrospective, thus reflective activity, directed into the past.

So potentially, there is a need for enriching embodied learning experiences on multi-touch interfaces with scripting for deliberate activities, reflective collaborative discourse, or both. In this paper, we consider whether prompting learners to apply heuristics that make the task easier and / or prompting learners to verbalize underlying mathematical concepts, abstract from the immediate experience and hence making the task more difficult, can facilitate processes and outcomes of co-located CSCL using multi-touch devices. While positive effects can be assumed, the opposite might happen: Do well-meant prompts actually interfere with intuitive embodied learning environments? While prompts theoretically would improve cognition and problem solving strategies, their sudden appearance during gameplay can also seriously ruin the experience (Wouters et al., 2015). Scaffolding could be both: counterproductive, thus ruining and disrupting the bodily, hands-on learning experience, or complementary, thus improving and supplementing it with processes of encoding that are necessary for memorization and learning.

Proportional reasoning

Proportional reasoning is defined as "reasoning with ratios, rates, and percentages" (Jitendra, Star, Rodriguez, Lindell, & Someki, 2011, p. 731) and is a central topic in mathematics (Boyer, Levine, & Huttenlocher, 2008). Children typically have difficulties with it; particularly handling fractions is a problem (Mix, Levine, & Huttenlocher, 1999). Reinholz, Trninic, Howison, and Abrahamson (2010) blame a lack of senso-motorical, embodied experiences for causing difficulties in handling proportions. Researchers identified typical misconceptions that children face in proportional reasoning tasks: First, application of counting strategies to proportions, especially when concrete units are presented (Boyer et al., 2008). Second, application of addition rules to proportions (Mix et al., 1999). Third, failure to form correct proportional representations from discrete units, that is not building a relative relationship between numerator and denominator of a fraction (Boyer et al., 2008; Mix et al., 1999). Fostering proportional reasoning requires effective learning set-ups. Research showed that collaboration in combination with hypothesis testing has the potential to live up to that (Ellis, Klahr, & Siegler, 1993). Recent developments also aim at including embodied learning experiences. Similar to the Mathematical Image Trainer (Reinholz et al., 2010), we have developed the "Proportion" iPad app that aims at improving children's proportional reasoning by letting them directly manipulate proportional relations.

Research question and hypotheses

This study clarifies the following research question utilizing a 2×2 design: To what extent can collaborative learning with tablets be supported by different types of prompts (facilitating strategy prompts / "STRAT" vs. problematizing verbalization prompts / "VERB"), regarding learning processes and outcomes?

- **Hypothesis 1:** STRAT prompts and VERB prompts and their combination will result in higher task focus. Facilitating the task with the STRAT prompts should help learners to make progress and stay on track. Problematizing the task with the VERB prompts also should direct learners' attention to relevant task features.
- **Hypothesis 2:** The STRAT prompts will induce more positive emotions; the VERB prompts will induce more negative emotions. Making the task easier (STRAT) is supposed to trigger positive emotions like enjoyment, because task progression is facilitated. Making the task harder (VERB) is supposed to result in negative emotions like frustration or anger, because task progression is slowed down by difficult verbal tasks.
- **Hypothesis 3:** The VERB prompts will enhance the quality of dialogic interactions. Explicitly requesting learners to externalize their knowledge and engage in discussions should result in higher transactivity and higher epistemic quality.
- **Hypothesis 4:** STRAT prompts and VERB prompts and their combination will result in higher learning gains, regarding both near and far transfer task types. The internalization of task solving strategies suggested by the STRAT prompts should have a positive impact on learning. Higher-order verbalizations as scaffolded by the VERB prompts impact learning positively by promoting deeper elaboration and multiple perspectives.

Methods

Sample

Participants (fourth graders; mean age: 10.34 years (SD=.55); 50% male) were acquired from seven primary schools in Germany. All participants had a consent form signed by their legal representatives, informing on data collection and analyses. In total, *n*=162 participants took part in the experiments, being tested in four different experimental conditions (control, STRAT, VERB, and VERB-STRAT). χ^2-tests and ANOVAs did not reveal statistically significant differences between conditions regarding the control variables gender, handedness, experience with or owning of a multi-touch device, pre-test or –questionnaire.

Material

The learning environment "Proportion" is an iPad application (Rick, 2012). The interface is designed to afford incorporation of hand / arm movements, i.e. aiming at actively experiencing and embodying proportional relations. The app consists of a fixed sequence of 21 levels with 5 to 23 tasks each. In Proportion, learners control two bars (orange and blue) that are positioned vertically next to each other. To solve the tasks, learners need to resize the bars, so that they are in the right relation to each other as indicated by the associated numbers. See figure 1a for an example: In this case, the bars need to be resized so that the left bar's height would be 2/3 compared to the right one. Once a task is solved, the numbers of the next task appear. An owl acts as a pedagogical agent and provides feedback, e.g. announces "correct" once a task is solved. The owl also voices the varied prompts; see figure 1b and 1c. The prompt versions (A, B, or C; see table 1) alternate in the same fashion for all dyads in the prompted conditions: After the first task of each level, one STRAT (in the STRAT condition), one VERB (in the VERB condition), or both prompts (in the VERB-STRAT condition) appear on the screen.

Table 1: Overview on prompts

Version	Level	STRAT	VERB
A	1, 4, 7, 10, 13, 16, 19	"Tip for all tasks: What is higher, orange or blue? First say it out loud, then you do!"	"Explain to your learning partner: What did one need to do in order to solve the task?"
B	2, 5, 8, 11, 14, 17, 20	"Tip for all tasks: First think and provide an estimate, then set the bars' correct height!"	"Describe to your learning partner: What could one learn in this task?"
C	3, 6, 9, 12, 15, 18, 21	"Tip for all tasks: If the task is hard and you're stuck, what might help is to discuss and talk!"	"Explain to your learning partner: What do all of these tasks have in common?"

Pre- and post- questionnaires and math tests were applied. The pre-questionnaire collected socio-demographic data, previous experiences with multi-touch devices, and attitudes towards math, school, and collaborative learning. The post-questionnaire measured participants' acceptance of the app, subjective learning gain, and aspects of the collaboration. The math test consisted of tasks related to fractions and proportions; the tasks were classified as requiring lower vs. higher levels of transfer (near transfer tasks vs. far transfer tasks). The near transfer tasks were designed to capture the strategies that were used to progress within Proportion. The far transfer tasks aimed at capturing knowledge on proportions and fractions more broadly. At maximum, one could reach 21 points in the math test: 13 for the far transfer tasks and 8 for the near transfer tasks.

(a) (b) (c) (d)

Figure 1. The interface of Proportion (a), example of displaying one STRAT (b) and one VERB (c) prompt, two children collaboratively using Proportion (d).

Seven iPads of the second generation were used for the experiments. Video cameras and microphones recorded participants' interactions with each other and Proportion.

Experimental procedure

The experimental procedure followed a pre-test – intervention – post-test design and has been carried out by one of several trained experimenters. Experiments took place inside the respective schools. Participants were randomly assigned to conditions by lot. After a general welcoming and introduction to the learning session, the participants individually filled in the pre-questionnaire and the pre- math test (10 minutes). Next, the students worked collaboratively with the Proportion app for 40 minutes, see figure 1d; this phase was video-taped. After a 5 minutes break, the participants individually filled in the post-questionnaire and post- math test (10 minutes). Altogether, one experiment cycle covered about two regular school lessons (90 minutes).

Variables

The dependent variables have been aggregated (i.e. averaged) on the dyad level as we cannot assume statistical independence of the dyadic learners. Regarding the analysis of learners' non-verbal behavior and dialogues, we chose to sample every second problem of every second of the up to 21 levels learners reached in the given 40 minutes. This allowed us to focus on continuous interactions (reactions to success, reactions to the prompts for the prompted conditions, and the problem solving process) on typical tasks throughout the learning experience rather than special cases of initial coordination or final conclusions. The samples started with the appearance of "Correct" of the first problem of the level, covered the prompt, extended over the second problem and ended when this problem has been solved. The video samples have been transcribed and video coded using coding schemes that have been developed based on previous work (e.g., Weinberger & Fischer, 2006) and the data at hand. Interrater reliability was measured using Krippendorff's α. Dependent variables analyzed for this contribution comprise task focus (measured as off-task behavior), negative and positive emotions (gestures), transactivity and epistemic quality, and learning gains in near and far transfer tasks (math-tests). Instances of off-task behavior, positive and negative emotions have each been summed up and their average occurrence per coded segment has been calculated, see table 2 for the coding criteria.

Table 2: Coding criteria for nonverbal behavior (video analysis)

Variable	Krippendorff's α	Indicators
Off-task behavior	.95	looking around the classroom, looking into the camera, interactions with participants outside the own group
Positive emotions	.81	clapping into one's own or the learning partner's hands, throwing hands up in the air, clenching the fist, showing thumbs-up
Negative emotions	.77	threatening the iPad, facing the palms upwards, dismissive hand gesture, face-palming

Regarding the transcripts of dialogues, the utterance / turn was the unit of analysis and one category has been assigned to each. Transactivity refers to the extent that participants react to and base their verbal contributions on their partners' contributions. Epistemic quality refers to the content of participants' utterances: Are the utterances off- or on-topic and are they a pure regulation of their interaction or (different levels) of actual task-related explanations?

Table 3: Coding criteria for transactivity and epistemic quality (transcripts of videos)

Variable	Krippendorff's α	Categories	Relative quality
Transactivity	.78	Externalization	0
		Externalization as reaction	0
		Acceptance	1
		Refusal	1
		Elicitation	2
		Integration	4
		Conflict-oriented consensus building	8
Epistemic quality	.90	Off-topic utterance	0
		On-topic: regulation of the interaction	0
		On-topic: concrete task-related regulation	1
		On-topic: abstract content-related regulation	4
		On-topic: Strategies / procedural knowledge	8

To illustrate the coding scheme, here is one example from the data: Student A: "This needs to go higher", student B: "You have to go lower. It needs to be three times the size". Regarding transactivity, we coded externalization for student A, and conflict-oriented consensus building for student B, as student B refutes the partner's suggestion, but also provides an alternative suggestion and justifies it. Regarding epistemic quality, we coded concrete task-related regulation for student A, because the remark is closely tied to what can be seen on the screen, and we coded abstract content-related regulation for student B, because the remark refers to more abstract knowledge ("three times the size"), going beyond what can be seen on the screen. Taking their relative quality into account (see e.g., Teasley, 1997; Weinberger & Fischer, 2006), the raw categories of transactivity and epistemic quality, see table 3, have each been aggregated to global scores.

Results

Hypothesis 1 stated a main effect for both STRAT and VERB as well as an interaction between the prompts on task focus. A two-factorial ANOVA revealed no statistically significant main effect of STRAT ($F(1,76)=.159$, $p=.691$), but a highly significant main effect of VERB: $F(1,76)=18.190$, $p=.000$, $\eta^2=.19$. However, this highly significant effect is contrary to our hypothesis, as off-task behavior is actually reinforced, and not reduced, with the presence of the VERB prompt, see figure 2. The interaction of the prompts was not statistically significant ($F(1,76)=2.451$, $p=.122$).

Hypothesis 2 stated more positive emotions for STRAT and more negative emotions for VERB. A two-factorial ANOVA revealed no statistically significant effect of STRAT on positive emotions: $F(1,76)=1.919$, $p=.170$. The effect of VERB on negative emotions is statistically significant with $F(1,76)=7.019$, $p=.010$, $\eta^2=.09$, see figure 2.

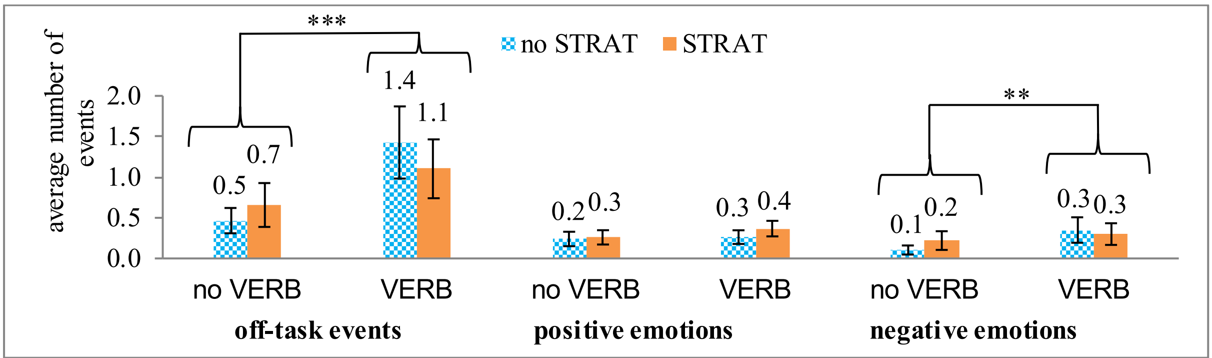

Figure 2. Average number of off-task events, and positive and negative emotions per coded segment; error bars represent 95% CIs; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Hypothesis 3 stated a main effect of VERB on the quality of dialogic interactions, regarding transactivity as well as epistemic quality. Two-factorial ANOVAs were conducted. Regarding transactivity, there was a significant main effect of VERB ($F(1,70)=7.241$, $p=.009$, $\eta^2=.094$), indicating higher transactivity in the presence of the VERB prompts than without it, see figure 3a. Also regarding the epistemic quality score, there was a significant main effect of VERB ($F(1,70)=9.437$, $p=.003$, $\eta^2=.119$), indicating a higher epistemic quality in the presence of the VERB prompts than without it, see figure 3b.

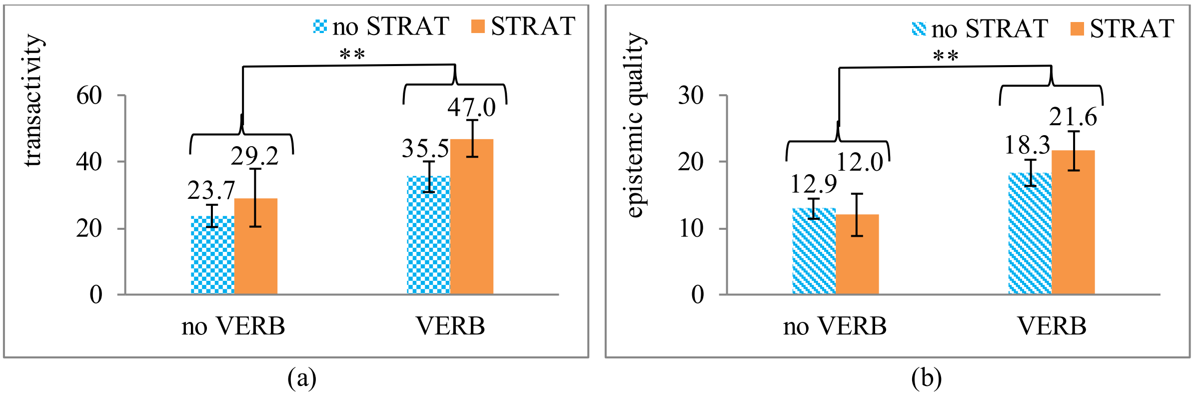

Figure 3. Scores in transactivity (a) and epistemic quality (b); error bars represent 95% CIs; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Hypothesis 4 stated that STRAT and VERB and their combination will increase learning gains, regarding both near and far transfer task types. Two-factorial repeated measures ANOVAs were conducted. Regarding near transfer tasks, there was no statistically significant interaction of STRAT × point in time ($F(1,77)=.585, p=.447$), VERB × point in time ($F(1,77)=.061, p=.805$) or STRAT × VERB × point in time ($F(1,77)=1.502, p=.224$). Only an improvement independent from conditions could be found (main effect of point in time): $F(1,77)=11.179$, $p=.001$, $\eta^2=.13$. Regarding far transfer tasks, there was no statistically significant interaction of STRAT × point in time ($F(1,77)=3.231, p=.076$), VERB × point in time ($F(1,77)=.112, p=.739$) or STRAT × VERB × point in time ($F(1,77)=.003, p=.953$). Contrary to near transfer tasks, no generic improvement from pre- to post-test (main effect of point in time) in the far transfer tasks could be found: $F(1,77)=.174, p=.677$.

Discussion

Collaborative learning with a shared multi-touch device might enable young learners to bodily experience mathematical properties. The learning experiences may be enhanced by different ways of scaffolding, which is being investigated in this study.

Hypothesis 1 claimed a main effect on task focus for the strategy prompts, the verbalization prompts as well as their combination. While the strategy prompts did not have a statistically significant effect on task focus, the verbalization prompts' influence was strong but reverse to our hypothesis: The verbalization prompts actually increased off-task behavior. An interaction effect of the prompts on task focused behavior could not be found. Hypothesis 1 needs to be rejected. On one hand, the strategy prompts could not measurably help learners to stay on track, on the other hand, while we intended to direct learners' attention to relevant aspects of the task with the verbalization prompts, we actually achieved the opposite: Learners got distracted more easily.

Those results get solidified when looking at hypothesis 2 which predicted more positive emotions as a consequence of the strategy prompts and more negative emotions as a consequence of the verbalization prompts. Again, presence of the strategy prompts did not make a difference regarding positive emotions; possibly, any prompt, even if it is there to help learners, might be unwelcome in an embodied learning environment, as it disrupts the ongoing immersive activities. As hypothesized, the verbalization prompts induced more negative emotions, which may be linked to an increase in perceived task difficulty. Hypothesis 2 is being rejected concerning positive emotions affected by the strategy prompts, but confirmed regarding negative emotions affected by the verbalization prompts. Taken together with the results of hypothesis 1 (verbalization prompts increase off-task behavior) those results are disconcerting. Participants being confronted with the verbalization prompts showed about twice as much negative emotions and off-task behavior as participants without it. There seems to be a thin line between enriching a game-like learning environment in a way that facilitates learning, while keeping students' engagement high (Deater-Deckard, El Mallah, Chang, Evans, & Norton, 2014). However helpful prompts like "Explain to your learning partner…" have proven to be in the past, they may still come at the cost of raising the difficulty too much, interrupting the flow, and disengaging learners in specific immersive, embodied CSCL experiences. Hence, these results merit further investigation of when, how, and why prompting may produce these problematic side effects.

Hypothesis 3 predicted higher levels of transactivity and higher epistemic quality caused by the verbalization prompts. Indeed, this could be confirmed. Verbalization prompts had positive medium-sized effects on students' dialogues on both, transactivity as well as epistemic quality. Similar findings have already been established with adult learners, e.g. showing positive effects of scripting on argumentation (Weinberger, Stegmann, & Fischer, 2010). Here we can show that prompts can raise discussion quality even with young learners (see also van Dijk, Gijlers, & Weinberger, 2014).

Hypothesis 4 claimed that strategy prompts and verbalization prompts and their combination will result in higher learning gains. Only for near transfer tasks, a medium-to-large statistically significant improvement from pre- to post- math test over all conditions could be found. There was no interaction with any of the prompts, there was also no significant improvement regarding the far transfer tasks.

The strategy prompts were widely ineffectual. We could not find support for the assumption that they would actually make the task easier. Motivational as well as cognitive factors could have played a role. We could not force participants to thoroughly read the prompts and in the light of the engaging and fun gaming activity, the prompts might have simply been ignored and the gaming activity continued instead. Also, limited cognitive capacities in processing, understanding and applying the suggested strategies might account for the limited effects of the strategy prompts.

The verbalization prompts impacted many variables, in beneficial and detrimental ways. They showed to enhance quality of talk, not to affect learning gain, and to have adverse effects on task focus and negative emotions, presumably having raised the difficulty (too much). Since the verbalization prompts did augment quality of learners' talk (see results of hypotheses 3), a positive effect on learning could have been expected. Different alternative explanations may apply why that was not the case. Verbalization of difficult to verbalize cognitive representations can corrupt performance in non-verbal tasks (verbal overshadowing effect; Schooler, 2002). Also, social preference might have played a role: Van Dijk et al. (2014) found that social aspects are crucial in young learners' collaborative learning, i.e. students who got along well with each other produced better results. In our study, the dyad formation was random, so maybe that is why participants did not always collaborate effectively. Regarding the topical embodiment perspective, however, our study does not find much support for embodiment to be a huge factor in learning. Likewise, limited effects of game-based learning environments have been attributed to the sole acquisition of intuitive knowledge that is never verbalized and translated to out-of-game contexts like formal tests (Wouters et al., 2015). More studies comparing Proportion to a comparable learning environment that does not feature embodied approaches could firstly, shed more light on the actual role of embodiment in learning and secondly, inspire the development of new learning environments that sensibly bring together embodied, hands-on learning experiences with phases of activities that serve abstraction and reflection.

In this study, we investigated how prompting for strategic behavior and/or reflective dialogue effects collaborative learning of proportional reasoning in an intuitive, embodied learning environment. Challenges of incorporating phases of reflection into bodily learning environments while keeping flow and engagement high have been highlighted. Future research could focus on how to further address this issue.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 358-376). Cambridge, UK: Cambridge University Press.

Alvarez, C., Brown, C., & Nussbaum, M. (2011). Comparative study of netbooks and tablet PCs for fostering face-to-face collaborative learning. *Computers in Human Behavior*, *27*(2), 834–844.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490.

Danish, J. A., Enyedy, N., Saleh, A., Lee, C., & Andrade, A. (2015). Science through technology enhanced play: Designing to support reflection through play and embodiment. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015: Vol. 1* (pp. 332–339). Gothenburg, Sweden: The International Society of the Learning Sciences.

Deater-Deckard, K., El Mallah, S., Chang, M., Evans, M. A., & Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11-14 year olds. *International Journal of Child-Computer Interaction*, *2*(3), 101–108.

Ellis, S., Klahr, D., & Siegler, R. S. (1993). *Effects of feedback and collaboration on changes in children's use of mathematical rules.* Paper presented at the Meetings of the Society for Research in Child Development, New Orleans.

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.

Gijlers, H., Weinberger, A., van Dijk, A. M., Bollen, L., & van Joolingen, W. (2013). Collaborative drawing on a shared digital canvas in elementary science education: The effects of script and task awareness support. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 427–453.

Jackson, S. L., Krajcik, J., & Soloway, E. (1998). The design of guided learner-adaptable scaffolding in interactive learning environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI* (pp. 187–194). Los Angeles, CA: ACM Press.

Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving students' proportional thinking using schema-based instruction. *Learning and Instruction*, *21*(6), 731–745.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*(4), 664–687.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(2), 211–224.

Mercier, E. M., Higgins, S. E., & da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, *9*(4), 397–432.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology*, *35*(5), 164–174.

Paus, E., Werner, C. S., & Jucks, R. (2012). Learning through online peer discourse: Structural equation modeling points to the role of discourse activities in individual understanding. *Computers and Education, 58*(4), 1127–1137.

Reinholz, D. L., Trninic, D., Howison, M., & Abrahamson, D. (2010). It's not easy being green: Embodied artifacts and the guided emergence of mathematical meaning. In P. Brosnan, D. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Vol. 6* (pp. 1488–1496). Columbus, OH: PME-NA.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.

Rick, J. (2012). Proportion: A tablet app for collaborative learning. In *Proceedings of the 11th Annual Interaction Design and Children Conference (IDC 2012)* (pp. 316–319). New York, NY: ACM Press.

Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: Impact on students' mathematics learning. *Educational Technology Research and Development*, *58*(4), 399–419.

Schneps, M. H., Ruel, J., Sonnert, G., Dussault, M., Griffin, M., & Sadler, P. M. (2014). Conceptualizing astronomical scale: Virtual simulations on handheld tablet computers reverse misconceptions. *Computers & Education*, *70*, 269–280.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, *16*(8), 989–997.

Teasley, S. (1995). The Role of Talk in Children's Peer Collaborations. *Developmental Psychology, 31*(2), 207–220.

Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361–384). Berlin, Germany: Springer.

Van Dijk, A. M., Gijlers, H., & Weinberger, A. (2014). Scripted collaborative drawing in elementary science education. *Instructional Science*, *42*(3), 353–372.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*(1), 21–39.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*(4), 506–515.

Wouters, P., van Oostendorp, H., ter Vrugte, J., Vandercruysse, S., de Jong, T., & Elen, J. (2015). The role of curiosity-triggering events in game-based learning for mathematics. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and studying domain-specific serious games. Advances in game-based learning* (pp. 191–207). Springer.

Acknowledgements

The authors would like to thank Jochen Rick for designing and programming the Proportion app and the various student assistants supporting the research project.